Helmut Hofer
Alberto Abbondandolo
Urs Frauenfelder
Felix Schlenk
Editors

Symplectic Geometry

A Festschrift in Honour of Claude Viterbo's 60th Birthday

Birkhäuser

Birkhäuser

Helmut Hofer • Alberto Abbondandolo
Urs Frauenfelder • Felix Schlenk
Editors

Symplectic Geometry

A Festschrift in Honour of Claude Viterbo's 60th Birthday

Previously published in *Journal of Fixed Point Theory and Applications* Volume 24, issue 2, June 2022

Birkhäuser

Editors
Helmut Hofer
Institute for Advanced Study
Princeton, NJ, USA

Urs Frauenfelder
Institut für Mathematik
Universität Augsburg
Augsburg, Bayern, Germany

Alberto Abbondandolo
Fakultät für Mathematik
Ruhr-Universität Bochum
Bochum, Nordrhein-Westfalen, Germany

Felix Schlenk
Institut de Mathématiques
Université de Neuchâtel
Neuchâtel, Switzerland

ISBN 978-3-031-19110-7

Mathematics Subject Classification (2020): 53Dxx

This book is published under the imprint Birkhäuser, www.birkhauser-science.com by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Volume 1

Volume 2

The original version of the book has been revised. A correction to this book can be found at https://doi.org/10.1007/978-3-031-19111-4_36

The two articles listed below were intended for inclusion in this volume, but unfortunately they were not ready at the time of printing. They may be accessed at https://link.springer.com/collections/jfjdhejgee.

Kai Cieliebak, Nancy Hingston, and Alexandru Oancea, "Loop coproduct in Morse and Floer homology"

Yakov Eliashberg and Thomas Kragh, "Non-trivial families of Legendrian submanifolds"

J. Fixed Point Theory Appl. (2022) 24:54
https://doi.org/10.1007/s11784-022-00969-2
Published online 11 July 2022

Journal of Fixed Point Theory and Applications

Dedication

Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder and Felix Schlenk

This special volume on the occasion of Claude Viterbo's 60th birthday is a tribute to his mathematical work.

Claude was born in Geneva on April 20, 1961, and later grew up in Paris. After studies at the École Normale Supérieure, he received his Ph.D. working with Ivar Ekeland and François Laudenbach in 1985.

After postdoctoral positions at the Courant Institute in New York (1986–1988) and the Mathematical Sciences Research Institute (MSRI) in Berkeley (1988–1989), he returned to Paris where he held professorships at the Université de Paris-Sud, the École Polytechnique and the École Normale Supérieure. He is now a professor at the Université Paris-Saclay in Orsay.

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder and Felix Schlenk.

Besides being a brilliant mathematician, Claude always showed a strong sense of service to the scientific community. He not only has been a member of numerous scientific boards and hiring committees in France as well as internationally, but also he has been the chairman of the Centre de Mathématiques de l'École Polytechnique (2000–2006), President of the Mathematics Department of the École Polytechnique (2009–2010) and chairman of the Mathematics Department of École Normale Supérieure (2013–2017).

Claude has broadly impacted the development of symplectic geometry/topology and Hamiltonian dynamics as a mathematician, as a mentor, and as a friend. He always considered mathematical research as both a social activity and a solitary one—a deep mathematical discussion with friends followed by a quiet contemplation.

He played a significant role in growing the symplectic community in Europe by his work with students and his service. This is reflected in the wide range of the contributions to this Festschrift, in which the authors express their appreciation, gratitude and friendship.

Showing the same exquisite taste as his Ph.D. advisors, Claude worked successfully on carefully chosen problems opening doors to important developments. In 1987, he proved the Weinstein conjecture for $\mathbb{R}^{2n}$, which was followed by joint work with Andreas Floer and Helmut Hofer utilizing pseudoholomorphic curve methods for more general cases of Weinstein's conjecture. In work with Hofer, before the existence of Gromov–Witten invariants, it was shown in 1992 that the Weinstein conjecture holds in compact symplectic manifolds provided the moduli spaces of rational curves are suitably structured.

In another work in 1992, he used generating functions to construct spectral invariants leading to an alternative construction of a bi-invariant metric on the group of compactly generated Hamiltonian diffeomorphisms in $\mathbb{R}^{2n}$. This contribution led to a powerful tool in symplectic geometry, in particular in the reformulation in terms of Floer homology by Schwarz and Oh.

Another important contribution is the 'Viterbo Transfer Map' introduced in 1999 in his paper on functors and computations in Floer homology. A year later, Claude formulated what is now called the 'Viterbo Conjecture', an intriguing relation between convex and symplectic geometry with far-reaching consequences.

There is also some unpublished work which had significant impact. For example, Claude's work in real algebraic geometry described in V. Kharlamov's Séminaire Bourbaki talk "Variétés de Fano Réelles [d'après C. Viterbo]". It is proved that the real locus of a strongly Fano manifold cannot carry a metric of negative sectional curvature. An alternative shorter proof by Eliashberg, based on SFT neck stretching, extends this to the uniruled case (negatively solving the higher dimensional Nash conjecture in the smooth case). The result is now called the Viterbo–Eliashberg Theorem. In an unpublished sequel to the 'functors and computations' paper, Claude explained that the Floer homology of a cotangent bundle is equal to the homology of the loop space of the underlying manifold. Many improvements were later given by several authors, among them Abbondandolo–Schwarz, Salamon–Weber, Cieliebak–Latschev, and Abouzaid.

His work in 2008 on commuting Hamiltonians and Hamilton–Jacobi multi-time equations with Franco Cardin initiated much work on Poisson rigidity like the symplectic function theory of Entov and Polterovich.

Questions about fillings of contact manifolds start with Gromov's pseudoholomorphic curves paper and in higher dimensions with the Eliashberg–Floer–McDuff result. The 2012 paper of Claude with his former student Alexandru Oancea on the topology of fillings was followed by developments by Ghiggini–Niederkrüger–Wendl, Geiges, Zehmisch et al, and Bowden–Gironella–Moreno.

In recent years, Claude's interests include the relationship between the Floer theoretic and sheaf theoretic approaches to symplectic geometry, stochastic homogenization for variational solutions of the Hamilton–Jacobi equations, as well as C^0-symplectic geometry, in particular the use of barcodes in the understanding of area-preserving homeomorphisms.

We are grateful to Clemens Heine and Patricia Zuberbühler for their help with preparing this Festschrift.

Alberto Abbondandolo, Urs Frauenfelder, Helmut Hofer and Felix Schlenk
April 20, 2022

Helmut Hofer
Princeton
USA
e-mail: hofer@ias.edu

Alberto Abbondandolo
Bochum
Germany

Urs Frauenfelder
Augsburg
Germany

Felix Schlenk
Neuchâtel
Switzerland

J. Fixed Point Theory Appl. (2022) 24:27
https://doi.org/10.1007/s11784-022-00931-2
Published online April 6, 2022

Journal of Fixed Point Theory and Applications

Symplectically convex and symplectically star-shaped curves: a variational problem

Peter Albers and Serge Tabachnikov

Abstract. In this article, we propose a generalization of the 2-dimensional notions of convexity resp. being star-shaped to symplectic vector spaces. We call such curves symplectically convex resp. symplectically star-shaped. After presenting some basic results, we study a family of variational problems for symplectically convex and symplectically star-shaped curves which is motivated by the affine isoperimetric inequality. These variational problems can be reduced back to two dimensions. For a range of the family parameter, extremal points of the variational problem are rigid: they are multiply traversed conics. For all family parameters, we determine when non-trivial first- and second-order deformations of conics exist. In the last section, we present some conjectures and questions and two galleries created with the help of a Mathematica applet by Gil Bor.

Mathematics Subject Classification. 49N99, 53D99, 53A15.

Keywords. Affine isoperimetric inequality, Symplectic space, Infinitesimal rigidity.

1. Introduction

In the paper, we make a step toward expanding some notions and results of equiaffine differential geometry of the plane to symplectic spaces.

Let γ be a smooth, closed, strictly convex, positively oriented plane curve. One can give the curve a parameterization $\gamma(t)$, such that $[\gamma'(t), \gamma''(t)] = 1$ for all t, where the bracket denoted the determinant made by two vectors. This is called an equiaffine parameterization and, accordingly, one defines the equiaffine length of the curve.

The affine isoperimetric inequality between the equiaffine length L and the enclosed area A asserts

$$L^3 \leq 8\pi^2 A,$$

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

with equality if and only if γ is an ellipse; see, e.g., [6,11]. Note that the inequality goes in the "wrong" direction, compared to the usual isoperimetric inequality!

Assume that the origin is inside the curve, then γ is also star-shaped, in addition to being convex. For any parameterization $\gamma(s)$, one has a well-defined (i.e., independent of the parameterization) differential 1-form and a cubic form

$$[\gamma(s), \gamma'(s)]\ \mathrm{d}s, \quad [\gamma'(s), \gamma''(s)]\ \mathrm{d}s^3,$$

such that

$$A = \frac{1}{2}\int [\gamma(s), \gamma'(s)]\ \mathrm{d}s, \quad L = \int \sqrt[3]{[\gamma'(s), \gamma''(s)]}\ \mathrm{d}s.$$

Thus, the affine isoperimetric inequality relates the integrals of these two 1-forms along a convex closed curve.

Let $\gamma(t)$ be a smooth closed curve in the standard symplectic vector space $(\mathbb{R}^{2n}, \omega)$. Call γ *symplectically star-shaped* if $\omega(\gamma(t), \gamma'(t)) > 0$ for all t, and *symplectically convex* if $\omega(\gamma'(t), \gamma''(t)) > 0$ for all t.

Remark 1.1. We point out that an alternative definition of symplectically star-shaped resp. convex is to require $\neq$ instead of $>$ above. That this is actually more general is explained in Sect. 2 where we construct examples of curves with all possible sign combinations.

Similarly to the plane case, we may define two differential 1-forms along a symplectically star-shaped and symplectically convex curve γ by

$$\omega(\gamma, \gamma')\ \mathrm{d}t, \quad \sqrt[3]{\omega(\gamma', \gamma'')}\ \mathrm{d}t.$$

Both forms are well defined and have no zeroes. Inspired by the affine isoperimetric inequality, we are interested in the relative extrema of $\int_\gamma \sqrt[3]{\omega(\gamma', \gamma'')}\ \mathrm{d}t$ constrained by $\int_\gamma \omega(\gamma, \gamma')\ \mathrm{d}t$. In fact, we consider a more general variational problem: describe the curves γ that are the relative extrema of $\int_\gamma \omega(\gamma', \gamma'')^a\ \mathrm{d}t$ constrained by $\int_\gamma \omega(\gamma, \gamma')\ \mathrm{d}t$ where a is a real exponent. We point out that the case $a = \frac{1}{3}$ corresponds to the affine isoperimetric inequality.

Before we come to our main results, we point out that the corresponding metric problem, extremizing the L^2-norm of the (metric) curvature on a class of plane curves, is a widely studied topic going back to Bernoulli and Euler and goes under the name of elastica, see [12]. In addition, we mention the recent article [13] in which the L^p-norms of the curvature are studied and [7] where the corresponding gradient flows are developed. The latter is a natural next step also for the affine context from this article. See [11] for an affine analog of the Euclidean curve shortening flow.

Main results We prove that such extremal curves of this variational problem lie in symplectic affine 2-planes, and therefore, the problem reduces to a 2-dimensional one (Proposition 3.4).

We then fix the constraint by giving the curve the centroaffine parameterization, i.e., we assume $[\gamma(t), \gamma'(t)] = 1$. Then, Hill's equation $\gamma''(t) = -p(t)\gamma(t)$ holds, and we may consider the functional $\mathcal{B}_a(\gamma) := \int [\gamma'(t), \gamma''(t)]^a\ \mathrm{d}t$.

For $a \in [\frac{1}{2}, 1]$, one has rigidity: the extremal curves are multiply traversed conics (Propositions 3.7 and 3.8). The same rigidity result holds, although for a different reason, in the case of the affine isoperimetric inequality, $a = \frac{1}{3}$, (Theorem 3).

In Theorem 1, we describe non-trivial infinitesimal deformations of multiply traversed conics in the class of extremal curves: if $a = \frac{1}{3}$, then the n-fold ellipse is infinitesimally rigid; otherwise, a non-trivial infinitesimal deformation of the n-fold ellipse exists if and only if

$$a = \frac{k^2 - 2n^2}{k^2 - 4n^2}$$

for some positive integer $k \neq n$.

Theorem 2 concerns the second-order deformations of conics: for $a < 0$, the circle γ_0 is a local minimum of $\mathcal{B}_a$; for $a \in (0, \frac{1}{3})$, it is a local maximum; for $a > \frac{7}{5}$, it is a local minimum; and in other cases,the Hessian is not sign-definite. The Hessian is degenerate (with 1-dimensional kernel) if and only if

$$a = \frac{k^2 - 2}{k^2 - 4}$$

for some positive integer k.

In Sect. 7, we present examples of extremal curves and formulate some conjectures about them.

This introduction would not be complete if we failed to mention another reason for our interest in centroaffine differential geometry, namely its close relation with the Korteweg–de Vries equation, discovered by Pinkall [9] and studied by a number of authors since then. When the exponent a equals 2, the extremal curves are periodic solutions to Lamé's equation thoroughly studied in this context in a recent paper [4].

2. Examples of symplectically convex and symplectically star-shaped curves

In this section, we construct curves γ with all possible sign combinations of the quantities $\omega(\gamma, \gamma') \neq 0$ and $\omega(\gamma', \gamma'') \neq 0$. In particular, we assume that all curves are immersed. We start with a remark concerning symplectically star-shaped curves. The sphere $S^{2n-1} \subset \mathbb{R}^{2n}$ carries a contact structure defined by the symplectic orthogonal complement to the position vector. A symplectically star-shaped curve projects to a transverse curve in S^{2n-1}. A similar remark applies to the contact $\mathbb{RP}^{2n-1}$, the projectivization of $\mathbb{R}^{2n}$. If $\omega(\gamma, \gamma') > 0$, then γ is positively transverse and $<$ corresponds to negatively transverse. A somewhat similar interpretation for the condition $\omega(\gamma', \gamma'') \neq 0$ is derived in Lemma 2.2 in case of $\mathbb{R}^4$.

We consider the unit sphere $S^3 \subset \mathbb{C}^2 = \mathbb{R}^4$ with its standard contact structure. The standard contact form at a point $q \in S^3$ is $\omega(q, \cdot)|_{T_q S^3}$. Let $\gamma(t)$ be a smooth closed Legendrian curve in S^3, i.e., $\omega(\gamma, \gamma') = 0$. Then,

$J\gamma'$ is a vector normal to γ inside the contact plane. Here, J is the complex structure on $\mathbb{C}^2$. The following lemma is well known; see, e.g., [3].

Lemma 2.1. *Pushing a closed Legendrian curve γ slightly in the direction of $J\gamma'$, resp. $-J\gamma'$, inside S^3 yields a negative, resp. positive, transverse curve. Thus, a Legendrian curve in $S^3 \subset \mathbb{C}^2$ can be deformed into a symplectically star-shaped curve.*

Proof. Let us assume that γ is parametrized by arc-length and set $\Gamma = \gamma + \varepsilon J\gamma'$. Then

$$\omega(\Gamma, \Gamma') = \varepsilon\big[\omega(\gamma, J\gamma'') + \omega(J\gamma', \gamma')\big] + O(\varepsilon^2).$$

If we denote by $\cdot$ the inner product, then $\omega(J\gamma', \gamma') = -\gamma' \cdot \gamma' = -1$. From $\gamma \cdot \gamma = 1$, we conclude $\gamma \cdot \gamma' = 0$, i.e., $\omega(\gamma, J\gamma') = 0$. Differentiating this equality gives $\omega(\gamma, J\gamma'') + \omega(\gamma', J\gamma') = 0$, and hence, $\omega(\gamma, J\gamma'') = \omega(J\gamma', \gamma')$. It follows that for sufficiently small $\varepsilon > 0$, one has $\omega(\Gamma, \Gamma') < 0$. That is, Γ is a negative transverse curve. The case of $\Gamma = \gamma - \varepsilon J\gamma'$ is similar. □

Let us say that a Legendrian curve γ in S^3 has an *inflection point* at $p \in \gamma$ if it is second-order tangent to its tangent great circle at p. A generic curve in S^3 does not have inflection points, but, as we shall see, a generic Legendrian curve has finitely many of them.

The tangent Gauss map sends a Legendrian curve in S^3 to the space of oriented Legendrian great circles, that is, to the oriented Lagrangian Grassmannian Λ_2^+. The image of the Gauss map is a smooth curve which has vanishing differential precisely at the points corresponding to inflection points of the Legendrian curve.

Lemma 2.2. *A necessary and sufficient condition for a Legendrian curve γ having an inflection point at $\gamma(t)$ is $\omega(\gamma'(t), \gamma''(t)) = 0$.*

Proof. The condition $\omega(\gamma', \gamma'') = 0$ is invariant under reparameterization of γ, so we may assume that γ is parameterized by arc length.

First, we claim that the orthogonal projection of γ'' to S^3 is $\gamma + \gamma''$. Indeed, $\gamma \cdot \gamma' = 0$ implies that $\gamma' \cdot \gamma' + \gamma \cdot \gamma'' = 0$; hence, $\gamma \cdot \gamma'' = -1$. Now, $\gamma \cdot (\gamma + \gamma'') = 1 - 1 = 0$, as claimed. Therefore, an inflection point is characterized by the vanishing of $\gamma + \gamma''$.

Next, we claim that the tangential acceleration vector $\gamma + \gamma''$ lies in the contact structure. Indeed, $\omega(\gamma, \gamma') = 0$; hence, after differentiating, $\omega(\gamma, \gamma'') = 0$. Therefore, $\omega(\gamma, \gamma + \gamma'') = 0$, as needed.

In addition, $\gamma \cdot \gamma = \gamma' \cdot \gamma' = 1$ implies that γ' and $\gamma + \gamma''$ are orthogonal to each other. Finally, since γ is Legendrian, i.e., $\omega(\gamma, \gamma') = 0$, the tangent vector γ' lies also in the contact structure. To summarize, we have two orthogonal vectors, γ' and $\gamma + \gamma''$, in the contact structure. In particular, $\omega(\gamma', \gamma + \gamma'') = \pm\|\gamma'\|\|\gamma + \gamma'\| = \pm\|\gamma + \gamma'\|$ since $\|\gamma'\| = 1$.

The Lemma now follows from $\omega(\gamma', \gamma'') = \omega(\gamma', \gamma + \gamma'') = 0$ if and only if $\gamma + \gamma'' = 0$, i.e., if and only if γ has an inflection point. □

It follows that if we find a closed Legendrian curve γ with $\omega(\gamma', \gamma'') \neq 0$, then, according to Lemma 2.1, its small push in the normal direction in the contact plane will yield a curve with $\omega(\gamma,\gamma') \neq 0$ and $\omega(\gamma',\gamma'') \neq 0$. Let us show how to construct such a Legendrian curve γ. See [2,8] for related matters.

Consider the Reeb field of the standard contact form on S^3 and let $pr : S^3 \to S^2$ be the respective Hopf fibration. The projection pr takes the Legendrian great circles in S^3 to the great circles in S^2. Closed Legendrian curves are projected to closed immersed curves in S^2 that bound a region with signed area that is a multiple of 2π and, conversely, such a spherical curve lifts (non-uniquely) to a closed Legendrian curve in S^3.

Since pr is a Riemannian submersion, it follows that the inflections of a Legendrian curve γ correspond to the inflections of the spherical curve $pr(\gamma)$. Accordingly, every closed spherical curve with everywhere positive geodesic curvature and bounding area $2\pi k$ for some $k \in \mathbb{Z}$ lifts to a Legendrian curve which is free from inflections. Starting with a closed spherical curve with everywhere positive geodesic curvature, the area condition is easily arranged by adding appropriately sized kinks to the curve.

Finally, we note that the map $(z_1, z_2) \mapsto (\bar{z}_1, \bar{z}_2)$ of $\mathbb{C}^2$ preserves the contact structure and the property of γ being Legendrian, but changes the sign of $\omega(\gamma'(t), \gamma''(t))$ to the opposite. Therefore, we can have all four combinations of signs of the quantities $\omega(\Gamma,\Gamma')$ and $\omega(\Gamma',\Gamma'')$ where Γ is a small push-off as in Lemma 2.1.

Remark 2.3. Another interpretation of the above construction is by looking at a different Hopf fibration $\widetilde{pr} : S^3 \to S^2$, whose fibers are Legendrian. This Hopf fibration takes Legendrian great circles to circles (of some, possibly zero radius) on S^2. The projection of a smooth Legendrian curve γ is a wave front, possibly with cusps, and the inflections of γ correspond to the vertices of the spherical curve $\widetilde{pr}(\gamma)$. If $\widetilde{pr}(\gamma)$ is smooth and convex, it must have at least four vertices (the 4-vertex theorem), without convexity at least two vertices, but if $\widetilde{pr}(\gamma)$ has cusps, it may be vertex-free.

3. Toward the solution of the variational problem

In this section, we describe the setting of our variational problem and prove that extremizers are contained in symplectic affine 2-planes. This allows us to reduce the problem to the 2-dimensional case.

The problem is to find the closed symplectically star-shaped and symplectically convex curves that are extrema of $\mathcal{B}_a(\gamma) := \int \omega(\gamma',\gamma'')^a \; \mathrm{d}t$, $a \in \mathbb{R} \setminus \{0\}$, subject to the constraint given by $\mathcal{A}(\gamma) := \int \omega(\gamma,\gamma') \; \mathrm{d}t$. More precisely, we consider the space $\mathcal{P}_T$ of T-periodic symplectically star-shaped and symplectically convex curves in $\mathbb{R}^{2n}$ and consider $\mathcal{A}, \mathcal{B}_a : \mathcal{P}_T \to \mathbb{R}$ and ask for extrema of $\mathcal{B}_a$ subject to $\mathcal{A} = c_0$, that is, we want to describe the set

$$\mathrm{Crit}(\mathcal{B}_a|_{\{\mathcal{A}=c_0\}}) \subset \mathcal{P}_T.$$

Let us note that both functionals, $\mathcal{A}$ and $\mathcal{B}_a$, are translation invariant, that is, do not depend on the choice of the origin (as long as the curve remains star-shaped). We start with a few simple observations.

Lemma 3.1. *For $c_0 > 0$ and $c_1 > 0$, there is a natural bijection (by rescaling) from $\mathcal{P}_T$ to itself inducing a bijections*

$$\{\mathcal{A} = c_0\} \cong \{\mathcal{A} = c_1\}$$

and

$$\mathrm{Crit}(\mathcal{B}_a|_{\{\mathcal{A}=c_0\}}) \cong \mathrm{Crit}(\mathcal{B}_a|_{\{\mathcal{A}=c_1\}}).$$

Proof. Let $\gamma_0 \in \mathcal{P}_T$ and consider

$$\gamma_1(t) := \left(\tfrac{c_1}{c_0}\right)^{\frac{1}{2}} \gamma_0(t) \in \mathcal{P}_T.$$

If $\gamma_0 \in \{\mathcal{A} = c_0\}$, then

$$\mathcal{A}(\gamma_1) = \tfrac{c_1}{c_0}\mathcal{A}(\gamma_0) = c_1,$$

i.e., $\gamma_1 \in \{\mathcal{A} = c_1\}$. From

$$\mathcal{B}_a\left(\left(\tfrac{c_1}{c_0}\right)^{\frac{1}{2}} \gamma\right) = \left(\tfrac{c_1}{c_0}\right)^{a} \mathcal{B}_a(\gamma), \tag{1}$$

it follows that the bijection $\gamma \mapsto \left(\tfrac{c_1}{c_0}\right)^{\frac{1}{2}} \gamma$ just rescales $\mathcal{B}_a$ by a fixed factor, i.e., induces the claimed bijection $\mathrm{Crit}(\mathcal{B}_a|_{\{\mathcal{A}=c_0\}}) \cong \mathrm{Crit}(\mathcal{B}_a|_{\{\mathcal{A}=c_1\}})$. □

Remark 3.2. Equation (1) implies that, if $\mathcal{B}_a$ has critical points at all, they appear in $\mathbb{R}_{>0}$-family and the critical value is necessarily 0.

Lemma 3.3. *The bijection $\mathcal{P}_T \to \mathcal{P}_1$ given by $\Gamma(t) := \gamma(t/T)$ preserves $\mathcal{A}$ and rescales $\mathcal{B}_a$ by $\frac{1}{T^{3a-1}}$.*

Proof. We consider the bijection $\mathcal{P}_T \to \mathcal{P}_1$ given by $\Gamma(t) := \gamma(t/T)$. Then

$$\begin{aligned}\mathcal{A}(\Gamma) &= \int_0^1 \omega(\Gamma(t), \Gamma'(t))\ \mathrm{d}t = \int_0^1 \omega\big(\gamma(\tfrac{t}{T}), \tfrac{1}{T}\gamma'(\tfrac{t}{T})\big)\ \mathrm{d}t \\ &= \int_0^T \omega\big(\gamma(s), \gamma'(s)\big)\ \mathrm{d}s = \mathcal{A}(\gamma),\end{aligned}$$

as it has to be since $\mathcal{A}(\Gamma) = \mathcal{A}(\gamma)$ is twice the enclosed area. Similarly, we compute

$$\begin{aligned}\mathcal{B}_a(\Gamma) &= \int_0^1 \omega(\Gamma'(t), \Gamma''(t))^a\ \mathrm{d}t \\ &= \int_0^1 \tfrac{1}{T^{3a}}\omega\big(\gamma'(\tfrac{t}{T}), \gamma''(\tfrac{t}{T})\big)^a\ \mathrm{d}t = \tfrac{1}{T^{3a-1}} \int_0^T \omega\big(\gamma'(s), \gamma''(s)\big)^a\ \mathrm{d}s \\ &= \tfrac{1}{T^{3a-1}}\mathcal{B}_a(\gamma),\end{aligned}$$

as claimed. □

Now, we begin to study the relative extrema of $\mathcal{B}_a$ constrained by $\mathcal{A}$. We call the extrema *critical curves*. The previous two lemmata imply that we may consider curves of fixed period and with fixed constraint.

Proposition 3.4. *A critical curve is contained inside a symplectic affine 2-plane of $(\mathbb{R}^{2n}, \omega)$.*

Proof. We first derive the equation for a critical curve using a Lagrange multiplier $\lambda \in \mathbb{R}$. For that, let $v(t)$ be a vector field along γ and consider an infinitesimal deformation $\gamma_\varepsilon = \gamma + \varepsilon v$ of γ. Then, γ is critical if for some λ for all v

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\bigg|_{\epsilon=0} \left[\lambda \int \omega(\gamma_\varepsilon, \gamma_\varepsilon') \,\mathrm{d}t + \int \omega(\gamma_\varepsilon', \gamma_\varepsilon'')^a \,\mathrm{d}t\right] = 0.$$

The derivative of the first integral in direction of v computes to

$$\int (\omega(v, \gamma') + \omega(\gamma, v')) \,\mathrm{d}t = 2 \int \omega(v, \gamma') \,\mathrm{d}t,$$

where we used integration by parts and skew symmetry of the symplectic form in the equality sign.

Similarly, setting $F := \omega(\gamma', \gamma'')^{a-1}$, the derivative of the second integral can be expressed as

$$\begin{aligned} a \int F\big(\omega(v', \gamma'') + \omega(\gamma', v'')\big) \,\mathrm{d}t &= a \int \big(F' \omega(v', \gamma') + 2F\omega(v', \gamma'')\big) \,\mathrm{d}t \\ &= a \int \big(F'' \omega(\gamma', v) + 3F' \omega(\gamma'', v) + 2F\omega(\gamma''', v)\big) \,\mathrm{d}t. \end{aligned}$$

Since v is arbitrary and ω is non-degenerate, the criticality condition implies that the vectors $\gamma', \gamma'', \gamma'''$ are linearly dependent, and since $F \neq 0$, one has $\gamma''' = f\gamma' + g\gamma''$ for some periodic functions $f(t), g(t)$.

It follows that the bivector $\gamma' \wedge \gamma''$ satisfies the differential equation $(\gamma' \wedge \gamma'')' = g(\gamma' \wedge \gamma'')$, that is, remains proportional to itself. Hence, the curve γ' is planar. It follows by integration that the curve γ lies in an affine 2-plane, parallel to the plane spanned by γ' and γ''. □

Proposition 3.4 allows us to reduce the problem to two dimensions as follows. A critical curve γ is contained in an affine 2-plane. Let V be the two-dimensional linear space parallel to this affine 2-plane, and W be its symplectic orthogonal. Then, we may write $\gamma = \gamma_1 \oplus \gamma_2 \in V \oplus W$. Since $\gamma' \in V$ we conclude that γ_2 is constant. We claim that γ_1 is also a critical curve. Indeed, since W is symplectically orthogonal to V, we compute

$$0 \neq \omega(\gamma, \gamma') = \omega(\gamma_1 \oplus \gamma_2, \gamma_1') = \omega(\gamma_1, \gamma_1'),$$

i.e., γ is symplectically star-shaped and $\mathcal{A}(\gamma) = \mathcal{A}(\gamma_1)$. Trivially, γ_1 is also symplectically convex with $\mathcal{B}_a(\gamma) = \mathcal{B}_a(\gamma_1)$. Now, that we have reduced the problem to two dimensions, we denote the symplectic form ω in the plane by brackets $[\cdot, \cdot]$.

Convention 3.5. *To fix the constraint $\mathcal{A}$, from now on, we always parametrize a star-shaped curve, so that $[\gamma, \gamma'] = 1$, that is, γ is in its centroaffine parameterization.*

In centroaffine parameterization, we have $\gamma'' = -p\gamma$. The function p is called the *centroaffine curvature* of the curve γ. For some computations below, we record that $p = [\gamma', \gamma''] > 0$ and $\gamma''' = -p'\gamma - p\gamma'$.

Thus, we can reformulate our goal as describing all extremal curves of the functional $\mathcal{B}_a(\gamma) := \int p^a(t)\mathrm{d}t$ on the space of periodic curves in centroaffine parameterization with $p > 0$. Here, $a \in \mathbb{R} \setminus \{0\}$ is fixed. To be quite precise, we point out that we, of course, consider the space of all periodic curves in centroaffine parameterization of a *fixed period*, mostly, 2π or a multiple of 2π. As explained in the introduction, the case $a = \frac{1}{3}$ is of special interest, since this corresponds to the affine isoperimetric inequality.

Lemma 3.6. *The properties of a curve being symplectically star-shaped and convex are invariant under point-wise multiplication by an element in $SL(2, \mathbb{R})$ and sufficiently small translation. For any value of a, the functional $\mathcal{B}_a$ is invariant under this $SL(2, \mathbb{R})$ action. In addition, if $a = \frac{1}{3}$, the functional $\mathcal{B}_{\frac{1}{3}}$ is invariant under translations. This is the only value of a with this additional symmetry.*

Proof. Point-wise multiplication of a curve with a matrix from $SL(2, \mathbb{R})$ does not change its centroaffine parameterization nor its centroaffine curvature.

A sufficiently small translation of a curve γ keeps the properties of being symplectically star-shaped and convex but in general not the centroaffine parameterization of γ. After reparametrization, the shifted curve $\bar{\gamma} = \gamma + c$ is in centroaffine parameterization, i.e., $\bar{\gamma}(\tau) = \gamma(t(\tau)) + c$ and $[\bar{\gamma}, \frac{d\bar{\gamma}}{\mathrm{d}\tau}] = 1$. Its centroaffine curvature $\bar{p}$ satisfies

$$\bar{p}(\tau) = \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^3 p(\tau);$$

indeed

$$\bar{p} = \left[\frac{\mathrm{d}\bar{\gamma}}{\mathrm{d}\tau}, \frac{\mathrm{d}^2\bar{\gamma}}{\mathrm{d}\tau^2}\right] = \left[\frac{\mathrm{d}t}{\mathrm{d}\tau}\frac{\mathrm{d}\bar{\gamma}}{\mathrm{d}t}, \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^2 \frac{\mathrm{d}^2\bar{\gamma}}{\mathrm{d}t^2} + \frac{\mathrm{d}^2 t}{\mathrm{d}\tau^2}\frac{d\bar{\gamma}}{\mathrm{d}t}\right] = \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^3 \left[\frac{\mathrm{d}\bar{\gamma}}{\mathrm{d}t}, \frac{\mathrm{d}^2\bar{\gamma}}{\mathrm{d}t^2}\right]$$
$$= \left(\frac{\mathrm{d}t}{\mathrm{d}\tau}\right)^3 p,$$

since $\frac{\mathrm{d}\bar{\gamma}}{\mathrm{d}t} = \frac{d\gamma}{\mathrm{d}t} = \gamma'$. We conclude that precisely for $a = \frac{1}{3}$ the functional $\mathcal{B}_a$ satisfies

$$\mathcal{B}_a(\bar{\gamma}) = \int \bar{p}^a(\tau)\mathrm{d}\tau = \int p^a(t)\mathrm{d}t = \mathcal{B}_a(\gamma)$$

for all symplectically star-shaped and convex curves γ. □

We now derive a critical point equation for $\mathcal{B}_a$ in terms of the function $F = p^{a-1}$.

Proposition 3.7. *For $a \neq 1$, the extremal curves γ of the functional $\int p^a(t)\mathrm{d}t$ satisfy*

$$F''' = -2(2+b)F^b F', \tag{2}$$

where $F = p^{a-1}$ and $b = \frac{1}{a-1}$. If $a = 1$ or $a = \frac{1}{2}$, then p is constant and γ is therefore a conic.

Proof. First, we describe the vector fields v along γ that preserve the centroaffine parameterization of γ. Write such a field as $v = g\gamma + f\gamma'$. Since the deformation by v is assumed to preserve the centroaffine parameterization, we conclude that $[\gamma, v'] + [v, \gamma'] = 0$ and hence $2g + f' = 0$. Thus, the vector field v has the form $-\frac{f'}{2}\gamma + f\gamma'$; the deformations keeping the centroaffine parameterization depend on one periodic function.

As in the proof of Proposition 3.4, linearizing $\int p^a(t)\mathrm{d}t$ in v leads to

$$\int F\big([\gamma', v''] + [v', \gamma'']\big)\ \mathrm{d}t = 0$$

for every vector field v as above, where we recall that $F = p^{a-1}$. Using integration by parts twice and recalling that $\gamma''' = -p'\gamma - p\gamma'$, we rewrite the integral as

$$\int \Big(\big(F'' - 2pF\big)[\gamma', v] - \big(3pF' + 2p'F\big)[\gamma, v]\Big)\ \mathrm{d}t.$$

Using $[\gamma, v] = f$, $[\gamma', v] = \frac{1}{2}f'$ and another integration by parts, this integral becomes

$$\int \Big(-\tfrac{1}{2}F''' - 2pF' - p'F\Big)f\ \mathrm{d}t.$$

The critically condition is that this integral vanishes for all f, and we conclude that the integrand is zero.

If $a = 1$, then $F \equiv 1$, and hence, the criticality condition simply becomes $p' = 0$. Therefore, γ solves $\gamma'' = \text{const} \cdot \gamma$, i.e., γ is a conic.

Otherwise, recall that $p = F^b$; hence, $p' = bF^{b-1}F'$. Substitute this to the integrand and collect terms to obtain the claimed formula (2).

If $a = \frac{1}{2}$, then $b = -2$, and Eq. (2) reduces to $F''' = 0$. Since F is periodic and positive, F is necessarily a constant, and so is $p = F^{-2}$. Thus, again, γ is a conic in this case. □

Proposition 3.8. *For $a \in (\frac{1}{2}, 1)$, Eq. (2) has only constant solutions.*

Proof. We can integrate Eq. (2) to

$$F'' = -\frac{2(b+2)}{b+1}F^{b+1} + c \tag{3}$$

with some constant c. Note that $b \neq -1$. Since F is a positive and periodic function, we get at the minimum $m := \min F$ and the maximum $M := \max F$ of F the usual inequalities for F'', leading to

$$0 \geq -\frac{2(b+2)}{b+1}M^{b+1} + c \qquad \text{and} \qquad 0 \leq -\frac{2(b+2)}{b+1}m^{b+1} + c.$$

Thus, we arrive at the two inequalities

$$\begin{cases} \dfrac{b+2}{b+1}m^{b+1} \leq \dfrac{b+2}{b+1}M^{b+1} \\ m \leq M. \end{cases} \tag{4}$$

If $b+1 > 0$ or $b+2 > 0 > b+1$, then the first inequality in (4) is consistent with the second. However, the first inequality together with $0 > b+2$ implies

$m \geq M$ and we conclude $m = M$. That is, $0 > b+2$ implies that F is constant. Now, recall that $b = \frac{1}{a-1}$. Therefore, F is constant if $\frac{1}{2} < a < 1$, as claimed. $\square$

Lemma 3.9. *For $a \notin [\frac{1}{2}, 1]$, Eq. (2) has non-constant positive, periodic solutions.*

Remark 3.10. The solutions we construct in the proof of the lemma below are, by construction, often times multiply covered.

Proof. As a preparation, we consider the Hamiltonian formulation of the ODE (3). Hamilton's equations for the Hamiltonian function ($c \in \mathbb{R}$ is a constant which we will choose appropriately below)

$$H(Q,P) := \frac{1}{2}P^2 + \frac{2}{b+1}Q^{b+2} - cQ : \mathbb{R}^2 \longrightarrow \mathbb{R}$$

are given by

$$\begin{cases} \dot{P} = -\dfrac{\partial H}{\partial Q} = -\dfrac{2(b+2)}{b+1}Q^{b+1} + c \\ \dot{Q} = \dfrac{\partial H}{\partial P} = P, \end{cases}$$

and are, with $Q = F$ and $P = F'$, equivalent to the ODE (3). We compute

$$dH(Q,P) = P\,dQ + \left(\frac{2(b+2)}{b+1}Q^{b+1} - c\right)dQ$$

and

$$\mathrm{Hess}H(Q,P) = \begin{pmatrix} 2(b+2)Q^b & 0 \\ 0 & 1 \end{pmatrix}.$$

Therefore, (Q,P) is a critical point of H if and only if

$$\begin{cases} P = 0 \\ \dfrac{2(b+2)}{b+1}Q^{b+1} = c, \end{cases}$$

and then is a local minimum if

$$(b+2)Q^b > 0.$$

Our assumption $a \notin (\frac{1}{2}, 1)$ is equivalent to $b \in (-2,-1) \cup (-1,0) \cup (0,\infty)$ (since $b = \frac{1}{a-1}$ and $a \neq 0$).

We are searching for non-constant positive, periodic solutions of equation (2); i.e., we are looking for non-constant periodic orbits of H with $Q = F > 0$.

In case $b \in (0,\infty)$, we choose $c \in \mathbb{R}$ very large and positive. Then, the equation $\frac{2(b+2)}{b+1}Q^{b+1} = c$ determines a local minimum of F at, say, $(Q_0, P_0 = 0)$ with Q_0 large and positive. Therefore, the linearized dynamics given by

$$\mathrm{Hess}H(Q_0,0) = \begin{pmatrix} 2(b+2)Q_0^b & 0 \\ 0 & 1 \end{pmatrix}$$

is a very fast rotation of $\mathbb{R}^2$. That is, the linearized dynamics is periodic with very small periodic. Since $(Q_0, 0)$ is a local minimum, any level set $\{H = E\}$ with E slightly larger than the energy $H(Q_0, 0) = \frac{2}{b+1}Q_0^{b+2} - cQ_0$ of $(Q_0, 0)$ consists of a small circle (and potentially other connected components). This small circle is then a periodic orbit with period approximately that of the linearized dynamics. Since $b \neq 0$, the Hamiltonian H is not quadratic and therefore varying the energy value E also changes the period of the periodic orbit in the level set $\{H = E\}$ near $(Q_0, 0)$; i.e., by iterating and varying the level set, we can arrange any period we wish.

In case $b \in (-1, 0)$, we choose $c \in \mathbb{R}$ positive and small. Then, the equation $\frac{2(b+2)}{b+1}Q^{b+1} = c$ determines a local minimum of F at $(Q_0, P_0 = 0)$ with Q_0 small and positive. The linearized dynamics is again given by

$$\text{Hess}H(Q_0, 0) = \begin{pmatrix} 2(b+2)Q_0^b & 0 \\ 0 & 1 \end{pmatrix}.$$

This is still a very fast rotation of $\mathbb{R}^2$, since $b < 0$, $b + 2 > 0$ and $Q_0 > 0$ is small. The argument proceeds as in the previous case.

In case $b \in (-2, -1)$, we choose $c \in \mathbb{R}^2$ negative and large in absolute value. The equation $\frac{2(b+2)}{b+1}Q^{b+1} = c$ determines a local minimum of F at $(Q_0, P_0 = 0)$ with Q_0 small and positive. The linearized dynamics is yet again a very fast rotation of $\mathbb{R}^2$, since $b < 0$, $b + 2 > 0$, and $Q_0 > 0$ is small. The argument proceeds as in the first case. $\square$

Remark 3.11. Equation (2) can be further integrated, which is in terms of the Hamiltonian formulation just expressing the preservation of H along solutions

$$(F')^2 = -\frac{4}{b+1}F^{b+2} + 2cF + d. \tag{5}$$

For $a = 2$, resp. $a = \frac{3}{2}$, we have $b + 2 = 3$, resp. $b + 2 = 4$, and hence F is the Weierstrass elliptic function. Since $p = F^b$, in the first case, p is also the Weierstrass function, and in the second case, p is its square. In the former case, the respective equation $\gamma'' = -p\gamma$ is called the Lamé equation, and it was thoroughly studied, see, e.g., [14].

4. Infinitesimal rigidity of multiple conics as critical curves

As a multiple conic, we take the unit circle traversed n times. This is a critical curve of the functional $\mathcal{B}_a(\gamma) = \int_0^{2\pi n} p^a \mathrm{d}t$, and we ask whether it admits a non-trivial infinitesimal deformation in the class of critical curves.

The functional $\mathcal{B}_a$ is invariant under the action of $SL(2, \mathbb{R})$; the respective deformations comprise a 3-dimensional space, and we consider them as trivial. In addition, if $a = \frac{1}{3}$, the functional is invariant under parallel translations; see Lemma 3.6. In this case, we add this 2-dimensional space to the deformations that we consider as trivial. The n-fold circle is infinitesimally rigid if it does not admit non-trivial infinitesimal deformations in the class of critical curves.

Let $\gamma_0(t) = (\cos t, \sin t)$ be the unit circle traversed n times, and let $\gamma_1 = -\frac{f'}{2}\gamma_0 + f\gamma_0'$ be a vector field along it that defines its infinitesimal deformation. We assume that the period is $2\pi n$, and accordingly, $f(t)$ is a $2\pi n$-periodic function.

The trivial deformations are described in the following lemma.

Lemma 4.1. *The infinitesimal action of* $SL(2,\mathbb{R})$ *corresponds to the functions* f *that are linear combinations of* $1, \cos(2t), \sin(2t)$*, and parallel translations to the functions* f *that are linear combinations of* $\cos t, \sin t$*.*

Proof. The deformed curve is $\gamma_0 + \varepsilon\gamma_1$. The case $f = 1$ corresponds to the rotation of the unit circle.

Let $f(t) = \sin(2t)$. In this case, computing up to ε^2, the curve $\gamma_0 + \varepsilon\gamma_1$ is the ellipse $(1+2\varepsilon)x^2 + (1-2\varepsilon)y^2 = 1$, and likewise for $f = \cos(2t)$.

If $f(t) = 2\sin t$, then, again mod ε^2, the curve $\gamma_0 + \varepsilon\gamma_1$ is the unit circle $(x+\varepsilon)^2 + y^2 = 1$, and likewise for $f = 2\cos t$. □

We are ready to describe the infinitesimal rigidity of the circle.

Theorem 1. *If* $a = \frac{1}{3}$*, then the* n*-fold unit circle is infinitesimally rigid. Otherwise, a non-trivial infinitesimal deformation of the* n*-fold unit circle exists if and only if*

$$a = \frac{k^2 - 2n^2}{k^2 - 4n^2}$$

for some positive integer $k \neq n$*.*

Proof. The calculations below are modulo ε^2.

Let $\Gamma = \gamma_0 + \varepsilon\gamma_1$. We have $\Gamma'' = -(p_0 + \varepsilon p_1)\Gamma$, and hence

$$p_0 + \varepsilon p_1 = [\Gamma', \Gamma''] = [\gamma_0' + \varepsilon\gamma_1', \gamma_0'' + \varepsilon\gamma_1''] = 1 + \varepsilon([\gamma_0', \gamma_1''] + [\gamma_1', \gamma_0'']).$$

Therefore

$$F = (p_0 + \varepsilon p_1)^{a-1} = 1 + \varepsilon(a-1)([\gamma_0', \gamma_1''] + [\gamma_1', \gamma_0'']).$$

We calculate

$$\gamma_1' = -\left(f + \frac{1}{2}f''\right)\gamma_0 + \frac{1}{2}f'\gamma_0', \;\; \gamma_1'' = -\left(\frac{3}{2}f' + \frac{1}{2}f'''\right)\gamma_0 - f\gamma_0';$$

hence

$$q := [\gamma_0', \gamma_1''] + [\gamma_1', \gamma_0''] = 2f' + \frac{1}{2}f'''.$$

Compare with [9], where the Korteweg-de Vries equation is interpreted as a flow on centroaffine curves.

The case of $a = 1$ was considered earlier (the only solution of the variational problem is a conic); therefore, we assume that $a \neq 1$.

Since the perturbed curve is critical, Eq. (3) holds. Write the constant in this equation as $c_0 + \varepsilon c_1$. Then, (3) implies

$$q'' = -2(b+2)q + \frac{c_1}{a-1}.$$

Since q'' and q have zero average, we conclude that $c_1 = 0$. Therefore, $q'' = -2(b+2)q$.

This equation has periodic solutions only when $b+2 > 0$, and then, $q(t)$ is a linear combination of $\cos(\sqrt{2(b+2)}t)$ and $\sin(\sqrt{2(b+2)}t)$. For q to be $2\pi n$-periodic, one has to have $\sqrt{2(b+2)} = \frac{k}{n}$, that is

$$b = \frac{k^2 - 4n^2}{2n^2} \quad \text{or} \quad a = \frac{k^2 - 2n^2}{k^2 - 4n^2} \tag{6}$$

for positive integers k, n. Note that since $b \neq 0$, one has $k \neq 2n$.

Let us show that if condition (6) holds, the desired infinitesimal deformations of a multiple circle exist.

We find $f(t)$ from the equation

$$2f' + \frac{1}{2}f''' = A\cos\left(\frac{kt}{n}\right) + B\sin\left(\frac{kt}{n}\right).$$

It follows that, modulo the kernel of the differential operator $\frac{1}{2}d^3 + 2d$, the function f is a linear combination of $\cos\left(\frac{kt}{n}\right)$ and $\sin\left(\frac{kt}{n}\right)$. The kernel of the differential operator contributes trivial deformations, and we end up with a 2-dimensional space of deformations.

It remains to see when these deformations are trivial. Since $k \neq 0$ and $k \neq 2n$, the only "suspicious" case is $k = n$. In this case, $a = \frac{1}{3}$, and indeed, the first harmonics give trivial deformations, corresponding to parallel translations. □

Remark 4.2. We note that in agreement with Lemma 3.8

$$a = \frac{k^2 - 2n^2}{k^2 - 4n^2}$$

does not take values in $[\frac{1}{2}, 1]$: if $k > 2n$, then $a > 1$, and if $k < 2n$, then $a < \frac{1}{2}$.

5. Second-order deformations of conics

Here, we investigate how the functional $\mathcal{B}_a(\gamma) = \int p(t)^a \mathrm{d}t$ changes under a second-order deformation of the unit circle. We recall that the functional is $SL(2,\mathbb{R})$-invariant; see Lemma 4.1.

We assume that the period is 2π and that the curves have the centroaffine parameterization $[\gamma, \gamma'] = 1$. Let $\Gamma = \gamma_0 + \varepsilon\gamma_1 + \varepsilon^2\gamma_2$, ignoring the higher order terms in ε. As before, $\gamma_0(t) = (\cos t, \sin t)$, and hence, $\gamma_0'' = -\gamma_0$, and $\gamma_1 = -\frac{f'}{2}\gamma_0 + f\gamma_0'$, where $f(t)$ is a 2π-periodic function.

The condition $[\Gamma, \Gamma'] = 1$ implies

$$[\gamma_0, \gamma_2'] + [\gamma_1, \gamma_1'] + [\gamma_2, \gamma_0'] = 0. \tag{7}$$

We have $\Gamma'' = -P\Gamma$, and hence

$$P = [\Gamma', \Gamma''] = [\gamma_0' + \varepsilon\gamma_1' + \varepsilon^2\gamma_2', \gamma_0'' + \varepsilon\gamma_1'' + \varepsilon^2\gamma_2''] = 1 + \varepsilon q(t) + \varepsilon^2 r(t),$$

where

$$q = [\gamma_0', \gamma_1''] + [\gamma_1', \gamma_0''],\ \ r = [\gamma_0', \gamma_2''] + [\gamma_1', \gamma_1''] + [\gamma_2', \gamma_0''].$$

Then

$$P^a = 1 + \varepsilon a q + \varepsilon^2 a \left(r + \frac{a-1}{2} q^2 \right).$$

As we already know

$$q = [\gamma_0', \gamma_1''] + [\gamma_1', \gamma_0''] = 2f' + \frac{1}{2} f'''.$$

We need to calculate

$$\int_0^{2\pi} \left(r + \frac{a-1}{2} q^2 \right) \mathrm{d}t.$$

Integrating by parts and using $\gamma_0''' = -\gamma_0'$, we get

$$\int r \mathrm{d}t = \int ([\gamma_0', \gamma_2''] + [\gamma_1', \gamma_1''] + [\gamma_2', \gamma_0'']) \mathrm{d}t = \int (2[\gamma_2, \gamma_0'] + [\gamma_1', \gamma_1'']) \mathrm{d}t.$$

Integrate Eq. (7)

$$0 = \int ([\gamma_0, \gamma_2'] + [\gamma_1, \gamma_1'] + [\gamma_2, \gamma_0']) \mathrm{d}t = \int (2[\gamma_2, \gamma_0'] + [\gamma_1, \gamma_1']) \mathrm{d}t;$$

therefore

$$\int r \mathrm{d}t = \int ([\gamma_1', \gamma_1''] - [\gamma_1, \gamma_1']) \mathrm{d}t.$$

We calculate

$$[\gamma_1, \gamma_1'] = f^2 + \frac{1}{2} f f'' - \frac{1}{4} f'^2, \ [\gamma_1', \gamma_1''] = f^2 + \frac{1}{2} f f'' + \frac{3}{4} f'^2 + \frac{1}{4} f' f''';$$

hence

$$\int r \mathrm{d}t = \int \left(f'^2 + \frac{1}{4} f' f''' \right) \mathrm{d}t = \int \left(f'^2 - \frac{1}{4} f''^2 \right) \mathrm{d}t.$$

Next

$$\int q^2 \mathrm{d}t = \int \left(2f' + \frac{1}{2} f''' \right)^2 \mathrm{d}t = \int \left(4f'^2 - 2f''^2 + \frac{1}{4} f'''^2 \right) \mathrm{d}t.$$

In the case of most interest, $a = \frac{1}{3}$, and we obtain

$$\int \left(r + \frac{a-1}{2} q^2 \right) \mathrm{d}t = \int \left(r - \frac{1}{3} q^2 \right) \mathrm{d}t = -\frac{1}{12} \int (4f'^2 - 5f''^2 + f'''^2) \mathrm{d}t. \quad (8)$$

Lemma 5.1. *The integral in (8) is non-negative, and it equals zero if and only if f is a first harmonic.*

Proof. Let

$$f'(t) = \sum_k c_k e^{ikt}, \ c_{-k} = \bar{c}_k$$

be the Fourier expansion. Then

$$\int_0^{2\pi} (4f'^2 - 5f''^2 + f'''^2) \mathrm{d}t = 2 \sum_{k>0} (4 - 5k^2 + k^4) |c_k|^2.$$

We have $4 - 5k^2 + k^4 = (k^2-1)(k^2-4) \geq 0$, and the sum is positive unless the only non-zero term is for $k = 1$, that is, f is a first harmonic. $\square$

Now, consider the general case

$$\int \left(r + \frac{a-1}{2} q^2 \right) \mathrm{d}t = \int \left((2a-1)f'^2 - \frac{4a-3}{4} f''^2 + \frac{a-1}{8} f'''^2 \right) \mathrm{d}t.$$

In terms of the Fourier coefficients of f', this is

$$2 \sum_{k>0} \left(2a - 1 - \frac{4a-3}{4} k^2 + \frac{a-1}{8} k^4 \right) |c_k|^2.$$

The expression in the parentheses equals

$$\frac{1}{8}(k^2-4)\left[(a-1)k^2 - 2(2a-1) \right]. \tag{9}$$

We also note that the quadratic term of P^a contains the factor a.

Theorem 2. *For $a < 0$, the circle γ_0 is a local minimum of $\mathcal{B}_a$; for $a \in (0, \frac{1}{3})$, it is a local maximum; for $a > \frac{7}{5}$, it is a local minimum, and in other cases, the Hessian is not sign-definite. The Hessian is degenerate (with 1-dimensional kernel) if and only if $a = \frac{k^2-2}{k^2-4}$ for some positive integer k.*

Proof. Let $a = 1$. Then, the sign of (9) is that of $-(k^2 - 4)$, which is positive for $k = 1$ and negative for $k \geq 3$.

Let $a > 1$. Then, the sign of (9) is positive for sufficiently large k. The Hessian is positive-definite if this sign is positive for all k. When $k = 1$, the first factor in (9) is negative, and so is the second one: $1 - 3a$. When $k \geq 3$, the first factor is positive, and the second one is positive if and only if $9(a-1) - 2(2a-1) > 0$, that is, $a > \frac{7}{5}$.

Let $0 < a < 1$. Then, the sign of (9) is negative for sufficiently large k. The Hessian is negative-definite if this sign is negative for all k. When $k = 1$, the first factor in (9) is negative, and the second factor equals $1 - 3a$. Thus, (9) is negative if and only if $a < \frac{1}{3}$. If this inequality is satisfied, then, for $k \geq 3$, the second factor in (9) is $2 - 4a - (1-a)k^2 < 0$.

If $a < 0$, then the analysis of the preceding paragraph still holds, but the factor a of the quadratic term in P^a changes the sign to the opposite.

Finally, the Hessian is degenerate when (9) is zero for some k. Solving this for a yields the last claim of the theorem. □

The last statement of the theorem agrees with Theorem 1. The numbers $a = \frac{k^2-2}{k^2-4}$ form the sequence

$$\frac{1}{3}, \frac{7}{5}, \frac{7}{6}, \frac{23}{21}, \ldots$$

that converges to 1. Each time that a crosses an element of this "spectrum"; the signature of the Hessian changes by 1.

6. The case of $a = \frac{1}{3}$

We recall that we attempt to describe extremal curves of the functional $\mathcal{B}_a = \int p^a(t)\mathrm{d}t$ on the space of periodic curves in centroaffine parameterization with $p > 0$. The case $a = \frac{1}{3}$ corresponds to the affine isoperimetric inequality. In particular, the functional then is translation invariant; see Lemma 3.6. In this case, $b = -\frac{3}{2}$, and Eq. (3) reads

$$F'' = 2F^{-\frac{1}{2}} + c. \tag{10}$$

Since $F'' \leq 0$ at the maximum and $F > 0$, we conclude that $c < 0$.

Next, integrate Eq. (10) to

$$(F')^2 = 8F^{\frac{1}{2}} + 2cF + d \tag{11}$$

(this is the equation of a level curve of the Hamiltonian, see Sect. 3).

Let $F(t) = G^2(t)$ where $G(t)$ is also a positive periodic function. Then, (11) becomes

$$(GG')^2 = \frac{c}{2}G^2 + 2G + d \tag{12}$$

(again renaming the constants).

The right-hand side of (12) is a quadratic polynomial in G, and it has at least two roots, because G is a periodic function that attains maximum and minimum. Since a quadratic polynomial has at most two roots, G oscillates between its maximum and minimum, and has no other critical values.

Example 6.1. Let us examine the case of a parallel translated n times traversed circle, which is a critical curve

$$\gamma = (A + r\cos\alpha, B + r\sin\alpha),$$

where $\alpha(t)$ is a function of the centroaffine parameter t. We assume that the range of t is $[0, 2\pi]$, that of α is $[0, 2\pi n]$, and the radius of the circle is r.

The range of the centroaffine parameter is twice the (algebraic) area bounded by the curve, and hence, $r = n^{-\frac{1}{2}}$.

We calculate

$$[\gamma, \gamma_\alpha] = r^2 + rA\cos\alpha + rB\sin\alpha,$$

and since $[\gamma, \gamma_t] = 1$, we have

$$\frac{\mathrm{d}t}{d\alpha} = r^2 + rA\cos\alpha + rB\sin\alpha.$$

This can be integrated

$$t = r^2\alpha + rA\sin\alpha - rB\cos\alpha + C,$$

or

$$\alpha + A\sin\alpha + B\cos\alpha = nt + C',$$

with the constants renamed. Since

$$A\sin\alpha + B\cos\alpha = -\sqrt{A^2 + B^2}\cos(\alpha + \theta)$$

with

$$\sin\theta = \frac{A}{\sqrt{A^2 + B^2}}, \ \cos\theta = -\frac{B}{\sqrt{A^2 + B^2}},$$

we can change the parameter α to obtain a simplified equation

$$\alpha(t) - A\cos\alpha(t) = nt + C \tag{13}$$

(once again renaming the constants). Then, $d\alpha/\mathrm{d}t = n(1 + A\sin\alpha)^{-1}$.

Next

$$p = [\gamma_t, \gamma_{tt}] = [\gamma_\alpha, \gamma_{\alpha\alpha}] \left(\frac{d\alpha}{\mathrm{d}t}\right)^3 = n^2(1 + A\sin\alpha)^{-3}.$$

Since $G = p^{-\frac{1}{3}}$, we conclude that

$$G(t) = n^{-\frac{2}{3}}(1 + A\sin\alpha(t)). \tag{14}$$

Let us continue with the general case. Let $0 < m < M$ be the minimum and the maximum of G, and let $c = -2k^2$. Write the right-hand side of (12) as $k^2(G - m)(M - G)$, then the differential equation becomes

$$GG' = k\sqrt{(G - m)(M - G)}, \tag{15}$$

where we allow the square root to take values at its positive and negative branches.

Set

$$\mu = \frac{M + m}{2},\ \varepsilon = \frac{M - m}{2}.$$

Note that since $G > 0$, we have $\varepsilon < \mu$.

Since G oscillates between m and M, let us make another substitution

$$G(t) = \mu + \varepsilon \sin\varphi(t), \tag{16}$$

where $\varphi(t)$ is not necessarily a periodic function anymore. Since G is 2π-periodic, $\varphi(t + 2\pi) = \varphi(t) + 2\pi n$ where n is an integer.

Substitute (16) into (15) to obtain $\varphi'(\mu + \varepsilon\sin\varphi) = k$. This differential equation integrates to

$$\mu\varphi(t) - \varepsilon\cos\varphi(t) = kt + C. \tag{17}$$

Since $0 < \varepsilon < \mu$, the left-hand side is a monotonic function of φ; therefore, this functional equation uniquely determines the function $\varphi(t)$.

Substituting $\varphi(t + 2\pi)$ in (17), we find that $\mu(\varphi(t + 2\pi) - \varphi(t)) = 2\pi k$. Since $\varphi(t + 2\pi) = \varphi(t) + 2\pi n$ with $n \in \mathbb{Z}$, we have $k = \mu n$. Thus, we have

$$\varphi(t) - \varepsilon\cos\varphi(t) = nt + C, \tag{18}$$

where, as before, we renamed the constants.

Theorem 3. *If $a = \frac{1}{3}$, then the relative extrema of the functional $\mathcal{B}_a$ constrained by $\mathcal{A}$ are multiply traversed conics.*

Proof. Let $\gamma(t)$ be a 2π-periodic critical curve. Then, Eq. (18) holds for some n, defining function $\varphi(t)$. Observe that Eq. (18) is identical to Eq. (13) from Example 6.1. Therefore, $\varphi(t)$ coincides with the function $\alpha(t)$. Similarly, Eqs. (16) and (14) coincide, and so, function $G(t)$ coincides with that from Example 6.1.

It follows that the centroaffine curvature $p(t)$ of the curve $\gamma(t)$ is the same as that of the parallel translated n-fold circle $\gamma_0(t)$. Consider the contact element $(\gamma(0), \gamma'(0))$. Acting on $\gamma_0(t)$ by an element of $SL(2,\mathbb{R})$, we can arrange for $(\gamma_0(0), \gamma_0'(0))$ to coincide with $(\gamma(0), \gamma'(0))$. The action of $SL(2,\mathbb{R})$

does not change the centroaffine curvature; hence, the two curves, γ and γ_0, satisfy the same second-order differential equation and have the same initial data. Therefore, they coincide. $\square$

Remark 6.2. The Lambert W function is the inverse function of the complex function $z = we^w$; see [5]. The function φ defined by equation (18) is related to the Lambert function in the following way.

Let us assume that $C = 0, n = 1$ in (18). Consider the complex function given by the equation $\xi = \eta - \varepsilon e^{i\eta}$. If η is real then $\mathfrak{Re}\ \xi = \eta - \varepsilon \cos\eta$, the expression that defines the function $\varphi(t)$.

Let $z = we^w$, and set $z = -i\varepsilon e^{i\xi},\ w = -i\varepsilon e^{i\eta}$. Then, $\ln z = w + \ln w$, that is, $\xi = \eta - \varepsilon e^{i\eta}$. Therefore, the inverse function $\eta(\xi)$ is conjugated to the Lambert function $w(z)$ by the exponential function.

7. Pictures and open problems

First, in Figs. 1, 2, 3, 4, we present extremal curves obtained in numerical experiments using a Mathematica applet created by Gil Bor.

Let us comment on a common geometrical feature of these curves. Recall the notion of the osculating circle of a smooth curve in Euclidean geometry:

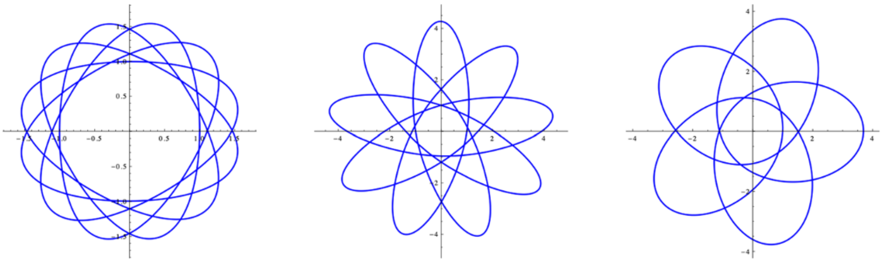


FIGURE 1. Curves having $a = -2, a = -1, a = 0.2$ and the rotation numbers $5, 5, 4$, respectively

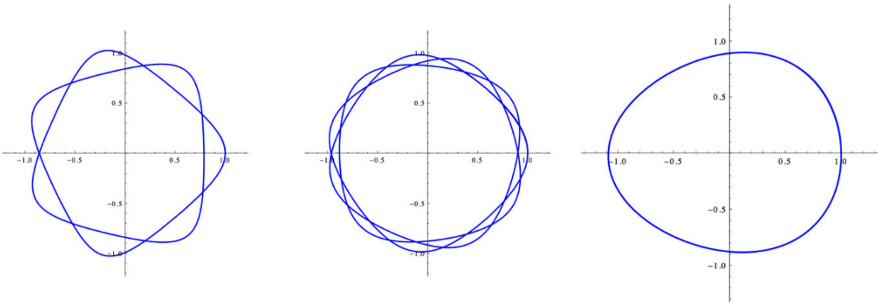


FIGURE 2. These curves have $a = 1.2, a = 1.2, a = 1.4$, respectively

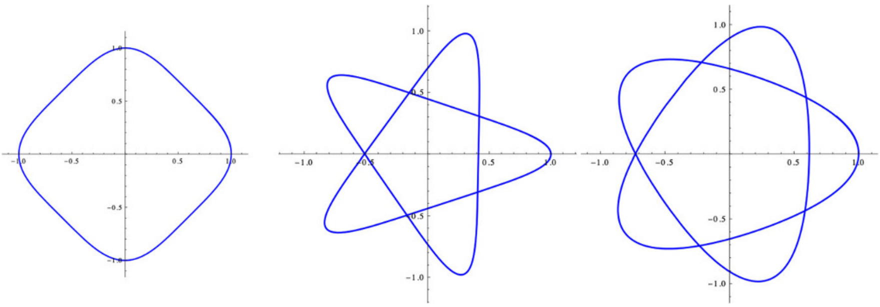


Figure 3. These curves have $a = 1.5, a = 1.5, a = 1.75$, respectively

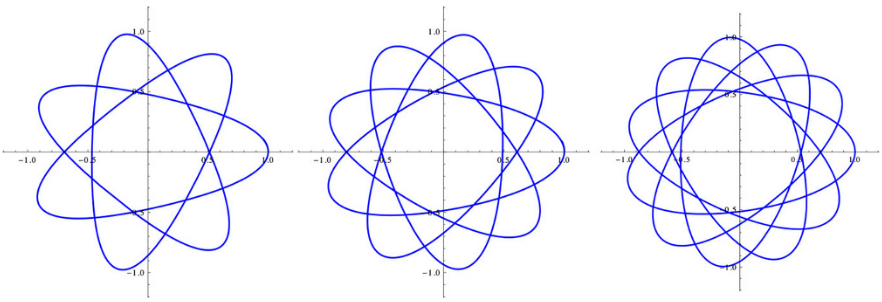


Figure 4. These curves have $a = 2, a = 2.5, a = 3$ and the rotation numbers $3, 4, 5$, respectively

this is a circle that is 2-order tangent to a curve at a given point, that is, it shares the curvature with the curve. Informally speaking, the osculating circle passes through three infinitesimally close points of the curve.

In centroaffine geometry, the role of osculating circles is played by osculating central conics. The space of central conics is 3-dimensional, and for every point of a star-shaped curve, there exists a central conic that is 2-order tangent to it at this point. Central conics have constant centroaffine curvature, and the osculating central conic at point $\gamma(t_0)$ has the constant centroaffine curvature $p(t_0)$.

Similarly to the case of osculating circles, the osculating central conic goes from one side of the curve to the other side if $p'(t_0) \neq 0$. If p has a non-degenerate maximum or minimum at point t_0, that is, $\gamma(t_0)$ is a centroaffine vertex, then the osculating central conic lies on one side of the curve near this point.

In the following lemma, we prove that the centroaffine curvature of *extremal curves* takes only two values at their centroaffine vertices, its maximum and minimum.

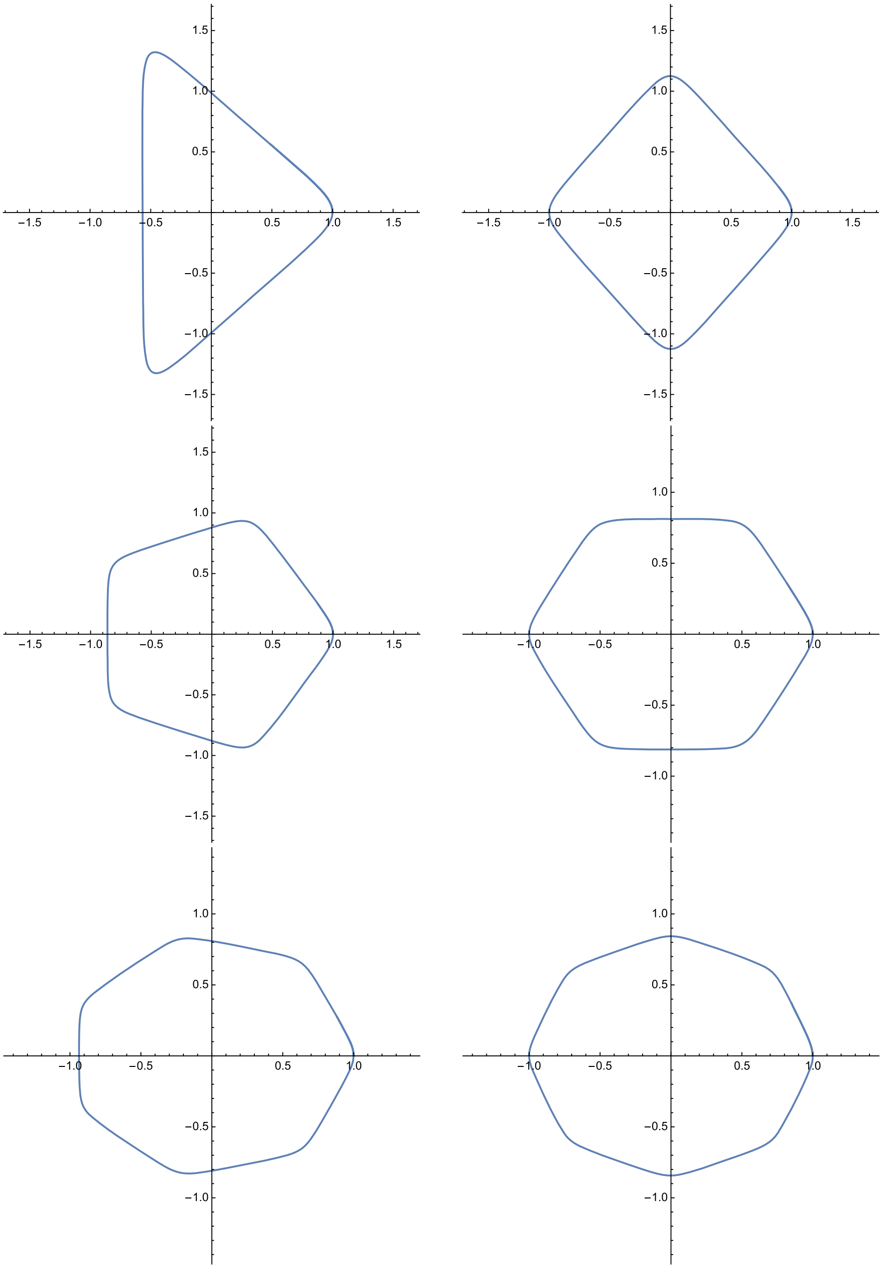


FIGURE 5. $a = \frac{7}{6}$, $a = \frac{23}{21}$, $a = \frac{17}{16}$, $a = \frac{47}{45}$, $a = \frac{31}{30}$, $a = \frac{79}{77}$

Lemma 7.1. *The function F, introduced in* Sect. 3, *and hence the centroaffine curvature p, has only two critical values, its maximum and minimum.*

Proof. Since the curve is closed, the function F attains its maximum and minimum. It has no other critical values, because the right-hand side of Eq. (5), as a function of F, has no more roots than the number of sign changes among its three coefficients; see [10] (part 5, chapter 1, §6, No 77). □

Looking again at the above figures it is fairly obvious that the case $a = 1.4$ in Fig. 2, the "egg" lacks the same symmetry all other curves have. This seems related to Theorem 1 asserting that for all a of the form $a = \frac{k^2-2}{k^2-4}$, $k \in \mathbb{Z}_{>2}$, a circle (which is a critical curve for any a) admits a non-trivial infinitesimal deformation. Indeed, $1.4 = \frac{3^2-2}{3^2-4}$.

A more systematic computer experiment (again using the Mathematica applet by Gil Bor) leads to Fig. 5 where curves corresponding to $k = 4, 5, 6, 7, 8, 9$ are displayed. We recall that for any value a, the functional $\mathcal{B}_a$ is invariant under $SL(2, \mathbb{R})$. These correspond, of course, to trivial deformations.

Figure 5 leads us to the conjecture that for $a = \frac{k^2-2}{k^2-4}$, there exists an extremal curve which is a "rounded $(k-1)$-gon". Unfortunately, our software is currently not powerful enough to verify this. We hope to return to this point in the future. It is worth pointing out that these $(k-1)$-gons seem to approach a circle for k very large. This is, at least, consistent with $a \to 1$ when $k \to \infty$ and the circle is indeed rigid for $\mathcal{B}_1$. As a rather special case, the egg (case $a = 1.4$ in Fig. 2) should be considered a rounded 2-gon.

We collect some further questions.

- What is the minimal rotation number of a periodic solution of $\mathcal{B}_a$ in dependence of a? It seems that this minimal rotation number goes to infinity with a.
- In general, what happens if $a \to \infty$ respectively, $a \to -\infty$?
- Is there some kind of duality for positive and negative values of a?
- Is there an appropriate gradient flow of $\mathcal{B}_a$ similar to the metric case, see [7] and [13]?

Acknowledgements

We are very grateful to Gil Bor for writing a Mathematica program that made it possible to experiment with extremal curves. PA was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through Germany's Excellence Strategy EXC-2181/1-390900948 (the Heidelberg STRUCTURES Excellence Cluster), the Transregional Colloborative Research Center CRC/TRR 191 (281071066), and the Research Training Group RTG 2229 (281869850). ST was supported by NSF grant DMS-2005444.

Appendix

In this appendix, we describe a curious symmetry of Eq. (5). In the discovery of this symmetry, we were motivated by Bohlin's theorem as described in Appendix 1 of [1].

Consider the family of equations

$$\left(\frac{\mathrm{d}F}{\mathrm{d}t}\right)^2 = uF^q + vF + w, \tag{19}$$

where $F(t)$ is an unknown function and u, v, w, q are parameters. We are looking for changes of independent and dependent variables

$$F(t) = G(\tau)^{\mu}, \ \frac{\mathrm{d}\tau}{\mathrm{d}t} = G^{\lambda}$$

that preserve the form of Eq. (19), but possibly change the parameters u, v, w, and q.

Theorem 4. *These changes of variables form a group, the group of permutations S_3. The orbit of the exponent q is*

$$\left\{q, \frac{1}{q}, 1-q, \frac{1}{1-q}, \frac{q}{q-1}, \frac{q-1}{q}\right\}.$$

We note that this is precisely how the permutations of four points in the projective line affect their cross-ratio.

Proof. Denote the new parameters by $\bar{u}, \bar{v}, \bar{w}$, and $\bar{q}$.

Using the chain rule, one obtains the differential equation on G

$$\left(\frac{\mathrm{d}G}{\mathrm{d}\tau}\right)^2 = \frac{u}{\mu^2}G^{\mu q - 2\mu - 2\lambda + 2} + \frac{v}{\mu^2}G^{2-2\lambda-\mu} + \frac{w}{\mu^2}G^{2-2\lambda-2\mu}.$$

For this equation to have the same form as (19), one needs the following relation between the exponents to hold:

$$\{0, 1\} \subset \{\mu q - 2\mu - 2\lambda + 2, 2 - 2\lambda - \mu, 2 - 2\lambda - 2\mu\}.$$

Thus, one needs to consider six cases. We present one of them

$$\begin{cases} \mu q - 2\mu - 2\lambda + 2 = 0 \\ 2 - 2\lambda - \mu = 1. \end{cases}$$

Hence

$$\mu = \frac{1}{1-q}, \ \bar{q} = 2 - 2\lambda - 2\mu = \frac{q}{q-1}, \ \bar{u} = w(q-1)^2, \ \bar{v} = v(q-1)^2,$$
$$\bar{w} = u(q-1)^2.$$

The other five cases are analyzed in a similar way. □

Returning to Eq. (5), one has $q = b + 2 = \frac{2a-1}{a-1}$. The S_3-orbit of the exponent a is as follows:

$$\left\{a, 1-a, \frac{2a-1}{3a-1}, \frac{a}{3a-1}, \frac{2a-1}{3a-2}, \frac{a-1}{3a-2}\right\}.$$

In particular, the orbit of 1 is $\{1, 1/2, 0\}$ and the orbit of $1/3$ is $\{1/3, 2/3, \infty\}$, and these are special values in our study. All other orbits consist of six elements.

We end with the following remarks. First, the function G (or equivalently F) needs to be positive in order for $\frac{\mathrm{d}\tau}{\mathrm{d}t} = G^\lambda$ being an actual change of coordinates. This is always satisfied in our situation. Moreover, F is periodic if and only if G is. However, in the transition from a solution F of (5), a special case of (19), to a curve γ, it seems hard to see if γ is again periodic. Therefore, the groups of coordinate changes above might or might not map closed curves to closed curves.

References

[1] Arnold, V.: Huygens and Barrow, Newton and Hooke. Pioneers in Mathematical Analysis and Catastrophe Theory from Evolvents to Quasicrystals. Birkhäuser, Basel (1990)

[2] Arnold, V.: The geometry of spherical curves and quaternion algebra. Russ. Math. Surv. **50**, 1–68 (1995)

[3] Bennequin, D.: Entrelacements et équations de Pfaff. Third Schnepfenried geometry conference, Vol. 1 (Schnepfenried, : 87–161, Astérisque, 107–108, p. 1983. Soc. Math, France, Paris (1982)

[4] Bialy, M., Bor, G., Tabachnikov, S.: Self-Bäcklund curves in centroaffine geometry and Lamé's equation. arXiv:2010.02719

[5] Corless, R., Gonnet, G., Hare, D., Jeffrey, D., Knuth, D.: On the Lambert W function. Adv. Comput. Math. **5**, 329–359 (1996)

[6] Lutwak, E.: Selected Affine Isoperimetric Inequalities. Handbook of Convex Geometry, Vol. A, pp. 151–176. North-Holland, Amsterdam (1993)

[7] Okabe, S., Pozzi, P., Wheeler, G.: A gradient flow for the p-elastic energy defined on planar curves. Math. Ann. **378**, 777–828 (2020)

[8] Ovsienko, V., Tabachnikov, S.: Sturm theory, Ghys theorem on zeroes of the Schwarzian derivative and flattening of Legendrian curves. Selecta Math. (N.S.) **2**, 297–307 (1996)

[9] Pinkall, U.: Hamiltonian flows on the space of star-shaped curves. Res. Math. **27**, 328–332 (1995)

[10] Pólya, G., Szegö, G.: Problems and Theorems in Analysis. II. Theory of Functions, Zeros, Polynomials, Determinants, Number Theory, Geometry. Springer, Berlin (1998)

[11] Sapiro, G., Tannenbaum, A.: On affine plane curve evolution. J. Funct. Anal. **119**, 79–120 (1994)

[12] Singer, D.: Lectures on elastic curves and rods. Curvature and variational modeling in physics and biophysics. In: AIP Conf. Proc., 1002, Amer. Inst. Phys., Melville, pp. 3–32 (2008)

[13] Wanatave, K.: Planar p-elastic curves and related generalized complete elliptic integrals. Kodai Math. J. **37**, 453–474 (2014)

[14] Whittaker, E.T., Watson, G.N.: A Course of Modern Analysis. An Introduction to the General Theory of Infinite Processes and of Analytic Functions: with

an Account of the Principal Transcendental Functions, 4th edn. Cambridge University Press, New York (1962)

Peter Albers
Mathematisches Institut
Universität Heidelberg
69120 Heidelberg
Germany
e-mail: peter.albers@uni-heidelberg.de

Serge Tabachnikov
Department of Mathematics
Pennsylvania State University
University Park
PA 16802
USA
e-mail: tabachni@math.psu.edu

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:42
https://doi.org/10.1007/s11784-022-00959-4
Published online April 29, 2022

Journal of Fixed Point Theory and Applications

C^0-robustness of topological entropy for geodesic flows

Marcelo R. R. Alves, Lucas Dahinden, Matthias Meiwes and Louis Merlin

To Claude Viterbo on the occasion of his 60th birthday.

Abstract. In this paper, we study the regularity of topological entropy, as a function on the space of Riemannian metrics endowed with the C^0 topology. We establish several instances of entropy robustness (persistence of positive entropy after small C^0 perturbations). A large part of this paper is dedicated to metrics on the two-dimensional torus, for which our main results are that metrics with a contractible closed geodesic have robust entropy (thus, generalizing and quantifying a result of Denvir–Mackay) and that metrics with robust positive entropy on the torus are C^∞ generic. Moreover, we quantify the asymptotic behavior of volume entropy in the Teichmüller space of hyperbolic metrics on a punctured torus, which bounds from below the topological entropy for these metrics. For general closed manifolds of dimension at least 2, we prove that the set of metrics with robust and large positive entropy is C^0-large in the sense that it is dense and contains cones and arbitrarily large balls.

Mathematics Subject Classification. Primary 37B40, 37D40, 53D25.

M. R. R. Alves was supported by the ERC consolidator grant 646649 "SymplecticEinstein" and by the Senior Postdoctoral fellowship of the Research Foundation—Flanders (FWO) in fundamental research 1286921N.

L. Dahinden was supported by Deutsche Forschungsgemeinschaft (DFG) under Germany's Excellence Strategy EXC-2181/1-390900948 (the Heidelberg STRUCTURES Excellence Cluster).

M. Meiwes was supported by RWTH Aachen University and the Chair for Geometry and Analysis of the RWTH Aachen University.

L. Merlin was supported by RWTH Aachen University and the Chair for Geometry and Analysis of the RWTH Aachen University.

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

Contents

1. Introduction

1.1. Context

In this paper, we study robustness properties for the topological entropy of Riemannian geodesic flows with respect to the C^0-topology on the space of Riemannian metrics.

The space of metrics. Let Q be a closed manifold and $\mathfrak{G}(Q)$ be the space of C^∞-smooth Riemannian metrics on Q. For $g, g' \in \mathfrak{G}(Q)$, we say that $g \prec g'$ if $g_x(v,v) \leqslant g'_x(v,v)$ for all x, v, and for $C \in \mathbb{R}$ we define $Cg \in \mathfrak{G}(Q)$ by $Cg_x(v,w) = C \cdot g_x(v,w)$, for all $x \in Q, v, w \in T_xQ$. We consider on $\mathfrak{G}(Q)$ the metric d_{C^0} defined by

$$d_{C^0}(g,g') = \inf\left\{\log(C) \;\middle|\; \frac{1}{C}g \prec g' \prec Cg;\, C > 0\right\}. \tag{1.1}$$

The metric d_{C^0} defines the C^0 topology on $\mathfrak{G}(Q)$. d_{C^0} is a variant of the Riemannian Banach-Mazur distance from [41], see below for further discussion. From a purely metric point of view, d_{C^0} is natural since geometric quantities such as the volume of subsets of (Q,g), the diameter of (Q,g), and the Riemannian distance function d_g on Q, are all continuous with respect to

d_{C^0}.[1] In studying these quantities, it is, therefore, more natural to consider the d_{C^0}-distance than finer distance functions involving also derivatives of g.

Topological entropy. The topological entropy h_{top} is a numerical invariant of a dynamical system that measures orbit complexity, see Sect. 1.4 for its definition. In this paper we study the topological entropy h_{top} of Riemannian geodesic flows φ_g, $g \in \mathfrak{G}(Q)$, seen as a function on the metric space $(\mathfrak{G}(Q), d_{C^0})$. We will show, especially when Q is two-dimensional, that h_{top} is more robust under perturbations of the metric than one could imagine at first sight. The reason why this robustness is not clear is that the geodesic vector field depends on the first derivatives of g, and therefore does not change continuously with d_{C^0}: a d_{C^0}-small change of the metric can result in a large change of the geodesic flow, meaning that a priori it is reasonable to believe that a dynamical quantity such as the topological entropy would be subject to a large change. This view is reinforced by the lack of continuity in topologies much stronger than C^0: in [34], it is shown that on the class of C^r maps, $r < \infty$, h_{top} fails to be upper semi-continuous in the C^∞ topology and even smooth perturbations of smooth diffeomorphisms on closed 3-manifolds can collapse topological entropy to 0, see [16, Section 2] and also [30].

Our investigations are part of the more general study of how the topological entropy of the geodesic flow behaves with respect to perturbations of the metric. This is a long-standing problem, and greatly depends on the topology considered on the space of metrics: see for example [25] and [15]. Nowadays, a satisfactory answer is given for metrics of negative curvature and C^1 perturbations, even for some non-compact manifolds [39].

Two continuous invariants. The first motivation to study the continuity properties of h_{top} on $(\mathfrak{G}(Q), d_{C^0})$ is that there are two functions on $(\mathfrak{G}(Q), d_{C^0})$ which bound h_{top} from below and which are clearly continuous in $(\mathfrak{G}(Q), d_{C^0})$: the volume entropy h_{vol} and the exponential growth rate Γ_{Morse} of the Morse homology of the based loop space or the free loop space. The fact that h_{vol} is a lower bound for h_{top} is due to Manning [27], and the fact that Γ_{Morse} is a lower bound for h_{top} is due to Paternain [36] (in the case of based loop spaces) and [29] (in the case of free loop spaces). These two functions either vanish on all of $\mathfrak{G}(Q)$ or are positive for every element of $\mathfrak{G}(Q)$. If one of these functions does not vanish, then the topological entropy is *robust* for all $g \in \mathfrak{G}(Q)$ in the sense that for any $g \in \mathfrak{G}(Q)$, there is $c > 0$ and an open neighborhood $\mathcal{U}_g$ of g in $(\mathfrak{G}(Q), d_{C^0})$ such that

$$h_{\mathrm{top}}(\varphi_{g'}) > c \text{ for all } g' \in \mathcal{U}_g. \tag{1.2}$$

This shows that for manifolds with positive h_{vol} or positive Γ_{Morse} the topological entropy of geodesic flows cannot be destroyed by a C^0-small perturbations. Our main results show that some of this robustness persists for manifolds with vanishing h_{vol} and Γ_{Morse}. See for example Theorem 10, which shows that C^∞ generic metrics on the two-dimensional torus have robust h_{top}.

[1] Moreover, the logarithms of these quantities are Lipschitz with respect to d_{C^0}.

Homogeneity. We proceed to discuss the differences between h_{top} and the functions h_{vol} and Γ_{Morse}. Recall that the volume entropy $h_{\mathrm{vol}}(g)$ of (Q,g) measures the exponential growth rate of the volume of Riemannian balls with respect to the radius on the universal cover of (Q,g). To prove that $h_{\mathrm{vol}}(g)$ is continuous on $(\mathfrak{G}(Q),d_{C^0})$, we first observe that h_{vol} is monotonous: if $g,g'\in\mathfrak{G}(Q)$ and $g\leqslant g'$ then $h_{\mathrm{vol}}(g)\geqslant h_{\mathrm{vol}}(g')$. Furthermore, h_{vol} is homogeneous: $h_{\mathrm{vol}}(Cg)=C^{-\frac{1}{2}}h_{\mathrm{vol}}(g)$. The continuity of h_{vol} on $(\mathfrak{G}(Q),d_{C^0})$ follows easily from these two properties. The function Γ_{Morse} is also monotonous and homogeneous on $\mathfrak{G}(Q)$, and this was explored in [16] to study C^0-robustness of h_{top} of geodesic flows, and more generally Reeb flows. Since monotonous and homogeneous functions on $\mathfrak{G}(Q)$ are either 0 or always positive, the homogeneous function h_{top} cannot be monotonous for metrics on the torus. Moreover, Theorem 12 shows that it is possible to increase h_{top} arbitrarily with C^0-small perturbations on $(\mathfrak{G}(Q),d_{C^0})$. In particular, this implies that C^0-near any metric we can find another metric that doubles its entropy. It follows that

Corollary 1. (From Theorem 12) *The topological entropy of geodesic flows is not homogeneous on any closed manifold of dimension at least 2.*

The set of high entropy metrics. The results of this paper show that on the other hand in many situations h_{top} cannot be arbitrarily decreased by small perturbations on $(\mathfrak{G}(Q),d_{C^0})$. Theorem 34 shows that on the 2-torus, a generic metric has robust positive topological entropy and Theorem 12 shows that on any manifold of dimension at least 2 the set of robust high entropy metrics is C^0-dense.

Our results and this discussion suggest the following conjecture:

Conjecture 2. *If Q is a closed surface, then $h_{\mathrm{top}}(g)$ is robust whenever it does not vanish.*

Although all methods presented in this paper yield robust lower bounds which are not sharp, we ask also the following question, which, if answered in the affirmative, implies Conjecture 2.

Question 3. Is, for any closed surface Q, h_{top} lower semi-continuous on $(\mathfrak{G}(Q), d_{C^0})$?

The recent results in [3] provide some reason to expect a positive answer to this question.

The C^0 distance d_{C^0} is a variant of the Riemannian Banach–Mazur distance d_{RBM} defined by Stojisavljević and Zhang [41], a pseudo-metric on $\mathfrak{G}(Q)$.[2] d_{RBM} is defined as $d_{\mathrm{RBM}}(g,g'):=\inf d_{C^0}(g,\varphi^*g')$, where the infimum is taken over all diffeomorphisms $\varphi:Q\to Q$. The pseudo-metric d_{RBM} itself is an adaption to $\mathfrak{G}(Q)$ of the symplectic Banach–Mazur distance, which was first proposed by Ostrover and Polterovich to study the symplectic geometry of Liouville domains and studied, e.g., in [41,42]. In [41] the authors

[2]We note that the continuity and local robustness results for h_{top} in this paper still hold when considering d_{RBM} instead of d_{C^0}.

investigate the large-scale geometry of $(\mathfrak{G}(Q), d_{\mathrm{RBM}})$ and one of their result is that for $Q = T^2$ and every $n \in \mathbb{N}$, there is a quasi-isometric embedding $\Phi_n : (\mathbb{R}^n, |\cdot|_\infty) \to (\mathfrak{G}(T^2), d_{\mathrm{RBM}})$[3], so informally speaking $(\mathfrak{G}(T^2), d_{\mathrm{RBM}})$ is "very large" in the metric sense. The construction in [41] can be easily modified to have its image in the set of large entropy metrics by adding a C^0-small modification, cf. Corollary 13. For different points of view on the study of C^0-stable properties of Riemannian metrics, we refer the reader to [11,33].

Mañé's formula for the topological entropy. In [31], Mañé established the following remarkable formula for the topological entropy of the geodesic flow of a C^∞-smooth Riemannian metric g on a manifold Q:

$$h_{\mathrm{top}}(\phi_g) = \lim_{T\to+\infty} \frac{1}{T} \int_{Q\times Q} \log(\mathcal{N}_T^g(p,q))d\omega_g(p)d\omega_g(q). \tag{1.3}$$

Here, $\mathcal{N}_T^g(x,y)$ denotes the number of geodesic chords of g from p to q with length $< T$, $d\omega_g(p)$ means integration in the variable p with respect to the Riemannian volume form ω_g on Q associated with g, and $d\omega_g(q)$ means integration in the variable q with respect to ω_g. This formula gives a characterization of $h_{\mathrm{top}}(\phi_g)$ in terms of the purely geometric quantity which appears on the right side of (1.3).

The right-hand side of (1.3) is the exponential growth of the average number of geodesics connecting two points of Q. Using Mañé's formula, our results provide surprising robustness of this exponential growth in case Q is a surface. For example, Theorem 34 implies that for a C^∞-generic metric g on T^2 this exponential growth is positive and cannot be completely destroyed by C^0-small perturbation of g. This is far from obvious, given that the counting function $\mathcal{N}_T^g(p,q)$ can undergo dramatic changes when we make C^0-small perturbations.

1.2. Results, strategy, and layout of the paper

To find a metric g with robust entropy, we proceed in two steps. First, we prove a forcing type argument, which is a geometric feature of g that implies positivity of the entropy. One of the best-known examples is a theorem of [18] stating that a metric on the torus that admits a contractible closed (non-constant) geodesic must have positive entropy. Then, we show that the forcing situation is $\mathcal{C}^0$ robust, that is, persists after d_{C^0}-small perturbations of g.

The analogy with [18] is not incidental. In Sect. 2, we describe how a contractible closed geodesic in the two-torus forces robust topological entropy. Here and throughout the paper, $l_g(\gamma)$ denotes the length of the curve γ with respect to the metric g.

Definition 4. Let $\Pi : \mathbb{R}^2 \to T^2 = \mathbb{R}^2/\mathbb{Z}^2$ be the canonical projection and g_{flat} the push forward by Π of the euclidean metric. For any metric $g \in \mathfrak{G}(T^2)$

[3]In fact, it is an embedding into the space of metrics with fixed volume and diameter bounded by a fixed constant.

define

$$D(g) = e^{d_{C^0}(g_{\text{flat}},g)} = \inf\left\{C \mid \frac{1}{C}g \prec g_{\text{flat}} \prec Cg;\, C > 0\right\}.$$

Note that $\frac{1}{\sqrt{D(g)}}l_g(\gamma) \leqslant l_{g_{\text{flat}}}(\gamma) \leqslant \sqrt{D(g)}l_g(\gamma)$ for all curves γ.

Theorem 5. *Let g_0 be a Riemannian metric on T^2 with a contractible closed geodesic γ_0. Then, g_0 has robust topological entropy.*

Moreover, the following holds. Given $\delta > 0$ there is $\epsilon > 0$ such that for all g with $d_{C^0}(g, g_0) < \epsilon$

(1) $h_{\text{top}}(g) > \left(\left\lceil \frac{\sqrt{D(g_0)}\Lambda}{2} + \delta\right\rceil \sqrt{D(g_0) + \delta}\right)^{-1} \log 3$, *where* $\Lambda = l_{g_0}(\gamma_0)$ *if* g_0 *is bumpy and* $\Lambda = 2(l_{g_0}(\gamma_0) + \sqrt{D(g_0)})$ *in general.*

(2) $h_{\text{top}}(g) > \min\left\{\frac{1}{\sqrt{4\text{area}_{g_0}(T^2)+L^2}}, \frac{2}{3L}\right\}\log 2$, *where* $L = l_{g_0}(\gamma_0)$ *if* g_0 *is bumpy and* $L = \max\{4\sqrt{4\text{area}_{g_0}(T^2) + l_{g_0}(\gamma_0)^2}, 3l_{g_0}(\gamma_0)\}$ *in general.*

The proofs of the two parts of this theorem have similar, yet different core ideas. For a clean exposition, we prove the first part of this theorem first in the case where g_0 is bumpy in Sect. 2.1, then in Sect. 2.2 in the degenerate case. We then prove the second part of the theorem in Sect. 2.3.

A corollary of Theorem 5 (2) which we believe to be interesting in its own right is the following:

Corollary 6. *Let g be a Riemannian metric on the 2-torus whose area is $\leqslant 1$ and which has a contractible closed geodesic whose length is $\leqslant 1$. Then, $h_{\text{top}}(\phi_g) \geqslant \frac{1}{20}$.*

An interesting question is to know whether this result remains true without any assumption on the length of the contractible closed geodesic, i.e., if there is a lower bound for the h_{top} of geodesic flows of Riemannian metrics on the 2-torus with area 1 and which have a contractible closed geodesic.

One result which is needed for Theorem 5 and throughout the paper is presented in Appendix A, where we prove C^0-robustness of the length spectrum of a Riemannian manifold with bumpy metric. In our proof, we aim at using as little technology as possible. The main statement is the following, for the definition of topological non-degeneracy see Definition 48.

Proposition 7. *Suppose that $0 < e \in (a, b)$ is the only energy value in $[a, b]$ of a closed geodesic on a closed manifold and that all geodesics with energy e are (topologically) non-degenerate. Then, any C^0-close Riemannian metric has a closed geodesic in the same homotopy class with energy close to e.*

Remark 8. For the first bounds in Theorem 5, the condition for g_0 to be bumpy is stronger than necessary; it would suffice to ask that only γ_0 is non-degenerate. Correspondingly, in Proposition 7 one may ask that γ_0 is topologically non-degenerate and isolated in the loop space instead of its energy value to be isolated in the energy spectrum. However, this would significantly complicate the proof of the proposition, cf. Remark 49, and since we also treat the degenerate case, this additional complication would yield no significant improvement for Theorem 5.

In Sect. 3, we turn our attention to the *volume entropy* (to be defined in Paragraph 1.4) for hyperbolic metrics on the one-holed torus. The volume entropy is a lower bound for the topological entropy, so that a metric with robust h_{vol} has in particular robust h_{top}. Consider again the set of metrics on the two-torus that admit a non-constant simple contractible closed geodesic. This geodesic is in particular separating, one component of its complement is a disk, the other is a one-holed torus. In Sect. 3, we assume that the metric on the one-holed torus is hyperbolic (constant curvature -1). Denote by $\mathcal{H}$ this set of metrics on the torus. We prove the following result.

Theorem 9. *Any Riemannian metric $g \in \mathcal{H}$ has robust h_{vol} and, thus, robust h_{top}.*

Following the general approach (forcing plus robustness), in Sect. 4, we show that a certain configuration of curves in the torus, which we call a ribbon, forces robust topological entropy. Moreover, this condition is C^∞ generic, leading to a series of results that are more precisely stated in Theorems 30 and 34.

Theorem 10. *Four closed geodesics on the two-torus that form a ribbon force robust h_{top}. A C^∞ generic metric possesses four closed geodesics that form a ribbon and has, thus, robust h_{top}.*

In Sect. 5, we find a robust (albeit non-generic) forcing condition, which we call retractable neck with entropic body on a Riemannian manifold of any dimension. The following theorem is more precisely stated in Theorem 42:

Theorem 11. *Let the closed Riemannian manifold (M, g) (of any dimension) have a retractable neck and entropic body. Then, g has robust h_{top}.*

This condition is readily arranged by an explicit construction, see Example 43, which allows us to prove the following statements on the size of the space of metrics with large entropy, which we denote by $\mathfrak{G}^e(Q) = \{g \in \mathfrak{G}(Q) \mid h_{\mathrm{top}}(g) > e\}$.

Theorem 12. *Let Q be a closed manifold Q of dimension at least 2. For any $e > 0$ and for any metric $g \in \mathfrak{G}(Q)$, there is a C^0-continuous path $g(s) : (0, \infty) \to \mathfrak{G}^e(Q)$ such that for all $s \in (0, \infty)$:*

- *$d_{C^0}(g(s), g) < s$,*
- *and $B_{r(s)}(g(s)) \subset \mathfrak{G}^e$, where $B_{r(s)}(g(s))$ is the d_{C^0}-ball of radius $r(s) = \log \frac{s+3}{2+(s+1)e^{-s/8}}$ around $g(s)$.*

Please note that in our construction, $r(s)$ does not depend on e, but the path $g(s)$ does.

Corollary 13. • *For any $e > 0$, $\mathfrak{G}^e(Q) = \{g \in \mathfrak{G}(Q) \mid h_{\mathrm{top}}(g) > e\}$ is C^0-dense in $\mathfrak{G}(Q)$.*

- *As* $\lim_{s\to\infty} r(s) = \infty$, *we find arbitrarily big balls of arbitrarily large entropy.*
- *For* $Q = T^2$ *let* $\overline{\mathfrak{G}}(T^2)$ *be the set of metrics with volume 1 and diameter* $\leqslant 101$. *Then, for every* $n \in \mathbb{N}$ *there is a quasi-isometric embedding* $\Phi_n : (\mathbb{R}^n, |\cdot|_\infty) \to (\overline{\mathfrak{G}}^e(T^2), d_{\mathrm{RBM}})$.

1.3. Related developments

In an ongoing joint project of the authors Alves, Dahinden and Meiwes with Abror Pirnapasov, we are generalizing some of the results of the present paper, such as item (1) of Theorems 5 and 34, to the category of Reeb flows. Reeb flows on contact 3-manifolds are a generalization of geodesic flows of Riemannian metrics on surfaces, and the C^0-distance on the space of contact forms considered in [16] generalizes to the space of Reeb flows on unit tangent bundles of surfaces (endowed with the geodesic contact structure) the C^0-distance on the space of Riemannian metrics of surfaces that we consider here. For this generalization, one must use symplectic topological methods developed in [7] to study h_{top} of Reeb flows.

On the other hand, Corollary 6 cannot be generalized to the category of Reeb flows. Using the methods of [1], one can construct Reeb flows on the 3-torus $(T^3, \xi_{\mathrm{geo}})$ endowed with the geodesic contact structure contradicting any reasonable generalization of Corollary 6.

The questions considered in the present paper were inspired by [16] and [14].

1.4. Setup and definitions

Let (Q, g) be a compact Riemannian manifold. Throughout this paper, we will be interested in ergodic properties of its geodesic flow. Let T^1Q denote the unit tangent bundle of Q. For a vector $v \in T^1Q$, we consider the geodesic γ_v defined by the initial condition $\gamma_v'(0) = v$.

The geodesic flow of (Q, g), denoted φ_g^t (we sometimes omit t or g when the context is clear) is defined by

$$\begin{aligned} \varphi_g^t : T^1Q &\longrightarrow T^1Q \\ v &\longmapsto \gamma_v(t) \end{aligned}.$$

When Q is a manifold with boundary, we restrict φ^t to the forward recurrent set in T^1Q.

Entropies. Denote by $\Gamma_t f(t) := \limsup_t \frac{1}{t} \log f(t)$ the exponential growth in t of a function $f(t)$. We use the following definition of topological entropy:

Definition 14. Let $\varphi : (M, d) \to (M, d)$ be a continuous self map of a compact metric space. Define the dynamical metric

$$d_k(x, y) = \sup_{0 \leqslant l \leqslant k} d(\varphi^l x, \varphi^l y).$$

A (δ, k)-separated set is a set whose points have pairwise d_k-distance $\geqslant \delta$. The topological entropy of φ is then defined as follows:

$$h_{\mathrm{top}}(\varphi) = \lim_{\delta\to 0} \Gamma_k \sup_{\Delta} |\Delta|,$$

where the supremum runs over all (δ, k)-separated sets. There exists a "dual" characterization of the entropy with spanning sets instead of separated sets. Note that for any fixed $\delta > 0$ any sequence Δ_k of (δ, k)-separated sets provides the lower bound

$$h_{\text{top}}(\varphi) \geqslant \Gamma_k|\Delta_k|$$

on the entropy. This is how we obtain all lower bounds to the entropy in this paper. The difficulty is to find a growing sequence, i.e., one that provides a positive lower bound.

Topological entropy is well known to be independent of the choice of metric d (generating the same topology). Further, if $\varphi : \mathbb{R} \times M \to M$ is an autonomous flow, then it is well known that

$$h_{\text{top}}(\varphi^1) = \frac{1}{T} h_{\text{top}}(\varphi^T)$$

for any $T > 0$.

We say the topological entropy $h_{\text{top}}(g)$ of a geodesic flow on a Riemannian manifold is the topological entropy of the time-1 map of the geodesic flow on the unit sphere bundle. $\square$

Volume entropy is a related invariant, more geometric in nature. Let M be a closed manifold (possibly with non empty boundary) and $\widetilde{M}$ its universal cover that we endow with the metric pulled-back from M. We choose a basepoint $x \in \widetilde{M}$.

Definition 15. By a result of [27], the following limit exists and is independent of the basepoint x. We call it the volume entropy of M.

$$h_{\text{vol}} = \lim_{R \to \infty} \frac{\log \operatorname{Vol} B(x, R)}{R},$$

where $B(x, R)$ is the ball of radius R centered at x in the universal cover $\widetilde{M}$. $\square$

A comparison between h_{vol} and h_{top} is done in [27]: we have

$$h_{\text{vol}} \leqslant h_{\text{top}}$$

for any closed manifold M. Note that the result also applies to manifolds with boundary. Even if this case is not stated in [27], the same proof extends to this case. Equality happens, e.g., when M has nonpositive curvature.

In the context that we have in mind, the above inequality simply means that a metric with robust h_{vol} also has robust h_{top}.

Length, energy, area and loop spaces. Let Q be a closed manifold. Let g be Riemannian metric on Q. The *length* resp. the *energy* of a smooth curve $x : [0, 1] \to Q$ with respect to g are

$$l_g(x) = \int_0^1 \sqrt{g(x'(t), x'(t))}\, \mathrm{d}t,$$

$$\mathcal{E}_g(x) = \frac{1}{2} \int_0^1 g(x'(t), x'(t))\, \mathrm{d}t.$$

Note that, for $C > 0$, $l_g(x) = \frac{1}{\sqrt{C}} l_{Cg}(x)$, and $\mathcal{E}_g(x) = \frac{1}{C}\mathcal{E}_{Cg}(x)$, where Cg is the metric $Cg(v,w) := C(g(v,w))$.

If Q is a surface and $\Sigma \subset Q$ a subsurface, denote by $\mathrm{area}_g(\Sigma) = \int_\Sigma \sigma_g$ the *area* of Σ, where σ_g is the Riemannian volume form of g.

We use the following notations for various versions of loop spaces:

$$\begin{aligned}
\mathcal{L}Q &:= \{\gamma : S^1 \to Q \,|\, \gamma \text{ smooth}\},\\
\mathcal{L}Q_g^{<(\leqslant)a} &:= \{\gamma \in \mathcal{L}Q \,|\, \mathcal{E}_g(\gamma) < (\leqslant) a\},\\
\mathcal{L}_\alpha Q &:= \{\gamma \in \mathcal{L}Q \,|\, [\gamma] = \alpha\},\\
\mathcal{L}_\alpha Q_g^{<(\leqslant)a} &:= \{\gamma \in \mathcal{L}_\alpha Q \,|\, \mathcal{E}_g(\gamma) < (\leqslant) a\}.
\end{aligned}$$

Robustness. We are interested in conditions under which the entropy of the geodesic flow is robust under C^0-perturbations of the metric. Let us formulate the robustness property in which we are interested. Let Q be closed manifold, equipped with a metric g.

Definition 16. Let $\varepsilon > 0$. We say that g has ε-robust h_{top} if there is $c > 0$ such that for all metrics g' with $d_{\mathrm{RBM}}(g, g') < \varepsilon$ we have $h_{\mathrm{top}}(g') \geqslant c$. We say that g has robust h_{top} if it has ε-robust h_{top} for some $\varepsilon > 0$.

Remark 17. It seems that there is no established terminology for this property in the literature. In other places, g is said to have *stable* h_{top} or to be *entropy non-collapsing*.

A preliminary robustness lemma

We describe the (classic) mechanism that deduces positive topological entropy from many homotopically different periodic orbits.

Two free homotopy classes α, β are coprime if they are not multiples of a common class γ. Equivalently, α and β do not possess representatives a and b such that multiples na and mb are homotopic.

Lemma 18. *Let S_g be a compact Riemannian manifold. Let $\mathcal{P}_g$ be a set of periodic $g-$geodesics in pairwise coprime homotopy classes and let $\{\mathcal{P}_g^L\}_{L\in\mathbb{R}}$ be the filtration by g-length.*

Suppose that $\Gamma_L(\#\mathcal{P}_g^L) \geqslant \gamma$. Then, $h_{\mathrm{top}}(g) \geqslant \gamma$.

Proof. This follows from [2, Theorem 1]. It can also be obtained using the argument used to prove Manning's inequality in [27]. □

2. A robust version of the Denvir–MacKay theorem

The aim of this section is to prove Theorem 5. This builds upon [18], where it is shown that the existence of a contractible simple closed geodesic implies the existence of many other geodesics. We first prove in Sect. 2.1 item (1) of Theorem 5 in case γ_0 is topologically non-degenerate. For the notion of topological non-degeneracy see Definition 48. In Sect. 2.3, we provide an explicit robustness constant in terms of the area and of the minimal length of a contractible closed geodesic that is item (2) in Theorem 5.

2.1. Proof of item (1) of Theorem 5 in case γ_0 is non-degenerate

The strategy is to combine the robustness of contractible closed geodesics with the proof that a contractible closed orbit implies positive topological entropy.

Before presenting the proof, we recall some preliminary notions which will be needed in the proof. We fix, once and for all, a covering map $\Pi : \mathbb{R}^2 \to T^2$, such that the group G of deck transformations associated to Π is the group of translations of $\mathbb{R}^2$ by vectors (m, n), where m and n are integers. In other words

$$G = \{\mathcal{T}_{(m,n)} \mid m \text{ and } n \in \mathbb{Z}\},$$

where $\mathcal{T}_{(m,n)} : \mathbb{R}^2 \to \mathbb{R}^2$ is given by $\mathcal{T}_{(m,n)}(x, y) = (x+m, y+n)$. It is clear that for any choice of real numbers a and b the unit square $[a, a+1] \times [b, b+1] \subset \mathbb{R}^2$ is a fundamental domain for the covering Π.

Let σ be an immersed contractible closed curve in T^2 with only transverse self-intersections. A *lift* of σ is an immersed closed curve $\widetilde{\sigma}$ in $\mathbb{R}^2$ such that for any parametrization $\mathfrak{f} : S^1 \to \widetilde{\sigma}$ of $\widetilde{\sigma}$ the composition $\Pi \circ \mathfrak{f}$ is a parametrization of σ.

We will need the following elementary lemma.

Lemma 19. *Let g be a Riemannian metric on T^2 and fix a number $l > 0$. Then, if σ is an immersed closed curve in T^2 with only transverse self-intersections and g-length $< l$, every lift $\widetilde{\sigma}$ of σ is contained in a square of the form $[a, a + \lceil \frac{\sqrt{D(g)}l}{2} \rceil] \times [b, b + \lceil \frac{\sqrt{D(g)}l}{2} \rceil]$.*

Recall that we defined $D(g) = e^{d_{C^0}(g_{\text{flat}}, g)}$.

Proof. Since the g-length of σ is $< l$, the g_{flat}-length of σ is $< \sqrt{D(g)}l$, which implies that any lift $\widetilde{\sigma}$ of σ has length $< \sqrt{D(g)}l$ with respect to the flat metric in $\mathbb{R}^2$. Let p_{left} be a leftmost point of $\widetilde{\sigma}$ and p_{right} be a rightmost point of $\widetilde{\sigma}$. Since σ has length $< \sqrt{D(g)}l$ and has to travel from p_{left} to p_{right} and back to p_{left} by running a distance less than its length, we conclude that the Euclidean distance between p_{left} to p_{right} is $< \frac{\sqrt{D(g)}l}{2}$. We conclude that $\widetilde{\sigma}$ is contained in a vertical strip of $\mathbb{R}^2$ with width $< \frac{\sqrt{D(g)}l}{2}$.

Reasoning similarly for an uppermost point and a lowermost point of $\widetilde{\sigma}$ we conclude that $\widetilde{\sigma}$ must be contained in a horizontal strip of width $< \frac{\sqrt{D(g)}l}{2}$. This establishes the lemma. $\square$

We are now ready to proceed with the proof.

Step 1: Robustness of the length spectrum. If γ_0 is a contractible topologically non-degenerate closed geodesic of g_0, then for a sufficiently small $\delta > 0$, there is an $\epsilon > 0$ such that every Riemannian metric g whose d_{C^0}-distance to g_0 is $< \epsilon$ has a contractible closed geodesic γ with $|l_g(\gamma) - l_{g_0}(\gamma_0)| < \delta$ and $|l_{g_0}(\gamma) - l_{g_0}(\gamma_0)| < \delta$. Since $D(g)$ also varies continuously with respect to d_{C^0}, we can choose $\epsilon > 0$ to be small enough so that $D(g) < D(g_0) + \delta$. We believe this to be well known to experts, but since we are unaware of a published proof of this precise statement, we provide a proof using very low-tech methods in Appendix A.

Step 2: Simplifying the geodesic. We start with the contractible closed geodesic γ for g from Step 1. Denote by $\tilde{\gamma}$ a lift to the universal cover $\mathbb{R}^2$, which is a closed geodesic of $\widetilde{g} := \Pi^* g$ because γ is contractible. Denote by S the closure of the unbounded component of $\mathbb{R}^2 \setminus \tilde{\gamma}$, a set that is homeomorphic to a plane minus an open disk. Let α be one of the two non-trivial free homotopy classes of loops in S which contains embedded loops; these are the classes of curves in S which encircle once the disk which was removed from $\mathbb{R}^2$. Using the curve shortening flow on S and reasoning as in the proof of Lemma 2 of [6] one obtains a contractible simple closed geodesic γ_α of $\widetilde{g}$ in the homotopy class α, whose g-length is the minimal possible length for all curves in the homotopy class α.

Combining this discussion with Lemma 19, we conclude that $\widetilde{\gamma}$ is contained in a square of the form $[a, a + \lceil \frac{\sqrt{D(g)} l_{g(\gamma)}}{2} \rceil] \times [b, b + \lceil \frac{\sqrt{D(g)} l_{g(\gamma)}}{2} \rceil]$. For simplicity, we let $N := \lceil \frac{\sqrt{D(g)} l_{g(\gamma)}}{2} \rceil$.

Let $\widehat{G}$ be the subgroup of G generated by $\mathcal{T}_{N,0}$ and $\mathcal{T}_{0,N}$. Since γ_α is a simple closed curve in $\mathbb{R}^2$ contained in a fundamental domain $[a, a+N] \times [b, b+N]$ of $\widehat{G}$, it is a simple contractible closed geodesic in the quotient Riemannian manifold $(\widehat{T}, \widehat{g})$ that is obtained by quotienting $(\mathbb{R}^2, \widetilde{g} := \Pi^* g)$ by the action of $\widehat{G}$.

Step 3: Many free homotopy classes. Since $\widetilde{\gamma}$ is a simple contractible closed geodesic in $\widehat{T}$, it bounds a disk $\widehat{D}$ in $\widehat{T}$. The surface $S := \widehat{T} \backslash \widehat{D}$ is diffeomorphic to the torus minus a disk, so that $\pi_1 S$ is the free group with two generators. Let α_1 be the projection of $[a, a+N] \times \{b\}$ to $\widehat{T}$ and α_2 be the projection of $\{a\} \times [b, b+N]$ to $\widehat{T}$: $\{\alpha_1, \alpha_2\}$ is a basis of $\pi_1 S$.

We proceed to estimate the $\widehat{g}$-length of α_1 and α_2. It is clear that

$$l_{\widehat{g}}(\alpha_1) = N l_{\widetilde{g}}([a, a+1] \times \{b\}) \quad \text{and} \quad l_{\widehat{g}}(\alpha_2) = N l_{\widetilde{g}}(\{a\} \times [b, b+1]).$$

We know that $\max\{l_{\widetilde{g}}(\{a\} \times [b, b+1]), l_{\widetilde{g}}([a, a+1] \times \{b\})\} \leqslant \sqrt{D(g)}$.

Recall that the set of free homotopy classes of loops $\Omega(S)$ in S equals the set of conjugacy classes of $\pi_1 S$. Given a number n we denote by $\Omega_n(S)$ the set of elements in $\Omega(S)$ which have at least one representative in $\pi_1 S$ with word length $\leqslant n$ with respect to $\{\alpha_1, \alpha_2, \alpha_1^{-1}, \alpha_2^{-1}\}$. It is well known, see, e.g., [17, VI.A.], that the number of elements in $\pi_1 S$ of minimal word length $n > 0$ is $4 \cdot 3^{n-1}$. Since conjugation on cyclically reduced words in a free group corresponds to a cyclic permutation, one obtains $\#\Omega_n(S) \geqslant \frac{8 \cdot 3^{n-2}}{n}$.

Let $\rho \in \Omega_n(S)$. Because we can represent ρ as a word in α_1 and α_2 with word length $\leqslant n$, we can find curves in ρ whose length is $\leqslant N e^{D(g)} n$. Since $(S, \widehat{g})$ is a Riemannian surface with geodesic boundary, there exists a minimizing closed geodesic γ_ρ of $(S, \widehat{g})$ in ρ contained in the interior of S. The length of γ_ρ is $\leqslant N\sqrt{D(g)} n$

Step 4: Many geodesics lead to entropy. Let $\mathcal{N}^C(\widehat{g})$ be the set of prime minimizing closed geodesics of length $\leqslant C$ in $(S, \widehat{g})$ and $\widetilde{\mathcal{N}}^C(\widehat{g})$ be the set of minimizing closed geodesics of length $\leqslant C$ in $(S, \widehat{g})$. A simple argument shows that the exponential growths $\limsup_{C\to+\infty}\frac{\log(\#\mathcal{N}^C(\widehat{g}))}{C}$ of $\mathcal{N}^C(\widehat{g})$ and $\limsup_{C\to+\infty}\frac{\log(\#\widetilde{\mathcal{N}}^C(\widehat{g}))}{C}$ of $\widetilde{\mathcal{N}}^C(\widehat{g})$ are the same.

But

$$\limsup_{C\to+\infty}\frac{\log(\#\widetilde{\mathcal{N}}^C(\widehat{g}))}{C}\geqslant\frac{1}{N\sqrt{D(g)}}\limsup_{n\to+\infty}\frac{\log(\#\Omega_n(S))}{n}\geqslant\frac{\log 3}{N\sqrt{D(g)}}.$$

Let $\varphi_{\widehat{g}}$ be the geodesic flow of $\widehat{g}$ on $T^1\widehat{T}$. We consider the set of globally minimizing geodesics of $(S,\widehat{g})$. The lift $\mathcal{C}$ of this set is a compact $\varphi_{\widehat{g}}$-invariant set that is completely contained in T^1S. The lift of every minimizing closed geodesic of $(S,\widehat{g})$ belongs to $\mathcal{C}$. Reasoning as in the proof of Manning's inequality one obtains that the exponential growth of $\mathcal{N}^C(\widehat{g})$ is a lower bound for $h_{\text{top}}(\varphi_{\widehat{g}}|_{\mathcal{C}})$. Noting that two different prime closed geodesics of $(S,\widehat{g})$ give closed trajectories of $\varphi_{\widehat{g}}$ in $\mathcal{C}$ which are not homotopic in T^1S, one can also obtain that the exponential growth of $\mathcal{N}^C(\widehat{g})$ is a lower bound for $h_{\text{top}}(\varphi_{\widehat{g}}|_{\mathcal{C}})$ by using Theorem 1 of [2]. We can, thus, conclude that

$$h_{\text{top}}(\varphi_{\widehat{g}})\geqslant h_{\text{top}}(\varphi_{\widehat{g}}|_{\mathcal{C}})\geqslant\frac{\log 3}{N\sqrt{D(g)}}.$$

Since $h_{\text{top}}(\varphi_{\widehat{g}})=h_{\text{top}}(\varphi_g)$ we have shown that for every Riemannian metric g with $d_{C_0}(g,g_0)<\epsilon$ we have $h_{\text{top}}(\varphi_g)\geqslant\frac{\log 3}{N\sqrt{D(g)}}$, which completes the proof of the theorem. □

2.2. Proof of item (1) of Theorem 5 in case γ_0 is degenerate

In the degenerate case, we face the problem that the contractible closed geodesic may not persist after perturbation of the metric. It is easy to construct examples where it does not. However, in the following proof, we are able to show that there are different contractible geodesics that will persist. The proof follows the scheme of the non-degenerate case.

Proof. **Step 1: Simplifying the geodesic.** First, it is sufficient to consider the case where the lift $\widetilde{\gamma_0}$ to $\mathbb{R}^2$ is a simple contractible closed curve. This follows from [6] who proved that any Riemannian metric on T^2 with a contractible closed geodesic α with length l has a contractible closed geodesic α' whose length is $\leqslant l$ and whose lift to $\mathbb{R}^2$ is simple. For completeness, we sketch the proof of this fact. If a lift $\widetilde{\alpha}$ of α is simple there is nothing to be done. Otherwise, letting Σ be the closure of the unbounded component of $\mathbb{R}^2\backslash\widetilde{\alpha}$, we know that $\partial\Sigma$ is a geodesic polygon formed by pieces of $\widetilde{\alpha}$, and the length of $\partial\Sigma$ is $<l$. Perturbing $\partial\Sigma$ to a curve completely contained in the interior of Σ and applying the curve shortening flow one obtains the desired α'.

Reasoning as in the proof of the non-degenerate case, we obtain a square of the form $[a,a+\lceil\frac{\sqrt{D(g_0)}}{2}\rceil]\times[b,b+\lceil\frac{\sqrt{D(g_0)}}{2}\rceil]$. We let $(\widehat{T}^2,\widehat{g}_0)$ be the quotient of $(\mathbb{R}^2,\pi^*g_0)$ by the group of translations generated $\mathcal{T}_{0,M}$ and $\mathcal{T}_{M,0}$, where $M:=\lceil\frac{\sqrt{D(g_0)}}{2}\rceil$.

We denote by S the non-contractible component of $\widehat{T}^2\backslash\mathrm{im}(\gamma_0)$ and by D its complement. Note that $\partial D\subseteq\mathrm{im}(\gamma_0)$.

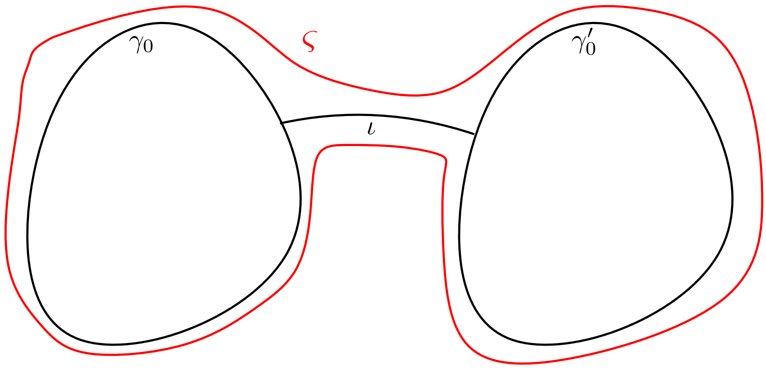


Figure 1. The construction of the curve ς

Step 2: Finding the geodesic.

Fix a lift $\widetilde{\gamma_0}$ and let $\widetilde{\gamma_0}'$ be the closest lift of γ_0 which does not intersect $\widetilde{\gamma_0}$. The distance between $\widetilde{\gamma_0}$ and $\widetilde{\gamma_0}'$ is $\leqslant \sqrt{D(g)}$: this is an elementary exercise that we leave to the reader. We let ι be the shortest geodesic of $\pi^* g_0$ connecting $\widetilde{\gamma_0}$ and $\widetilde{\gamma_0}'$. Let then ς be a simple closed curve in $\mathbb{R}^2$ contained in a small tubular neighborhood of $\widetilde{\gamma_0} \cup \widetilde{\gamma_0}' \cup \iota$, and such that the bounded component of $\mathbb{R}^2 \backslash \varsigma$ contains both $\widetilde{\gamma_0}$ and $\widetilde{\gamma_0}'$; see Fig. 1.

We denote by ρ the free homotopy class of loops in S which contain the projection of ς to S. The class ρ has the following properties:

(1) curves in ρ are not contractible in S,
(2) curves in ρ are contractible in T^2,
(3) curves in ρ are not homotopic to ∂S in S.

Given any $\varepsilon > 0$, there are curves in ρ with length $\leqslant 2(l_{g_0}(\gamma_0) + \sqrt{D(g_0)}) + \varepsilon$. This is obtained by considering curves homotopic to ς and contained in sufficiently small neighborhoods of $\widetilde{\gamma_0} \cup \widetilde{\gamma_0}' \cup \iota$; see Fig. 1.

We define

$$\sigma_\rho := \inf_{\gamma \in \rho} \mathcal{E}_{g_0}(\gamma).$$

By the remark above we know that $\sigma_\rho \leqslant (l_{g_0}(\gamma_0) + \sqrt{D(g_0)})^2$. The infimum σ_ρ is actually a minimum, and this can be proved using piecewise smooth geodesics and the strategy to prove [24, Theorem 1.5.1]. Let $\mathcal{A}_\rho$ be the set of curves in S whose g_0-energy is σ_ρ. They are all smooth geodesics of g_0.

Step 3: Repelling boundary. We show that energy minimizing curves have to be contained in the interior of S.

Lemma 20. *Given ρ as in the last step, there is $\delta_\rho > 0$ such that every loop in ρ that intersects ∂S has energy $> \sigma_\rho + \delta_\rho$.*

Proof. If this is false, then we find a sequence τ_m of C^∞ loops in ρ that touch ∂S with $\mathcal{E}_{g_0}(\tau_m) \to \sigma_\rho$. Then, it is possible to find a natural number K and replace τ_m by a sequence of piecewise smooth geodesics $\widehat{\tau}_m$ in ρ with at most K corners and satisfying $\mathcal{E}_{g_0}(\widehat{\tau}_m) \to \sigma_\rho$. [4] Then, $\widehat{\tau}_m$ converges up to a subsequence to a piecewise smooth geodesic τ that touches the boundary and that has energy $\mathcal{E}_{g_0}(\tau) = \sigma_\rho$. Because $\mathcal{E}_{g_0}(\tau) = \sigma_\rho$ it must be a smooth geodesic and since it is contained in S and touches ∂S it must be tangent to ∂S. This implies that τ is a geodesic tangent to the geodesic ∂S and thus, τ coincides with ∂S. But this is impossible because curves in ρ are not homotopic to the boundary. □

Step 4: Perturbing the metric. Consider now the open set $V_{\delta_\rho/2}$ of loops contained in S and belonging to ρ with energy in the interval $[\sigma_\rho, \sigma_\rho + \delta_\rho/2)$. From Step 3, all loops in $\overline{V_{\delta_\rho/2}}$ have image in the interior of S.

Denote by U_ε the set of Riemannian metrics g satisfying $(1-\varepsilon)g_0 < g < (1+\varepsilon)g_0$. For $g \in U_\varepsilon$ and any $\tau' \in W^{1,2}(S^1, T^2)$ we have

$$(1-\varepsilon)\mathcal{E}_{g_0}(\tau) < \mathcal{E}_g(\tau') < (1+\varepsilon)\mathcal{E}_{g_0}(\tau).$$

Choose $\varepsilon > 0$ small enough such that

$$(1+\varepsilon)\sigma_\rho < (1-\varepsilon)(\sigma_\rho + \frac{\delta_\rho}{2}). \tag{2.1}$$

Then, we can show that any $g \in U_\varepsilon$ has a contractible closed geodesic which belongs to $\overline{V_{\delta_\rho/2}}$. For this we reason as follows:

Let γ_ρ be a geodesic of g_0 in the class ρ and with $\mathcal{E}_{g_0}(\gamma_\rho) = \sigma_\rho$. The curve γ_ρ is in the interior of $\overline{V_{\delta_\rho/2}}$. We have

$$\mathcal{E}_g(\gamma_\rho) < (1+\varepsilon)\sigma_\rho.$$

Let ϕ_g be the negative gradient flow of $\mathcal{E}_g$. Then, $\phi_g^t(\gamma_\rho) \in \overline{V}_{\delta_\rho/2}$ for all $t \geqslant 0$ because $\mathcal{E}_g$ decreases along trajectories of ϕ_g and because for all $\tau' \in \partial\overline{V}_{\delta_\rho/2}$ we have

$$\mathcal{E}_g(\tau') \geqslant (1-\varepsilon)\mathcal{E}_{g_0}(\tau') = (1-\varepsilon)(\sigma_\rho + \delta_\rho/2) \geqslant (1+\varepsilon)\sigma_\rho > \mathcal{E}_g(\gamma_\rho).$$

It follows that $\phi_g^t(\gamma_\rho)$ cannot cross the boundary of $\overline{V_{\delta_\rho/2}}$ and is, thus, trapped in $\overline{V_{\delta_\rho/2}}$. By Palais–Smale $\phi_g^t(\gamma_\rho)$ converges to a geodesic γ of g which must then be in $\overline{V_{\delta_\rho/2}}$. The geodesic γ has image in the interior of S and belongs to ρ.

Step 5: Uniform lower bound on the entropy. To obtain a uniform lower bound on the topological entropy of metrics in U_ε one uses an argument with finite covers identical to the one used in the proof of the non-degenerate case. The crucial point is that the Riemannian metric g has a contractible closed geodesic whose length is very close to $\sigma_\rho \leqslant 2(l_{g_0}(\gamma_0) + \sqrt{D(g_0)})$. □

2.3. A lower bound on topological entropy in terms of surface area

The lower bound on the topological entropy obtained in the proofs of the first part of Theorem 5 depends on the number of fundamental domains a lift of a contractible closed geodesic in the torus (T^2, g) of a fixed length may intersect. A metric invariant that often appears naturally in the investigation of lower bounds on entropy is the surface area. In this section, we observe that

[4] To guarantee that we can find a uniform bound on the number of corners one uses that (S, g_0) has geodesic boundary and that its convexity radius is therefore positive.

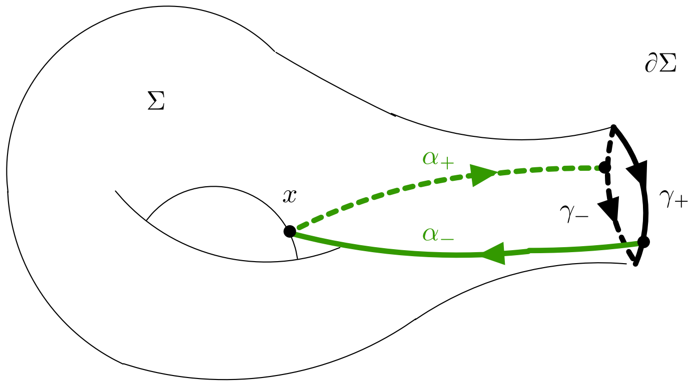


Figure 2. Situation in Sect. 2.3

there is a lower bound depending only on the surface area and the minimal length of contractible closed geodesics on (T^2, g_0). Since surface area C^0 continuously depends on the metrics on T^2 this gives an alternative robust bound.

The main additional idea is to first find for a genus 1 surface (Σ, g) with one (geodesic) boundary component γ two short curves u_+ and u_- that generate exponential growth rate $\log 2$ in the fundamental group of Σ, where the length of u_+ and u_- are bounded in terms of the length of γ and the area of (Σ, g).

Let Σ be a surface homeomorphic to a torus with connected boundary. We endow Σ with a Riemannian metric g such that the boundary curve is a geodesic. This is in particular the situation given by a closed geodesic on a torus, of which we remove the simply connected component of the complement of the geodesic. We fix an orientation of Σ, which induces an orientation on $\partial\Sigma$. Denote by L the length of $\partial\Sigma$ and by A the area of (Σ, g).

We consider the following set of curves:

$$\begin{aligned} B = \{\beta : I \to \Sigma \text{ with } \beta(\partial I) \subset \partial\Sigma \text{ and} \\ [\beta] \text{ non-trivial and not homotopic to } \partial\Sigma\} \end{aligned}$$

and the quantity

$$d = \inf \{l_g(\beta)\ ,\ \beta \in B\}\,.$$

The infimum is attained. Let α be a curve in B such that $l_g(\alpha) = d$. Let $x \in \Sigma$ be the point of α that divides it in two paths of equal length $d/2$, α_- the restriction of α to the path between $\alpha(0)$ and x; α_+ the restriction of α to the path between x and $\alpha(1)$.

We denote also by γ_- and γ_+ the paths in $\partial\Sigma$ (disjoint up to the endpoints) with $\gamma_-(0) = \gamma_+(0) = \alpha(1)$ and $\gamma_-(1) = \gamma_+(1) = \alpha(0)$ (one of which is constant if the endpoints of $\alpha(0) = \alpha(1)$ are the same). We choose these path such that γ_+ follows $\partial\Sigma$ in positive orientation and γ_- follows $\partial\Sigma$ in negative orientation. See Fig. 2.

We consider the two homotopy classes $\xi_\pm \in \pi_1(\Sigma, x)$:

$$\xi_\pm = [\alpha_+ \cdot \gamma_\pm \cdot \alpha_-] .$$

We need to argue that, first, we can find representatives of $\xi_\pm$ of small lengths and, second, that $\xi_\pm$ generate exponential growth in $\pi_1(\Sigma, x)$.

Representatives of small length. The strategy is to consider a family of closed curves such that each one together with $\partial\Sigma$ bounds an annulus in Σ. Integration on their lengths will show that there is an upper bound on the length of smallest representatives of $\xi_\pm$ in terms of A and L.

Lemma 21. *There exist representatives of $\xi_\pm$, both of length smaller than $M := \sqrt{L^2 + 4A}$.*

Proof. For $t < \frac{d}{2}$, we introduce the sets

$$\mathcal{G}^t = \{x \in \Sigma; \operatorname{dist}_g(x, \partial\Sigma) = t\} .$$

It follows from a classical result of Hartman [22], see also [38, Theorem 4.4.1], that there is a closed set of zero Lebesgues measure of exceptional parameters $E \subset [0, d/2]$ such that for all $t \in R := [0, d/2] \backslash E$, $\mathcal{G}^t$ consists of finitely many connected components, each being the image of a piecewise smooth closed simple path. In particular, the length $l_g(\mathcal{G}^t)$ of $\mathcal{G}^t$ is well defined for $t \in R$. Moreover, if we denote $\mathcal{B}^t := \{x \in \Sigma; \operatorname{dist}_g(x, \partial\Sigma) \leqslant t\}$, which is a subsurface in Σ bounded by $\partial\Sigma$ and $\mathcal{G}^t$, then for $t \in R$, $\frac{\mathrm{d}}{\mathrm{d}t}\operatorname{area}_g(\mathcal{B}^t) = l_g(\mathcal{G}^t)$.

It is straightforward to see that α is not contained in $\mathcal{B}^t$. Note also that, since $t < d/2$, $\mathcal{B}^t$ has genus zero. Take the collection of discs in Σ that any component of $\mathcal{G}^t$ may bound in the complement of $\mathcal{B}^t$, and attach it to $\mathcal{B}^t$. We denoted this surface by $\widetilde{\mathcal{B}^t}$.

Claim 22. $\widetilde{\mathcal{B}^t}$ is an annulus that is bounded by $\partial\Sigma$ and by one distinguished connected component γ^t of $\mathcal{G}^t$.

Proof. Assume by contradiction that the boundary of $\widetilde{\mathcal{B}^t}$ has more than one connected component besides $\partial\Sigma$. Then, there exists a path β, completely contained in $\mathcal{B}^t$, with endpoints in $\partial\Sigma$, and which is not homotopic in Σ relative to $\partial\Sigma$ to a path in $\partial\Sigma$. Let l_0 be the infimum of the lengths of such paths. We claim that $l_0 < d$, which contradicts the definition of d. For this, choose β as above to have length $l_0 \leqslant l_g(\beta) < l_0 + (d - 2t)$. There is a point y on β that divides β into two subpaths β_0 and β_1 of equal length $l_g(\beta)/2$. By the definition of $\mathcal{B}^t$, there is a path $\hat{\beta}$ from y to $\partial\Sigma$ in $\mathcal{B}^t$ of length $\leqslant t$. Either the concatenation of $\beta_0 \cdot \hat{\beta}$ or $\overline{\beta_1} \cdot \hat{\beta}$ is not homotopic relative $\partial\Sigma$ to a path in $\partial\Sigma$. Hence, $l_0 \leqslant l_g(\beta)/2 + t < l_0/2 + d/2$, and hence $l_0 < d$. □

Via the annulus that γ^t and $\partial\Sigma$ bound we choose the orientation of γ^t to be parallel to that of $\partial\Sigma$. For each $t \in R$, $t < \frac{d}{2}$, both paths, α_+ and α_-, intersect γ^t in some point, say z_+^t and z_-^t, respectively. Let α_-^t be the subpath of α_- from z_-^t to x, and α_+^t be the subpath of α_+ from x to z_+^t. Let γ_+^t and γ_-^t be two path on γ^t (disjoint up to the endpoints) with $\gamma_+^t(0) = \gamma_-^t(0) = z_+^t$ and $\gamma_+^t(1) = \gamma_-^t(1) = z_-^t$. We choose these path such that γ_+^t follows γ^t in positive orientation and γ_-^t follows γ^t in negative orientation. Consider the loops $u^t_\pm = \alpha^t_+ \cdot \gamma^t_\pm \cdot \alpha^t_-$. By the choice of orientations we have that $[u^t_\pm] = \xi_\pm \in \pi_1(\Sigma, x)$. The point $z^t_\pm \in \gamma^t$ are at distance t from $\partial\Sigma$, by the definition of γ^t. Hence, the length of $u^t_\pm$ satisfy

$$l_g(u^t_\pm) \leqslant d - 2t + l_g(\gamma^t).$$

We will show that there is $t \in R$ such that the length of both $u^t_\pm$ is $< M$. (Note here, that also $0 \in R$, and $l_g(u^0_\pm) < L+d$). Assume the contrary. This means that for all $t \in R$,

$$l_g(\gamma^t) \geqslant M + 2t - d.$$

Now, for any $0 \leqslant \sigma \leqslant d/2$,

$$\begin{aligned} A &\geqslant \mathrm{area}_g(\mathcal{B}^{d/2}) - \mathrm{area}_g(\mathcal{B}^{d/2-\sigma}) \geqslant \int_{d/2-\sigma}^{d/2} l_g(\gamma^t)\,\mathrm{d}t \\ &\geqslant \int_{d/2-\sigma}^{d/2} M + 2t - d\,\mathrm{d}t = \int_0^\sigma M - 2s\,\mathrm{d}s = M\sigma - \sigma^2, \end{aligned} \tag{2.2}$$

hence for all $0 \leqslant \sigma \leqslant d/2$,

$$0 \leqslant \sigma^2 - M\sigma + A = \left(\sigma - \frac{1}{2}(M - \sqrt{M^2 - 4A})\right)\left(\sigma - \frac{1}{2}(M + \sqrt{M^2 - 4A})\right),$$

which means that

$$M - \sqrt{M^2 - 4A} \geqslant d,$$

so $M - L \geqslant d$, and therefore $M \geqslant d + L > l_g(u^0_\pm)$, a contradiction to our assumption. □

Generating growth. With Lemma 21 we obtain

Lemma 23. *The number of free homotopy classes of loops in Σ that have a representative of length $\leqslant Mn$ is $\geqslant \frac{2^n-2}{n}$.*

Proof. $\pi_1(\Sigma, x)$ is a free group of two generators. It is straightforward to check, e.g., with the help of the loops α and β of Fig. 3 below, that we can choose two elements $a, b \in \pi_1(\Sigma, x)$ that freely generate $\pi_1(\Sigma, x)$ such that $\xi_- = a$ and $\xi_+ = b^{-1}ab$. For a given word w of length n in the letters ξ_+ and ξ_-, consider the word $\widetilde{w}$ in the letters a and b that we obtain when expressing ξ_- as a and ξ_+ as $b^{-1}ab$ and then reducing cyclically. It is straightforward to check that, if $w_1 \neq w_2$ are such words that are not of the form ξ_+^n or ξ_-^n, then $\widetilde{w}_1 \neq \widetilde{w}_2$. The homotopy classes of free loops in Σ correspond to conjugacy classes of elements in $\pi_1(\Sigma, x)$. The latter correspond to words in a and b up to cyclic reduction and cyclic permutation. Hence by Lemma 21 and the above considerations, the number of homotopy classes of loops in Σ that have a representative of length $\leqslant Mn$ is at least $\frac{2^n-2}{n}$.

With Lemmas 21 and 23 at hand, we prove the second part of Theorem 5.

Proof. Assume first that g_0 is bumpy. Then, as above, there is for all $\varepsilon > 0$ some $\delta > 0$ such that for all g with $d_{C^0}(g, g_0) < \delta$ there is a contractible closed geodesic γ for g with length $l_g(\gamma) < l_{g_0}(\gamma_0)+\varepsilon$. Furthermore, by a sufficiently small δ, one can additionally assume that $\mathrm{area}_g(T^2) < \mathrm{area}_{g_0}(T^2)+\varepsilon$. Therefore, in this case it is enough to prove the lower bound for $h_{\mathrm{top}}(\varphi_{g_0})$.

Denote by G the group of deck transformations for the covering $\Pi : \widetilde{T^2} \to T^2$, and choose a lift $\widetilde{\gamma}_0 : S^1 \to \widetilde{T^2}$ of γ_0.

We distinguish two cases:

(1) For all $\mathcal{T} \in G\backslash\{\mathrm{id}\}$, $\mathcal{T}(\mathrm{im}(\widetilde{\gamma}_0)) \cap \mathrm{im}(\widetilde{\gamma}_0) = \emptyset$,
(2) There is $\mathcal{T} \in G\backslash\{\mathrm{id}\}$, $\mathcal{T}(\mathrm{im}(\widetilde{\gamma}_0)) \cap \mathrm{im}(\widetilde{\gamma}_0) \neq \emptyset$.

If case (1) holds, then, as in Step 2 of the proof presented in Sect. 2.1 for item (1) of Theorem 5, we know that there is actually a contractible simple closed geodesic γ for g_0 with length $l_{g_0}(\gamma) \leqslant l_{g_0}(\gamma_0)$. Denote $D \subset T^2$ the disc that is bounded by γ. Then, by Lemma 23 applied to $\Sigma = T^2\backslash D$, and Lemma 18, we obtain that

$$h_{\mathrm{top}}(\varphi_{g_0}) > \frac{1}{\sqrt{4\mathrm{area}_{g_0}(\Sigma) + l_{g_0}(\gamma)^2}} \log 2 > \frac{1}{\sqrt{4A + l_{g_0}(\gamma_0)^2}} \log 2.$$

In case (2), fix $\mathcal{T} \in G\backslash\{\mathrm{id}\}$ with $\mathcal{T}(\mathrm{im}(\widetilde{\gamma}_0)) \cap \mathrm{im}(\widetilde{\gamma}_0) \neq \emptyset$, and let $k \in \mathbb{N}$, $k \geqslant 2$, with $k = \min\{l \in \mathbb{N} \,|\, \mathcal{T}^l(\mathrm{im}(\widetilde{\gamma}_0)) \cap \mathrm{im}(\widetilde{\gamma}_0) = \emptyset\}$. Consider the lifts $\widetilde{\gamma}_1 = \mathcal{T} \circ \widetilde{\gamma}_0$ and $\widetilde{\gamma}_k = \mathcal{T}^k \circ \widetilde{\gamma}_0$. Then, $\widetilde{\gamma}_1$ intersects both $\widetilde{\gamma}_0$ and $\widetilde{\gamma}_k$, hence $\mathrm{dist}_{\Pi^* g_0}(\mathrm{im}(\widetilde{\gamma}_0), \mathrm{im}(\widetilde{\gamma}_k)) < l_{g_0}(\gamma_0)/2$.

Choose $\mathcal{S} \in G$ such that for all $l \in \mathbb{Z}\backslash\{0\}$, and all $m \in \mathbb{Z}$, $\mathcal{S}^l(\mathrm{im}(\widetilde{\gamma}_0)) \cap \mathcal{T}^m(\mathrm{im}(\widetilde{\gamma}_0)) = \emptyset$. Denote $\widehat{T^2}$ the quotient of $\widetilde{T^2}$ by the action of the subgroup of G generated by $\mathcal{S}$ and $\mathcal{T}^k$, and $\widehat{\Pi} : \widehat{T^2} \to T^2$ the induced covering map. By the choice of $\mathcal{S}$, $\mathcal{T}$ and k, the projected curve $\widehat{\gamma}_0$ of $\widetilde{\gamma}_0$ to $\widehat{T^2}$ is a closed geodesic of $\widehat{\Pi}^* g_0$ that lies in one fundamental domain of the universal covering of $\widehat{T^2}$. So, as argued before, if $\widehat{\gamma}_0$ is not simple, we may replace it by a simple closed geodesic on $(\widehat{T^2}, \widehat{\Pi}^* g_0)$ with length $\leqslant l_{g_0}(\gamma_0)$ that encircles $\mathrm{im}(\widehat{\gamma}_0)$. Let $\widehat{D}$ be the disc bounded by $\widehat{\gamma}_0$ and let $\widehat{\Sigma} = \widehat{T^2}\backslash\widehat{D}$. With analogous notation as before,

$$\widehat{B} := \Big\{ \beta : I \to \Sigma \text{ with } \partial I \subset \partial\widehat{\Sigma} \text{ and } \\ [\beta] \text{ non-trivial and not homotopic to } \partial\widehat{\Sigma} \Big\},$$

one has $\widehat{d} := \inf \left\{ l_{\widehat{\Pi}^* g_0}(\beta) \,,\ \beta \in \widehat{B} \right\} \leqslant \mathrm{dist}_{\Pi^* g_0}(\mathrm{im}(\widetilde{\gamma}_0), \mathrm{im}(\widetilde{\gamma}_k)) < l_{g_0}(\gamma_0)/2$. Then, following the construction in the proof of Lemma 21 and Lemma 23, one proves that

$$h_{\mathrm{top}}(\widehat{\varphi}_{g_0}) > \frac{1}{\frac{3}{2}L} \log 2,$$

where $\widehat{\varphi}_{g_0}$ denotes the geodesic flow on $\widehat{T^2}$ with respect to the metric $\widehat{\Pi}^* g_0$. Since $h_{\mathrm{top}}(\widehat{\varphi}_{g_0}) = h_{\mathrm{top}}(\varphi_{g_0})$, the assertion follows.

In the case that g is degenerate, one combines the argument in the proof of the degenerate case of the first part of the theorem with the estimates obtained in the previous paragraph, and observes that there is a locally energy

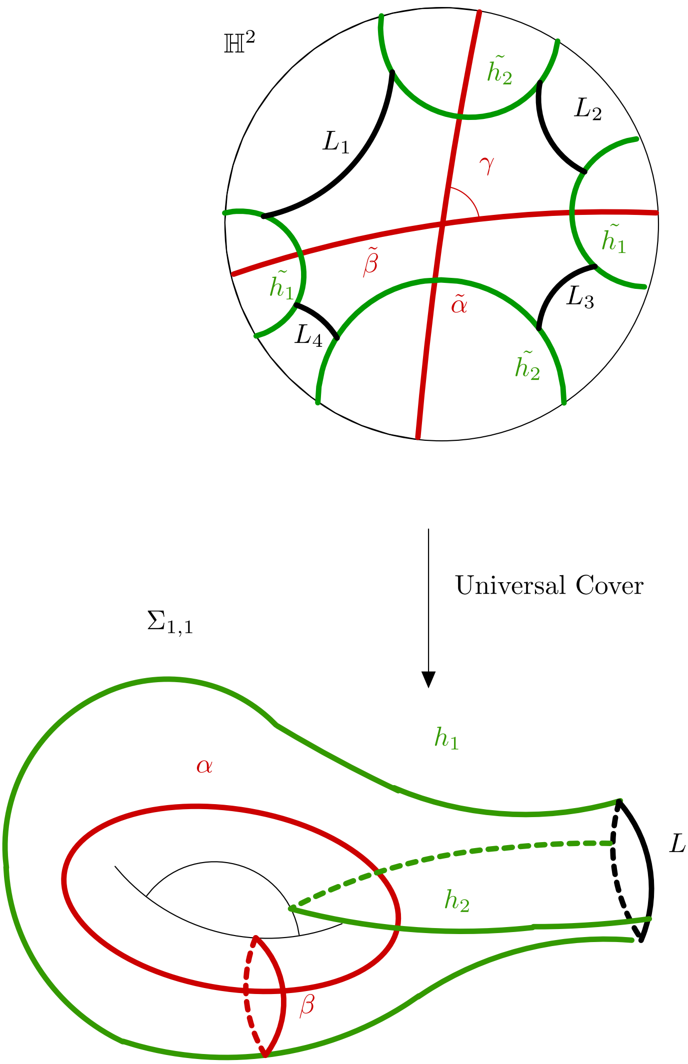


Figure 3. One-holed torus, generators, heights and boundary

minimizing contractible closed geodesic γ whose lift to the universal cover encircles two distinct lifts of γ_0 and whose length satisfies

$$l_{g_0}(\gamma) < \max\left\{4\sqrt{4\mathrm{area}_{g_0}(T^2) + l_{g_0}(\gamma_0)^2}, 3l_{g_0}(\gamma_0)\right\}.$$

The estimates for $h_{\mathrm{top}}(g)$ stated in the theorem follow as in the previous paragraph. $\square$

3. Robustness of volume entropy and hyperbolic geometry

We consider a torus with a boundary component $\Sigma_{1,1}$, endowed with a hyperbolic metric. Associated with this holed torus is its volume entropy, denoted h_{vol} (cf. Definition 15). Note that the universal cover is not all of $\mathbb{H}^2$ but only a geodesically convex subset of $\mathbb{H}^2$ bounded by the lifts of the boundary curve of length L. Hence h_{vol} is in general smaller than 1.

Both h_{vol} and L are considered as functions on the Teichmüller space of $\Sigma_{1,1}$.

We prove the following result.

Theorem 24. *If the volume entropy h tends to 0 in the Teichmüller space, then the boundary length L tends to ∞.*

In other terms, to keep the entropy bounded away from 0, we need to bound from above the boundary length. Note the analogy with Sect. 2 for the topological entropy. In this setting, the bound on the area is implicit because any hyperbolic metric on the one holed torus has area 2π by the Gauss–Bonnet formula.

Proof. We rely on Fig. 3: we choose α and β some generators of the fundamental group $\pi_1(\Sigma_{1,1}) = \mathbb{F}_2$, for which we denote by $2a$ and $2b$ the lengths of the respective geodesic realizations once we fix a hyperbolic metric on $\Sigma_{1,1}$. We also denote by γ the angle between the lifts of α and β.

We consider some *heights* associated to α and β: they are two geodesic curves starting and ending at the boundary and meeting, respectively, α and β orthogonally.

We then construct a fundamental domain for the action of π_1 on $\mathbb{H}^2$ as in Fig. 3: it is an octagon made of (pieces of) lifts of heights and boundary curves. Note that the lifts of α and β can be assumed to meet at their midpoints and that $L_1 + L_2 + L_3 + L_4 = 2L$. The fundamental domain is a union of four pentagons with one right angle and one angle γ or $\pi - \gamma$.

We may also assume that the two remaining angles in each pentagon is also a right angle. Indeed, doing so decreases the lengths of the L_i's and proving the result for four right-angled pentagons will yield the result.

For further use, assume that the remaining angle is γ in the pentagons containing as a side L_2 and L_4 and $\pi - \gamma$ in the two other pentagons.

In this case, h_{vol} and L are functions of a, b and γ. We will use two facts about these functions.

(1) The entropy $h = h_{\mathrm{vol}}$ satisfies the inequality

$$\frac{1}{1+e^{ha}} + \frac{1}{1+e^{hb}} \leqslant \frac{1}{2}.$$

This is proved in [8].

(2) The boundary length is expressed as

$$\cosh L_2 = \sinh a \sinh b - \cosh a \cosh b \cos \gamma.$$

This formula comes from a trigonometric formula in the pentagon with one side L_2, similar formulas allow to express the other L_is. This formula comes from, e.g., [12, p. 37 (iii)]. As $\gamma \to 0$, this formula implies that a and b tend to $+\infty$[5].

We now assume that $h \to 0$ and we want to see that one of the L_i's must tend to ∞. First notice that h is a continuous function of a, b and γ. So in order to make $h \to 0$, we need to make a, b and γ escaping compact sets. Note that γ cannot tend to 0 while keeping a and b bounded (unless $L_2 \to \infty$ and we are done) so a, b and γ escape compact sets if and only if a and b escape compact sets. We argue differently depending on how a and b escape compact sets. Without loss of generality, we assume that $a \leqslant b$. Note that two cases are immediately excluded because of (1): the case $a \to 0$ and $b \to 0$ and the case $a \to 0$ and b bounded.

1st case: a in bounded and $b \to \infty$. Formula (2) implies that

$$\cosh L_2 \sim \sinh a \frac{e^b}{2} - \cosh a \frac{e^b}{2} \cos \gamma.$$

To keep L_2 bounded, we need that $\cos \gamma \sim \tanh a$. But the same formula applied to the top left pentagon would imply that to keep L_1 bounded, we need that $\cos \gamma \sim -\tanh a$ and both are not possible simultaneously.

2nd case: $a \to \infty$ and $b \to \infty$. Then formula (2) implies that

$$\cosh L_2 \sim \frac{e^{a+b}}{4} (1 - \cos \gamma).$$

To keep L_2 bounded, we need that $\gamma \to 0$ and this is incompatible with keeping L_1 bounded.

3rd case: $a \to 0$ and $b \to \infty$. This is the most subtle case and we need to dig more into formula (1). Formula (2) implies already that

$$\cosh L_2 \sim \frac{a}{2}\frac{e^b}{2} - \frac{e^b}{2} \cos \gamma \quad \text{and} \quad \cosh L_1 \sim \frac{a}{2}\frac{e^b}{2} + \frac{e^b}{2} \cos \gamma.$$

Keeping both L_1 and L_2 bounded would imply that $e^b a$ is bounded, or $b \leqslant c - \log a$, where c is a constant.

On the other hand, formula (1) with $a \to 0$ and $b \to \infty$ becomes

$$b \geqslant \frac{1}{h} \log \left(\frac{4}{ha} \right) + o(a)$$

(see [8]). Combining the two formulas, we get

$$c - \log a \geqslant \frac{1}{h} \log \left(\frac{4}{ha} \right) + o(a).$$

Reordering, we have,

$$c + \left(\frac{1}{h} - 1 \right) \log a \geqslant \frac{1}{h} \log \left(\frac{4}{h} \right) + o(a)$$

---

[5]Indeed, if a and b are both bounded and, we see that $\cosh L_2$ is eventually negative as $\gamma \to 0$, which is absurd. If only a tends to $+\infty$ and b stays bounded, the same argument also applies because $\cosh a$ is arbitrarily close to $\sinh a$, when a tends to $+\infty$.

which is absurd since the left-hand side tends to $-\infty$ while the right hand side tends to $+\infty$ (when both a and h tend to 0).

Another, more intuitive, way to analyze this argument would be to remark that $b \to \infty$ is responsible for the fact that $h \to 0$ and conversely $a \to 0$ has the tendency to keep h away from 0. The opposition is settled by the relation ae^b bounded (which means a wins, a and b do not have a symmetric role). Hence, it is reasonable to reach a contradiction if we moreover assume that $h \to 0$. $\square$

We conclude this paragraph by an example showing that we cannot bound from below the volume entropy by an absolute constant. Those examples are well known, we only discuss them for completeness.

Proposition 25. *There exists a sequence of hyperbolic metrics on the torus with one boundary component whose volume entropy tends to zero.*

Proof. On the Poincaré disk model for the hyperbolic plane, we consider two orthogonal geodesics A and B meeting at the basepoint o. We denote by α and β the loxodromic isometries whose axis are A and B respectively and with the same translation length denoted $2a$ (the factor 2 makes the computations a bit easier).

A classical ping-pong-type argument shows that, when a is big enough, the group Γ_a generated by α and β is free, the quotient of the disk by Γ_a is a torus with a funnel and the convex core of the latter is a torus with one boundary component.

We will argue that, as a tends to infinity, the volume entropy of this torus tends to 0. To achieve this computation, we will use that the volume entropy is bounded above by the Hausdorff dimension of the limit set $\Lambda(\Gamma_a)$ [40] and actually compute the Hausdorff dimension.

Since the translation lengths of α and β are the same, the limit set is a self-similar Cantor set and we may compute its dimension for instance with [28, Theorem 4.14]. Since α is a Lipschitz map on the boundary, we look at "the quarter of the limit set" given by $\alpha(\Lambda(\Gamma_a))$.

The isometries α and β both move the point o to a point along their axis at Euclidean distance $\tanh a$. We deduce that the contraction ratios of α and β on $\alpha(\Lambda(\Gamma_a))$ are

$$r_\alpha = r_\beta \frac{\arccos\left(1 - \frac{(1-\tanh a)^2}{2}\right)}{\arccos\left(1 - \frac{\left(1-\tanh\left(\frac{a}{2}\right)\right)^2}{2}\right)} \underset{a\to\infty}{\sim} \frac{1+e^a}{1+e^{2a}}$$

(we use the spherical distance on $\alpha(\Lambda(\Gamma_a))$, which is bi-Lipschitz to the Euclidean distance). Finally, the Hausdorff dimension is given by ([28, Theorem 4.14])

$$\mathrm{Hdim}(\Lambda(\Gamma_a)) = \frac{-\log 3}{\log r_\alpha},$$

which tends to 0 (linearly) as $a \to \infty$. $\square$

4. Robustness from intersection patterns of a family of non-contractible geodesics on the two-torus

In this section, we discuss how a certain intersection pattern of closed (non-contractible) geodesics on T^2 implies robustness of topological entropy. Remarkably, and in contrast to the condition discussed in Sect. 2, this intersection pattern appears for a C^∞-generic metric. In other words, we obtain that topological entropy is C^0 robust for C^∞-generic metrics. The content in this section is motivated by the work of Bolotin and Rabinowitz [10] and Glasmachers and Knieper [20] and we use some of their results.

4.1. A definition of separation for lifts of two freely homotopic loops

Fix a free homotopy class α of loops in Q, $g \in \mathfrak{G}(Q)$. Let $\gamma, \gamma' \in \mathcal{L}_\alpha Q$ of energy $a = \mathcal{E}_g(\gamma), a' = \mathcal{E}_g(\gamma')$. Fix two lifts $\widetilde{\gamma}$ and $\widetilde{\gamma}' : \mathbb{R} \to \widetilde{Q}$ to the universal cover $\widetilde{Q}$ of Q. (This is understood as first lifting γ and γ' to maps $\mathbb{R} \to Q$ and then lifting to $\widetilde{Q}$.)

Definition 26. Define $b_0 = b_0(\widetilde{\gamma}, \widetilde{\gamma}')$ to be the infimum of the numbers $b > 0$ such that there is a continuous path in $\mathcal{L}_\alpha Q_g^{<b}$ from γ to γ' that lifts to a path from (the fixed lift) $\widetilde{\gamma}$ to (the fixed lift) $\widetilde{\gamma}'$. We define the *separation* of $\widetilde{\gamma}$ and $\widetilde{\gamma}'$ to be the non-negative real number $\operatorname{sep}_g(\widetilde{\gamma}, \widetilde{\gamma}') = \min\{\log(\frac{b_0}{a}), \log(\frac{b_0}{a'})\} \geqslant 0$. $\square$

The following robustness statement follows from the definitions.

Lemma 27. *Let $\delta > 0$, and g, g' be two metrics with $d_{C^0}(g, g') < \delta$. Then $\operatorname{sep}_{g'}(\widetilde{\gamma}, \widetilde{\gamma}') \leqslant \operatorname{sep}_g(\widetilde{\gamma}, \widetilde{\gamma}') + 2\delta$.*

Proof. Let $a = \mathcal{E}_g(\gamma)$ and $a' = \mathcal{E}_g(\gamma')$, and u_s a path from γ to γ' in $\mathcal{L}Q_g^{<b}$ for some $b > 0$ with $\min(\log(\frac{b}{a}), \log(\frac{b}{a'})) < \operatorname{sep}_g(\widetilde{\gamma}, \widetilde{\gamma}') + \epsilon$ for some $\epsilon > 0$, and which lifts to a path from $\widetilde{\gamma}$ to $\widetilde{\gamma}'$. Then $\mathcal{E}_{g'}(\gamma) > e^{-\delta} a$, $\mathcal{E}_{g'}(\gamma') > e^{-\delta} a'$, and $\mathcal{E}_{g'}(u_s) < e^\delta b$ for all s. In other words $\operatorname{sep}_{g'}(\widetilde{\gamma}, \widetilde{\gamma}') < \min\{\log(\frac{e^{2\delta} b}{a}), \log(\frac{e^{2\delta} b}{a'})\} < \operatorname{sep}_g(\widetilde{\gamma}, \widetilde{\gamma}') + 2\delta + \epsilon$, and since $\epsilon > 0$ was arbitrary the claim follows. $\square$

4.2. The two-torus and an intersection pattern

In the following, let $Q = T^2$ be the two-torus equipped with the standard orientation and α a non-trivial free homotopy class of loops in T^2. If a lift $\widetilde{\gamma}$ of a closed oriented curve γ representing α is embedded, then it divides the universal cover $\widetilde{T^2}$ in two connected components, the right $R(\widetilde{\gamma})$ and the left $L(\widetilde{\gamma})$ of $\widetilde{\gamma}$.

In the following definition, we formulate an intersection pattern of (lifts of) four closed curves of class α, see also Fig. 4 below.

Definition 28. We say that four oriented closed curves $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ in T^2 that represent α form a *ribbon* if

(0) their lifts to $\widetilde{T^2}$ are embedded,

and if for some choice of lifts $\widetilde{\gamma}_1, \widetilde{\gamma}_2, \widetilde{\gamma}_3, \widetilde{\gamma}_4$ to $\widetilde{T^2}$,

(1) $\widetilde{\gamma}_1$ is on the left of $\widetilde{\gamma}_3$ and $\widetilde{\gamma}_4$; $\widetilde{\gamma}_4$ is on the right of $\widetilde{\gamma}_1$ and $\widetilde{\gamma}_2$.

(2) $\widetilde{\gamma}_1$ and $\widetilde{\gamma}_2$ intersect, $\widetilde{\gamma}_2$ and $\widetilde{\gamma}_3$ intersect, and $\widetilde{\gamma}_3$ and $\widetilde{\gamma}_4$ intersect, and all intersections are transverse.

Let $\varepsilon > 0$. We say that $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ form an *ε-ribbon* (with respect to the metric g), if in addition to $(0), (1)$, and (2) (which is included in (4) below) the four lifts $\widetilde{\gamma}_1, \ldots, \widetilde{\gamma}_4$ satisfy that

(3) $\operatorname{sep}(\widetilde{\gamma}_i, \widetilde{\gamma}_j) \geqslant \varepsilon$, for all $i, j \in \{1, \ldots, 4\}$, $i \neq j$.

(4) two lifts $\widetilde{\tau}$ and $\widetilde{\tau}'$ of two closed curves τ, τ' of class α intersect whenever they satisfy one of the following
- $\operatorname{sep}_g(\widetilde{\tau}, \widetilde{\gamma}_1) < \varepsilon$ and $\operatorname{sep}_g(\widetilde{\tau}', \widetilde{\gamma}_2) < \varepsilon$.
- $\operatorname{sep}_g(\widetilde{\tau}, \widetilde{\gamma}_2) < \varepsilon$ and $\operatorname{sep}_g(\widetilde{\tau}', \widetilde{\gamma}_3) < \varepsilon$.
- $\operatorname{sep}_g(\widetilde{\tau}, \widetilde{\gamma}_3) < \varepsilon$ and $\operatorname{sep}_g(\widetilde{\tau}', \widetilde{\gamma}_4) < \varepsilon$.

□

The following Proposition 29 states that an ε-ribbon is robust with respect to the C^0 topology on the metrics. The main difficulty is to guarantee items (0) and (1) of a ribbon for a perturbed metric. Here results on the analysis of the curve-shortening flow are used [4,5,21].

Proposition 29. *Assume that there are four curves $\gamma_1, \ldots, \gamma_4$ that form an ε-ribbon for some $\varepsilon > 0$ with respect to g. Let $\delta > 0$ with $\varepsilon > 2\delta > 0$. Then for any metric g' with $d_{C^0}(g', g) < \delta$ there are four closed geodesics $\gamma_1', \ldots, \gamma_4'$ that form an $(\varepsilon - 2\delta)$-ribbon with respect to g'.*

Proof. The proof uses in an essential way the analysis of the curve shortening flow. Consider the space of embedded closed smooth curves $\Gamma = \{\gamma \in \mathcal{L}Q \mid \gamma \text{ embedded}\}$. The curve shortening flow is a continuous local semi-flow $\Phi^t : \Gamma \to \Gamma$, $\Phi^t(\gamma_0) = \gamma_t$ for $t \in [0, T_{\gamma_0})$, defined by $\frac{\partial \Phi}{\partial t} = k_t N_t$, where k_t is the geodesic curvature of γ_t and N_t the unit normal vector. We need the following properties (see [6]): Mutually non-intersecting curves γ_0 and γ_0' stay non-intersecting along the flow [4,5]. Assume that Q is compact with geodesic boundary, then for $\gamma \in \Gamma$, either the maximal T_γ is finite and $\Phi(\gamma)$ converges to a point, or $T_\gamma = +\infty$ and $\Phi(\gamma)$ converges to a geodesic [21]. The length is strictly decreasing under Φ^t.

Let now $\varepsilon > 0$ and $\gamma_1, \ldots, \gamma_4$ be the four curves representing α that form an ε-ribbon with respect to g for lifts $\widetilde{\gamma}_1, \ldots, \widetilde{\gamma}_4$, and let $\delta > 0$ and g' as in the proposition. Let $\mathcal{T}$ be the deck transformation corresponding to α on the universal covering $\widetilde{T^2}$ of T^2. The quotient $\widetilde{T^2}/\mathcal{T}$ by the action of $\mathcal{T}$ is an annulus, and we apply the curve shortening flow on $\widetilde{T^2}/\mathcal{T}$ with respect to the metric g' starting with the projection on $\widetilde{T^2}/\mathcal{T}$ of the four lifts. Note here that since the lifts in $\widetilde{T^2}$ are embedded and are invariant under the action of $\mathcal{T}$ their projections to $\widetilde{T^2}/\mathcal{T}$ are embedded (with respect to a suitable parametrization). The images of the curve shortening flow of these curves stay inside an annulus with geodesic boundary in $\widetilde{T^2}/\mathcal{T}$. By the properties mentioned above, the flow will converge to four embedded geodesics in $\widetilde{T^2}/\mathcal{T}$ that project to four geodesics $\gamma_1', \ldots, \gamma_4'$ in (T^2, g'). Moreover, the path of curves given by the curve shortening flow lifts to a path from $\widetilde{\gamma}_i$ to $\widetilde{\gamma}_i'$, $i = 1, 2, 3, 4$. These four curves and their lifts form an $(\varepsilon - 2\delta)$-ribbon. Properties

(0) and (1) follow immediately from the properties of the curve shorting flow. Reparametrizing uniformly by arc-length does not increase the energy along each of the four paths of closed curves given by the curve shortening flow, nor does it increase the energy along the obtained path. In particular the energy will be bounded from above by the energy of the starting curve. Hence for $i \neq j$, $\mathrm{sep}_{g'}(\widetilde{\gamma}_i', \widetilde{\gamma}_j') \geqslant \mathrm{sep}_{g'}(\widetilde{\gamma}_i, \widetilde{\gamma}_j)$. By Lemma 27, $\mathrm{sep}_{g'}(\widetilde{\gamma}_i, \widetilde{\gamma}_j) \geqslant \mathrm{sep}_{g}(\widetilde{\gamma}_i, \widetilde{\gamma}_j) - 2\delta \geqslant \varepsilon - 2\delta$.

To see property (4), consider two closed curves τ and τ' with lifts $\widetilde{\tau}$ and $\widetilde{\tau}'$ to $\widetilde{T^2}$ such that $\mathrm{sep}_{g'}(\widetilde{\tau}, \widetilde{\gamma}_1') < \epsilon - 2\delta$ and $\mathrm{sep}_{g'}(\widetilde{\tau}', \widetilde{\gamma}_2') < \varepsilon - 2\delta$. Since for every $b > \mathcal{E}_{g'}(\gamma_1)$, there is a path from γ_1 to γ_1' in $\mathcal{L}_\alpha Q^{<b}_{\alpha,g'}$, $\mathrm{sep}_{g'}(\widetilde{\tau}, \widetilde{\gamma}_1) < \varepsilon - 2\delta$, and hence $\mathrm{sep}_{g}(\widetilde{\tau}, \widetilde{\gamma}_1) < \varepsilon$. Similarly, $\mathrm{sep}_{g}(\widetilde{\tau}', \widetilde{\gamma}_2) < \varepsilon$. Therefore, $\widetilde{\tau}$ and $\widetilde{\tau}'$ intersect. Similarly one checks the remaining cases of property (4). $\square$

4.3. Robustness of entropy via ribbons

We will see that the intersection pattern of geodesics on T^2 considered above implies that the metric has robust topological entropy.

Theorem 30. *If (T^2, g) admits four closed geodesics $\gamma_1, \ldots, \gamma_4$ that form a ribbon, then the topological entropy of the geodesic flow φ_g is positive. Moreover, the topological entropy is bounded from below by $\frac{1}{L}\log 2$, where $L = \min\{l(\gamma_1) + l(\gamma_2), l(\gamma_3) + l(\gamma_4)\}$.*

Proof. See [20, Lemma 4.2] for a similar argument. In the following consider the strip $S = R(\widetilde{\gamma}_1) \cap L(\widetilde{\gamma}_4) \subset \widetilde{T^2}$. By our assumptions, we can choose $U_0 \subset S$ a connected component of $R(\widetilde{\gamma}_1) \cap L(\widetilde{\gamma}_2)$ and $U_1 \subset \widetilde{T^2}$ a connected component of $L(\widetilde{\gamma}_4) \cap R(\widetilde{\gamma}_3)$ such that $U_0 \cap U_1 \neq \emptyset$. Let $\mathcal{T} \neq id$ be the covering transformation corresponding to the free homotopy class α of the geodesics $\gamma_1, \ldots, \gamma_4$. Note that by the assumptions, for any $i, j \in \mathbb{Z}$, $i \neq j$, $\mathcal{T}^i U_0 \cap \mathcal{T}^j U_0 = \emptyset$, and $\mathcal{T}^i U_1 \cap \mathcal{T}^j U_1 = \emptyset$. For any bi-infinite sequence $\mathfrak{a} = (a_i)_{i\in\mathbb{Z}}$, $a_i \in \{0, 1\}$, consider the set $D(\mathfrak{a}) = S \backslash \bigcup_{i\in\mathbb{Z}} \mathcal{T}^i U_{a_i}$. For any periodic $\mathfrak{a}$ of period p we choose a closed curve γ in class $p\alpha$ which has a lift in $D(\mathfrak{a})$. We can assume that γ is a geodesic, e.g., by applying Lemma 33 below, and has minimal energy, and hence also minimal length, among such curves. In particular, the length of γ is bounded from above by pL, where $L = \min\{l(\gamma_1) + l(\gamma_2), l(\gamma_3) + l(\gamma_4)\}$. This follows since there exist both, a closed curve in class $p\alpha$ of length smaller than $p(l(\gamma_1) + l(\gamma_2))$, and one of length smaller than $p(l(\gamma_3) + l(\gamma_4))$ whose lifts are contained in the boundary of $D(\mathfrak{a})$.

These geodesics γ provide us with separating sets for the geodesic flow of g: Let u be a compact connected set in $\overline{S} \backslash \bigcup_{i\in\mathbb{Z}} \mathcal{T}^i(U_0 \cap U_1)$ such that $\overline{S} \backslash u$ has two components, one of which contains the sets $\mathcal{T}^i(U_0 \cap U_1)$, $i \leqslant 0$, and the other the sets $\mathcal{T}^i(U_0 \cap U_1)$, $i > 0$. Furthermore, choose two disjoint compact connected sets $v_0 \subset U_0 \backslash U_1$ and $v_1 \subset U_1 \backslash U_0$ such that $v_0 \cup (U_0 \cap U_1) \cup v_1$ also divides $\overline{S}$ into two components. Lift the geodesic flow of (T^2, g) to a flow on $T^1\widetilde{T^2}$ and denote it by $\widetilde{\varphi}_g$, i.e., $\widetilde{\varphi}_g$ is the geodesic flow of $(\widetilde{T^2}, \widetilde{g})$ of the lifted metric $\widetilde{g}$. Let $P : T^1\widetilde{T^2} \to T^1T^2$ be the covering map induced by the universal covering $\widetilde{T^2} \to T^2$. Furthermore, denote by $\widetilde{u}$, $\widetilde{v_0}$, and $\widetilde{v_1}$ the lifts to

$T^1\widetilde{T^2}$ of u, v_0, and v_1, respectively. It is easy to see that there is a constant $k > 0$, such that for every $t_0 > 0$ there is a covering of $\bigcup_{t\in[0,t_0]} \widetilde{\varphi}_g^t(\widetilde{u})$ by less than kt_0^2 open sets such that for any of the sets B of this covering and any $x, y \in B$, $d_{\widetilde{g}}(x, y) = d_g(P(x), P(y))$.

By the discussion above, for any p-periodic binary word $\mathfrak{a} = (a_i)_{i\in\mathbb{Z}}$ there is a lift $\widetilde{\gamma}$ of a closed geodesic in (T^2, g) of length $\leqslant pL$ that intersects u, that intersects $\mathcal{T}^i(v_j)$ if and only if $a_i = j$, and that intersects $\mathcal{T}^i(U_j)$ if and only if $a_i \neq j$. Hence for ε sufficiently small there is a set $X \subset \widetilde{u}$ of at least 2^p points such that for all $x, y \in X$, $\sup_{t\in[0,pL]} d_{\widetilde{g}}\left(\widetilde{\varphi}_g^t(x), \widetilde{\varphi}_g^t(y)\right) \geqslant \varepsilon$. Hence there is an (ε, pL)-separated set of φ_g of cardinality at least $2^p/(k(pL)^2)$, and therefore

$$h_{\text{top}}(\varphi_g) \geqslant \limsup_{p\to\infty} \frac{\log(2^p/(k(pL)^2))}{pL} = \frac{1}{L}\log 2.$$

□

Theorem 30 together with Proposition 29 show that the existence of four curves that form an ε-ribbon implies that the topological entropy is positive in a C^0 neighborhood of g. Of course, the intersection pattern of a ribbon might not be robust. Nonetheless, the next result asserts that the existence of four geodesics that form a ribbon implies another collection of four geodesics that form an ε-ribbon for some $\varepsilon > 0$, and hence the topological entropy is indeed positive in a C^0 neighborhood of g.

Theorem 31. *If there are four geodesics $\gamma_1, \ldots, \gamma_4$ that form a ribbon for a metric g, then there is $\varepsilon > 0$ and four geodesics $\eta_1, \ldots, \eta_4$ that form an ε-ribbon for the metric g.*

From Theorem 31, 30 and Proposition 29 it follows that

Corollary 32. *If there are four geodesics $\gamma_1, \ldots, \gamma_4$ that form a ribbon for a metric g, then there exists $\delta > 0$ such that for all g' with $d_{C^0}(g', g) < \delta$, the topological entropy of the geodesic flow $\varphi_{g'}$ is positive.*

In the proof of Theorem 31, we use the following lemma repeatedly. For that let A be an annulus such that each boundary component b of A is piecewise geodesic with at least one but finitely many non-smooth points, that all have outer angle $> \pi$. We say that a boundary component b with these properties is *admissible*.

Lemma 33. *Let A have admissible boundary and let α be the free homotopy class of curves in A provided by a choice of generator of $\pi_1(A) = \mathbb{Z}$. Then there is a simple closed geodesic γ in A with $\mathcal{E}_g(\gamma) = d := \inf\{\mathcal{E}_g(x) \,|\, x : S^1 \to A, [x] = \alpha\}$ and there is $\varepsilon > 0$ such that $\widetilde{d} = \inf\{\mathcal{E}_g(x) \,|\, x : S^1 \to A,\, [x] = \alpha, x \cap \partial A \neq \emptyset\} \geqslant e^{\varepsilon} d$.*

Proof. The statement of this lemma is well known and can be proved similarly to Lemma 20. To that end, note that since the boundary components are admissible, any closed piecewise geodesic γ in A that intersects a non-smooth point x at some boundary component of A can be replaced by a piecewise

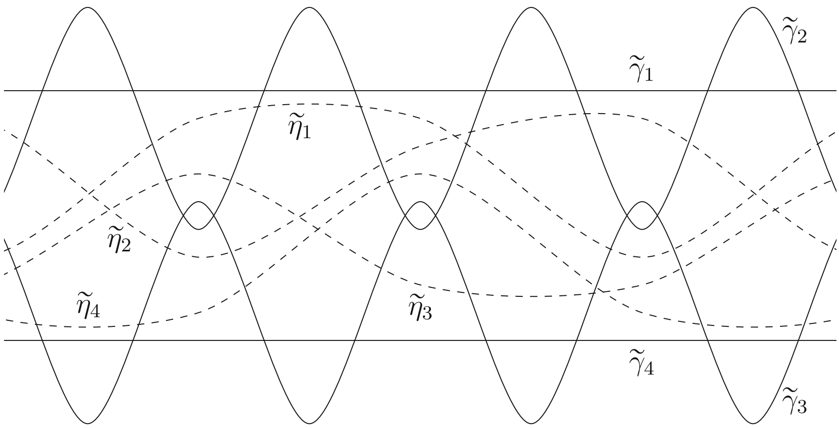


FIGURE 4. $\widetilde{\gamma}_1, \widetilde{\gamma}_2, \widetilde{\gamma}_3, \widetilde{\gamma}_4$ are lifts of four geodesics that form a ribbon. The paths $\widetilde{\eta}_1, \widetilde{\eta}_2, \widetilde{\eta}_3, \widetilde{\eta}_4$ are constructed in the proof of Theorem 31, and are lifts of four geodesics that form an ε-ribbon for some $\varepsilon > 0$

geodesic γ' in A whose image coincides with that of γ outside a neighborhood of b and that has strictly smaller energy. $\square$

Proof of Theorem 31. Let $\gamma_1, \dots, \gamma_4$ be four curves with lifts $\widetilde{\gamma}_1, \dots, \widetilde{\gamma}_4$ that form a ribbon. Choose U_0 and U_1 as in the proof before. Consider the four bi-infinite sequences $\mathfrak{a}^1, \dots, \mathfrak{a}^4$ of period 3 by extending periodically the four words 110, 011, 100, and 010, respectively. We find in two steps four closed geodesic $\eta_1, \dots, \eta_4$ representing 3α in T^2 with four lifts $\widetilde{\eta}_1, \dots, \widetilde{\eta}_4$ that lie in $D(\mathfrak{a}^1), \dots, D(\mathfrak{a}^4)$, respectively, and that form an ε-ribbon for some $\varepsilon > 0$.

Finding η_1 and η_2: Note that $D(\mathfrak{a}^1)$ and $D(\mathfrak{a}^2)$ are invariant under the shift $\mathcal{T}^3$, where $\mathcal{T}$ denotes the shift corresponding to α. Hence $D(\mathfrak{a}^1)$ and $D(\mathfrak{a}^2)$ project to annuli A_1 resp. A_2 in $\widetilde{T^2}/\mathcal{T}^3$. Their boundary components are admissible since the sequences $\mathfrak{a}^1$ and $\mathfrak{a}^2$ are non-constant. Fix choices $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ of the ε provided by Lemma 33 for A_1 resp. A_2. Let $\widehat{\eta}_1$ and $\widehat{\eta}_2$ be energy minimizing geodesics in A_1 and A_2 respectively, $\widetilde{\eta}_1, \widetilde{\eta}_2$ be some choice of lifts to $\widetilde{T^2}$, and η_1, η_2 their projections to T^2.

Finding η_3 and η_4: Consider now $D(\mathfrak{a}^3) \cap R(\widetilde{\eta}_1)$ and $D(\mathfrak{a}^4) \cap R(\widetilde{\eta}_1) \cap R(\widetilde{\eta}_2)$. These sets are invariant under $\mathcal{T}^3$ and project to annuli A_3 resp. A_4 in $\widetilde{T^2}/\mathcal{T}^3$. By the choice of the sequences $\mathfrak{a}^1, \dots, \mathfrak{a}^4$, one directly checks that the boundary components of A_3 and A_4 are admissible. Let $\varepsilon_3 > 0$ resp. $\varepsilon_4 > 0$ be choices of ε for A_3 resp. A_4 provided by Lemma 33. Let $\widehat{\eta}_3$, resp. $\widehat{\eta}_4$ be energy minimizing geodesics in A_3, resp. A_4, let $\widetilde{\eta}_3$, resp. $\widetilde{\eta}_4$ be some choice of lifts to $\widetilde{T^2}$, and η_3 resp. η_4 their projections to T^2.

By the intersection properties of the annuli A_1, A_2, A_3 and A_4, one sees that $\widetilde{\eta}_1$ and $\widetilde{\eta}_2$ intersect, that $\widetilde{\eta}_2$ and $\widetilde{\eta}_3$ intersect and that $\widetilde{\eta}_3$ and $\widetilde{\eta}_4$ intersect. By construction, $\widetilde{\eta}_4$ is on the right of $\widetilde{\eta}_1$ and $\widetilde{\eta}_2$, and $\widetilde{\eta}_3$ is on the right of $\widetilde{\eta}_1$. Hence, the geodesics $\eta_1, \dots, \eta_4$ form a ribbon.

Moreover, with $\varepsilon := \min\{\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4\}$, $\eta_1, \dots, \eta_4$ form an ε-ribbon. Indeed, let us see that $\mathrm{sep}_g(\widetilde{\eta}_1, \widetilde{\eta}_2) \geqslant \varepsilon_1$, the remaining conditions are checked analogously. Choose a path $u : [0,1] \to \mathcal{L}_{3\alpha}T^2$ that lifts to a path $\widehat{u}$ from $\widehat{\eta}_1$ to $\widehat{\eta}_2$. It is clear that there are $s \in [0,1]$ such that the lifted curve $\widehat{u}(s)$ touches the boundary of A_1, and let $s_1 \geqslant 0$ be the infimum in $[0,1]$ of such s. By compactness, $\widehat{u}(s_1)$ touches the boundary of A_1 and $\widehat{u}(s_1)$ is contained in A_1. Since $\widehat{\eta}_1$ has minimal energy among all curves in class 3α in A_1 we conclude with Lemma 33 that $\mathcal{E}_g(\widehat{u}(s_1)) \geqslant e^{\varepsilon_1}\mathcal{E}_g(\widehat{\eta}_1)$. Hence, $\mathrm{sep}_g(\widetilde{\eta}_1, \widetilde{\eta}_2) \geqslant \log\left(\frac{\mathcal{E}_g(\widehat{u}(s_1))}{\mathcal{E}_g(\widehat{\eta}_1)}\right) \geqslant \varepsilon_1$. □

4.4. Ribbons exist for C^∞ generic metrics

Minimal geodesics on higher genus surfaces and on the two-torus T^2, i.e., geodesics minimizing the length between any two of its points on the universal covering, were first systematically studied by Morse [32] and Hedlund [23]. Bolotin and Rabinowitz [10] obtained results about the existence of certain families of homoclinic and heteroclinic geodesics on T^2 that shadow minimal heteroclinics, using a renormalized length functional. As we will explain below, one can obtain from their work (more specifically [10, Theorem 4.2]) that under certain assumptions there is a (non-closed) geodesic in the universal cover which, after applying a family of covering transformations, provides a family of geodesics of a certain intersection pattern, similar to our ribbons. Furthermore, an argument analogous to the one in the proof of Theorem 31 then yields four closed curves that form a ribbon. The assumption in the theorem above is satisfied for a C^∞ generic metric. Hence one obtains the following

Theorem 34. *In the space of metrics on T^2 with positive topological entropy equipped with the C^∞ topology there is a co-meager set $\mathcal{S}$ such that any $g \in \mathcal{S}$ has robust topological entropy.*

We now discuss the result in [10] and its relation to the existence of curves that form a ribbon. We keep mainly the notations in [10].

Assume that the metric g on T^2 is not flat. Then, there is a simple free homotopy class of closed curves α in T^2 and two (possibly identical) minimal geodesics v_- and v_+ of class α that form an annulus $A \subset T^2$ that contains no minimal closed geodesics of class α in its interior, see [20]. Let $S \subset \widetilde{T^2}$ be the strip that is the preimage of A under the covering map. Let $\tau : \widetilde{T^2} \to \widetilde{T^2}$ be the translation corresponding to α. Let $\sigma : \widetilde{T^2} \to \widetilde{T^2}$ be a translation corresponding to a simple homotopy class such that, if $v_- = v_+$, then $\sigma(\widetilde{v}_-) = \widetilde{v}_+$, and otherwise $\sigma(\widetilde{v}_-)$ lies on the left of $\widetilde{v}_+$. Let $u_i, i \in \mathbb{Z}$, be the lifts in S of a shortest geodesic u connecting v_- and v_+, with $u_i = \tau^i u_0$. Consider the space α_i of (rectifiable) curves $x : [0,1] \to S$ with no constant pieces and such that $x(0) \in u_i$, $x(1) \in u_{i+1}$, and let

$$\Pi = \{y = (x_i)_{i\in\mathbb{Z}}\,|x_i \in \alpha_i, x_i(1) = x_{i+1}(0)\} \subset \prod_{i\in\mathbb{Z}} \alpha_i.$$

Let c be the (common) length of v_- and v_+. Define the *renormalized length functional* on Π as

$$J(y) := \sum_{i\in\mathbb{Z}} (\mathcal{L}(x_i) - c),$$

whenever the series is convergent, otherwise $J(y) := +\infty$. It is explained in [10] that J can be extended to paths $y : \mathbb{R} \to \widetilde{T^2}$, and in particular to $y : \mathbb{R} \to \widetilde{T^2}$ that are negative asymptotic to $\sigma^i(\widetilde{v}_-)$ and positive asymptotic to $\sigma^j(\widetilde{v}_+)$ for some $i, j \in \mathbb{Z}$. One defines a *barrier function* B_-^+ on S by

$$\begin{aligned} B_-^+(q) := \inf\{J(y)\,|\, y : \mathbb{R} \to S &\text{ goes through } q \\ &\text{and is asymptotic to } v_\mp \text{ as } t \to \mp\infty\}. \end{aligned}$$

One shows that B_-^+ is finite, and that the set of minimum points consists of $\widetilde{v}_- \cup \widetilde{v}_+$ and the set of minimal heteroclinics from $\widetilde{v}_-$ to $\widetilde{v}_+$, which is, moreover, non-empty. For minimal heteroclinics $h : \mathbb{R} \to S$ one has $B_-^+(h(t)) = J(h)$. Here *minimal heteroclinics* from $\widetilde{v}_-$ to $\widetilde{v}_+$ are globally minimizing geodesics $h(t)$ that are asymptotic to $\widetilde{v}_\mp$ as $t \to \mp\infty$. Parts of the Theorems 4.1 and Theorem 4.2 in [10] can be formulated as follows.

Theorem 35. [10] *Assume that B_-^+ is non-constant, and let $l \in \mathbb{N}$. Then, there is a heteroclinic $\widetilde{\gamma} : \mathbb{R} \to \widetilde{T^2}$ from $\widetilde{v}_-$ to $\sigma^l(\widetilde{v}_+)$, and minimal heteroclinics $h_i, i = 0, \ldots, l$ from $\widetilde{v}_-$ to $\widetilde{v}_+$ such that $\widetilde{\gamma}$ shadows $\sigma^i(h_i), i = 0, \ldots, l$.*

For the precise definition of shadowing, which is not important for our considerations, we refer to [10]. The time intervals in $\mathbb{R}$ for which $\widetilde{\gamma}$ shadows $\sigma^i(h_i)$ might be very far apart from each other, and so are the two time intervals in which $\widetilde{\gamma}$ is close to $\widetilde{v}_-$ resp. $\sigma^l(\widetilde{v}_+)$. Furthermore, the heteroclinics $\widetilde{\gamma}$ constructed in the theorem are embedded. The assumptions that B_-^+ is non-constant is equivalent to the assumption that there is no foliation of S by minimal heteroclinics from $\widetilde{v}_-$ to $\widetilde{v}_+$.

We now apply Theorem 35 and observe that it provides geodesics $\widetilde{\gamma} : \mathbb{R} \to \widetilde{T^2}$ such that together with certain translates, an intersection pattern similar to a ribbon appears, which we call ribbon*, and we proceed with the definition of this property, see Fig. 5 for an illustration. In the following, we say that two distinct geodesics η_1 and η_2 in $\widetilde{T^2}$ intersect positively (resp. negatively) at $\eta_1(t_1) = \eta_2(t_2)$ if the orientation given by the tangent vectors $(\eta_1'(t_1), \eta_2'(t_2))$ coincides (resp., does not coincide) with the orientation of T^2.

Definition 36. Let γ be a (non-closed) geodesic in T^2, and $\widetilde{\gamma}$ be a lift to $\widetilde{T^2}$. Assume $\widetilde{\gamma}$ is embedded. We say that $\gamma : \mathbb{R} \to T^2$ (or $\widetilde{\gamma}$), and five covering transformations $\theta_1, \theta_2, \theta_3, \theta_4, \mathcal{T} : \widetilde{T^2} \to \widetilde{T^2}$ form a ribbon* if for some parameters $s_j^i, t_j^i \in \mathbb{R}, s_j^i < t_j^i$, $j = 1, 4$; $i \in \mathbb{Z}$, and $u_j^i, v_j^i \in \mathbb{R}, u_j^i < v_j^i$, $j = 2, 3$; $i \in \mathbb{Z}$, the lifts $\widetilde{\gamma}_j^i = \mathcal{T}^i \circ \theta_j(\widetilde{\gamma})$, $j = 1, \ldots, 4$; $i \in \mathbb{Z}$ satisfy, for all $i \in \mathbb{Z}$, the following:

(0)
- $\widetilde{\gamma}_1^i$ and $\widetilde{\gamma}_1^{i+1}$ intersect negatively in $\widetilde{\gamma}_1^i(t_1^i) = \widetilde{\gamma}_1^{i+1}(s_1^{i+1})$,
- $\widetilde{\gamma}_4^i$ and $\widetilde{\gamma}_4^{i+1}$ intersect positively in $\widetilde{\gamma}_4^i(t_4^i) = \widetilde{\gamma}_4^{i+1}(s_4^{i+1})$.

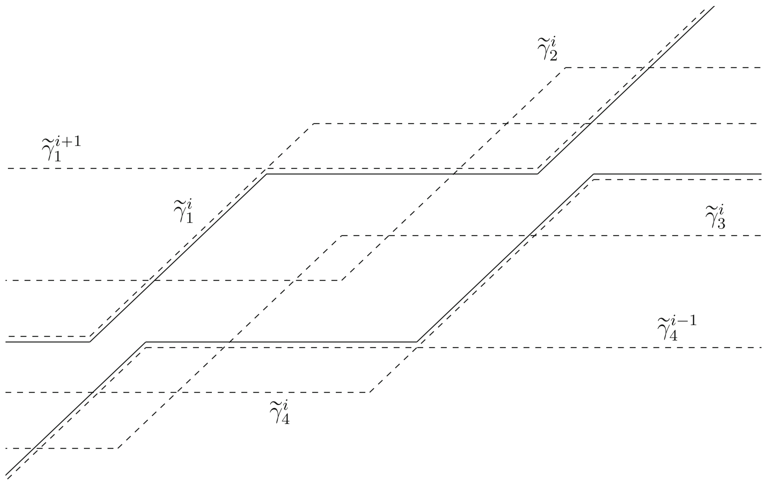


Figure 5. Schematic illustration of a heteroclinic from Theorem 35 with shifts that form a ribbon*. In non-horizontal parts there is a shadowing of minimal heteroclinics. Solid lines illustrate the course of η_1 and η_4 as given in Definition 36

(1) With $\eta_j^i := \widetilde{\gamma}_1^i|_{[s_j^i, t_j^i]}$, $j = 1, 4$, the piecewise geodesic $\eta_1 := \cdots \eta_1^{-1}\eta_1^0\eta_1^1 \cdots$ is on the left of the piecewise geodesic $\eta_4 := \cdots \eta_4^{-1}\eta_4^0\eta_4^1 \cdots$.

(2)
- $\eta_2^i := \widetilde{\gamma}_2^i|_{[u_2^i, v_2^i]}$ intersect η_1 only at the endpoints of η_2^i, first positively and then negatively, and does not intersect η_4.
- $\eta_3^i := \widetilde{\gamma}_3^i|_{[u_3^i, v_3^i]}$ intersect η_4 only at the endpoints of η_3^i, first negatively and then positively, and does not intersect η_1.

(3) η_2^i and η_3^i intersect.

(4) For all $i, j \in \mathbb{Z}$ with $i \neq j$, η_2^i and η_3^j are disjoint, η_2^i and η_2^j are disjoint, and η_3^i and η_3^j are disjoint.

$\square$

It is now easy to see, and we leave it to the reader to check, see also Fig. 5, that if $\widetilde{\gamma}$ is a heteroclinic from $\widetilde{v}_-$ to $\sigma^4(\widetilde{v}_+)$ obtained from Theorem 35, then for suitably chosen $n_2, n_3, n_4, n_5 \in \mathbb{N}$, $n_3 < n_2 < n_5 < n_4$, the geodesic $\widetilde{\gamma}$ together with the shifts $\theta_1 = \mathrm{id}$, $\theta_2 = \sigma^1 \circ \tau^{n_2}$, $\theta_3 = \sigma^{-2} \circ \tau^{n_3}$, $\theta_4 = \sigma^{-1} \circ \tau^{n_4}$, $\mathcal{T} = \sigma^3 \circ \tau^{n_5}$ form a ribbon*. Hence we conclude:

Proposition 37. *Assume that B_-^+ is non-constant. Then, there is a geodesic γ and deck transformations $\theta_1, \theta_2, \theta_3, \theta_4, \mathcal{T}$ on $\widetilde{T^2}$ that form a ribbon*.*

An analogous argument as in the proof of Theorem 31 yields

Proposition 38. *If there is a geodesic γ and shifts $\theta_1, \theta_2, \theta_3, \theta_4, \mathcal{T}$ that form a ribbon*, then there is $\varepsilon > 0$ and four closed geodesics $\tau_1, \ldots, \tau_4$ that form an ε-ribbon.*

Proof. The piecewise geodesics η_1 and η_4 form an infinite strip with piecewise geodesic boundary with outer angles $> \pi$ at the non-smooth points. Define for all $i \in \mathbb{Z}$ the non-empty sets $U_0^i = R(\eta_1) \cap L(\gamma_2^i)$, $U_1^i = L(\eta_4) \cap R(\gamma_3^i)$. By item (3) of the ribbon*, $U_0^i \cap U_1^i \neq \emptyset$. One can define for any binary bi-infinite sequence $\mathfrak{a}$ as in the proof of Theorem 30 (using the shift $\mathcal{T}$) sets $D(\mathfrak{a})$ which, by item (4) of the ribbon*, are infinite strips with piecewise geodesic boundary with outer angles $> \pi$ at the non-smooth points. Now one can proceed as in proof of Theorem 31 to obtain, for some $\varepsilon > 0$, closed geodesics $\tau_1, \ldots, \tau_4$ that form an ε-ribbon. □

Corollary 39. *Let g be a metric on T^2. If there is a geodesic γ and shifts $\theta_1, \theta_2, \theta_3, \theta_4, \mathcal{T}$ that form a ribbon*, then there is $\delta > 0$ such that for all g' with $d_{C^0}(g', g) < \delta$ the geodesic flow of g' has positive topological entropy.*

Note that for bumpy metrics closed minimal geodesics are hyperbolic, and note that bumpy is a C^∞ generic condition. In case B_-^+ is constant on S, then S is foliated by minimal heteroclinics from v_- to v_+, in other words the unstable manifold of v_- and the stable manifold of v_+ intersect, but not transversely. One can perturb in a neighborhood of any point of such a heteroclinic h_{vol}, as was shown by Donnay [19] (C^2 perturbation) and Petroll [37] (C^∞ perturbation) (for a sketch of the proof see also [13]), such that h_{vol} becomes a transverse heteroclinic connection from v_- to v_+. If such a perturbation is sufficiently small, v_- and v_+ stay to be adjacent (minimal) geodesics, where now B_-^+ is non-constant. So the assumptions in Theorem 35 hold C^∞ generically, which assures that Theorem 34 holds.

5. Robustness by retractable neck on general manifolds

Consider a closed Riemannian manifold (M, g) that is not necessarily a torus. As illustrated in Fig. 6, we assume that there exist nested nonempty open sets $U \subseteq V_1 \subseteq V_2 \subseteq W$ whose closures have smooth boundary such that

$$U \subseteq \overline{U} \subseteq V_1 \subseteq \overline{V}_1 \subseteq V_2 \subseteq \overline{V}_2 \subseteq W$$

and such that there is a retraction $\rho : M \backslash U \to M \backslash W$ that is homotopic to the identity relative $M \backslash W$ through the homotopy ρ_s. We define the two numbers

$$d_1 = \max_{x \in \partial V_2} l_g(\rho_s(x)),$$
$$d_2 = \mathrm{dist}_g(\partial V_1, \partial V_2).$$

We interpret this setup as follows: U is a head, that we intend to cut off. The set $W \backslash U$ is a neck that is further divided into lower neck $W \backslash V_2$, middle neck $V_2 \backslash V_1$ and upper neck $V_1 \backslash U$. The numbers d_i are the length of the lower neck d_1 and the length of the middle neck d_2, measured in a way that suits later proofs.

Assumption 40. *(Retractable neck and entropic body)* If the following statements are true, we say that the "neck" $W \backslash U$ is (c, k)-*retractable* for a number $c \in (0, 1)$ and $k \geqslant 3$.

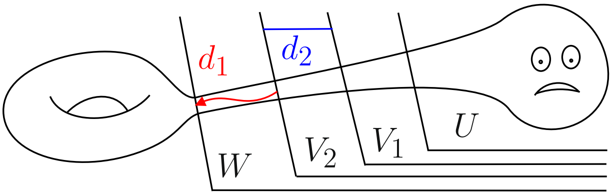


FIGURE 6. A torus with retractable neck before decapitation

- The retraction ρ is a contraction: for any curve $\gamma : I \to M\backslash U$, we have that $l_g(\rho \circ \gamma) \leqslant l_g(\gamma)$.
- The retraction ρ is a proper contraction in the middle and upper neck: for any curve $\gamma : I \to V_2\backslash U$ we have $l_g(\rho \circ \gamma) < c\, l_g(\gamma)$.
- The lower neck is substantially shorter than the middle neck:
$$\frac{d_1}{d_2} < \frac{1-c}{k}. \tag{5.1}$$

If the following statement is true, we say that M has an *entropic body.*

- There is a subset $\mathcal{P} \subseteq \widetilde{\pi}_1(M\backslash U)\backslash \iota\widetilde{\pi}_1(\partial U)$ of the free homotopy classes of $M\backslash U$ not homotopic to curves in ∂U, whose elements are mutually coprime such that the subsets
$$\mathcal{P}_g(T) = \{\alpha \in \mathcal{P} \mid \exists \gamma \in \alpha : l(\gamma) \leqslant T\}$$
grow exponentially: $\Gamma_T(\#\mathcal{P}_g(T)) > 0$.

□

Remark 41. Note that even though $\Gamma_T(\#\mathcal{P}_g(T))$ depends on the metric, the positivity
$$\Gamma_T(\#\mathcal{P}_g(T)) > 0$$
is a purely algebraic statement about the group growth of the free homotopy classes seen as $\widetilde{\pi}_1 = \pi_1/conj$. In particular, if π_1 has a subgroup isomorphic to $\mathbb{Z} * \mathbb{Z}$, then the assumption is satisfied.

Theorem 42. *Let the closed Riemannian manifold (M, g_0) have a (c, k)-retractable neck and an entropic body. Then, for C with $1 < C < \frac{k}{2+(k-2)c}$, g_0 has $\log C$-robust positive topological entropy and for $d_{C^0}(g, g_0) \leqslant \log C$ we have*
$$h_{\text{top}}(\varphi_g^t) \geqslant \frac{1}{\sqrt{C}}\Gamma(\#\mathcal{P}_{g_0}(T)).$$

Example 43. Let M^n be a manifold such that there exists a Riemannian metric whose topological entropy vanishes. Fix c and k. Consider a submanifold L such that $M\backslash L$ is an entropic body. We identify a tubular neighborhood of L with the normal disk bundle DL with radial coordinate $r \in [0, 3 + \frac{k}{2+(k-2)c}]$. We define the neck
$$U = \{r < 1\}, V_1 = \{r < 2\}, V_2 = \left\{r < 2 + \frac{k}{2+(k-2)c}\right\},$$

$$W = \left\{ r < 3 + \frac{k}{2+(k-2)c} \right\}$$

with the retraction defined by

$$\rho_s|_{M\setminus W} = id, \quad \rho_s|_W(r,x) = (r_s, x),$$

where $r_s = \min\left\{ r + \left(2 + \frac{k}{2+(k-2)c}\right) s, 3 + \frac{k}{2+(k-2)c} \right\}$. Endow the neck $M\setminus W$ with a metric g such that

$$g|_{r\in(1,3+\frac{k}{2+(k-2)c})} = f(r) g_{SL} + dr^2,$$

where g_{SL} is a metric of the normal sphere bundle over L and f is a function in r with $f(r) \geqslant f(3+\frac{k}{2+(k-2)c})$ for all $r \in (1, 3+\frac{k}{2+(k-2)c})$ and $c\sqrt{f(r)} \geqslant \sqrt{f(3+\frac{k}{2+(k-2)c})}$ for all $r \in (1, 2+\frac{k}{2+(k-2)c})$. With this metric, the neck has the (c,k)-retractable neck property. Thus, any extension of g to M satisfies the assumptions of the theorem.

In dimension 2, L is discrete. For S^2 three points and for T^2 one point suffice for having an entropic body. In dimension 3, L is a link. For S^3 the unlink with two components and for T^3 the unknot suffice. □

The following lemma is the C^0-robust property that we derive from a retractable neck.

Lemma 44. *Let (M, g_0) have a (c,k)-retractable neck. Let g be a metric with $d_{C^0}(g, g_0) \leqslant \log C$ for some number $1 < C < \frac{k}{2+(k-2)c}$. Let α be a homotopy class of a curve in $M\setminus U$ that is not homotopic to a curve in ∂U. Then, any g-length minimizer of α has image in $M\setminus V_1$.*

Proof. Let $\gamma \in \alpha$. Assume that there exists a T such that $\gamma(T) \in V_1\setminus U$. We claim that γ is not a length minimizer of α. We prove this by explicitly constructing a shorter curve homotopic to γ.

There is a maximal connected neighborhood $I \subseteq S^1$ of T such that $\gamma(I) \subseteq V_1\setminus U$. Since $\gamma \notin \iota\pi_1\partial U$, the interval I is not the entire circle. Because of maximality of $I = [t_1, t_2]$ the end points lie in the boundary $\gamma(t_1), \gamma(t_2) \in \partial V_1$. This implies that that $l_{g_0}(\gamma|_I) \geqslant 2d_2$ by definition of d_2.

We define the homotopy $\gamma_s(t)$, $s \in [0,1]$ as the concatenation

$$\gamma_s(t) = \gamma|_{S^1\setminus I} \circ \rho|_{[0,s]}(\gamma(t_1)) \circ \rho_s(\gamma|_I) \circ \overline{\rho|_{[0,s]}}(\gamma(t_2)).$$

Obviously $\gamma_0 \sim \gamma_1$. We rewrite the condition $C < \frac{k}{2+(k-2)c}$ as $\frac{1-cC}{2C} > \frac{1-c}{k}$. We compute

$$\begin{aligned} l_g(\gamma_0) - l_g(\gamma_1) &= l_g(\gamma|_I) - l_g(\rho\circ\gamma|_I) - l_g(\rho_s(\gamma(t_1))) - l_g(\overline{\rho_s}(\gamma(t_2))) \\ &\geqslant \frac{1}{\sqrt{C}} l_{g_0}(\gamma|_I) - \sqrt{C} l_{g_0}(\rho\circ\gamma|_I) - \sqrt{C} l_{g_0}(\rho_s(\gamma(t_1))) \\ &\quad - \sqrt{C} l_{g_0}(\overline{\rho_s}(\gamma(t_2))) \\ &> \sqrt{C}\left(\frac{1}{C} l_{g_0}(\gamma|_I) - c l_{g_0}(\gamma|_I) - l_{g_0}(\rho_s(\gamma(t_1))) - l_{g_0}(\overline{\rho_s}(\gamma(t_2))) \right) \end{aligned}$$

$$\begin{aligned}&\geqslant \sqrt{C}\left(\left(\frac{1}{C}-c\right)d_2-2d_1\right)=\sqrt{C}2\left(\left(\frac{1-cC}{2C}\right)d_2-d_1\right)\\&>2\sqrt{C}\left(\frac{1-c}{k}d_2-d_1\right)\\&>0.\end{aligned}$$

We conclude that $\gamma=\gamma_0$ is not a length minimizer. □

Proof of Theorem 42. Let $d_{C^0}(g,g_0)\leqslant \log C$ and $\alpha\in\mathcal{P}$. Since α is non-contractible, the infimal length of the homotopy class is positive, $l(\alpha):=\inf\{l(\gamma)\mid\gamma\in\alpha\}>0$. Let $\gamma_k:S^1\to M\backslash U$ be a sequence of smooth loops parametrized by constant speed with $l(\gamma_k)\to l(\alpha)$. Since $|\dot\gamma|\to l(\alpha)$ and since $M\backslash U$ is compact, we can apply Arzelà–Ascoli and find a subsequence that converges to a curve $\gamma_{\alpha,g}$ which satisfies $l(\gamma_{\alpha,g})\leqslant l(\alpha)$ by lower semi-continuity of the length functional. By minimality of $l(\alpha)$ this implies $l(\gamma_{\alpha,g})=l(\alpha)$. Thus, $\gamma_{\alpha,g}$ is a length minimizer. Lemma 44 tells us that the image of a length minimizer is contained in $M\backslash V_1$, which is in the interior of $M\backslash U$. We conclude that $\gamma_{\alpha,g}$ is a geodesic.

Thus, for every g with $d_{C^0}(g,g_0)<\log C$ and $\alpha\in\mathcal{P}$, there is a length minimizing geodesic $\gamma_{\alpha,g}:l_g(\alpha)S^1\to M\backslash U$, which we parametrize from now on by arc length for convenience. Note that $\gamma_{\alpha,g}$ lifts to a periodic orbit $(\gamma_{\alpha,g},\dot\gamma_{\alpha,g})$ of φ_g^t of period $l_g(\alpha)$.

The relation $l_g\leqslant\sqrt{C}l_{g_0}$ implies

$$\{\alpha\in\mathcal{P}\mid l_{g_0}(\alpha)\leqslant T\}\subseteq\{\alpha\in\mathcal{P}\mid l_g(\alpha)\leqslant\sqrt{C}T\}$$

and consequently the sets $\widetilde{\mathcal{P}}_g(T)=\{\gamma_{\alpha,g}\mid l_g(\gamma_{\alpha,g})<T\}$ satisfy

$$\Gamma(\#\widetilde{\mathcal{P}}_g(T))\geqslant\frac{1}{\sqrt{C}}\Gamma(\#\mathcal{P}_{g_0}(T)).$$

The desired statement now follows from Lemma 18. □

Proof of Theorem 12. Let (Q,g) be a $k\geqslant 2$ dimensional Riemannian manifold and let $e>0$ be arbitrary. We search metrics $g(s)\in\mathfrak{G}^e(Q)$ with $d_{C^0}(g,g(s))<s$ that have a $(c(s),k(s))$-retractable neck, where $\lim_{s\to 0}(c(s),k(s))=(1,3)$ and $\lim_{s\to\infty}(c(s),k(s))=(0,\infty)$. Theorem 42 then implies the statement qualitatively. The formula in Theorem 12 comes from the specific construction.

We first outline the argument: We prepare a small disk in which all geodesics considered will be contained. Then, we construct some heads inside such that the homotopy classes of curves in the disk minus the heads have positive algebraic growth. The growth of homotopy classes filtered by length will be at least the algebraic growth divided by the length of the longest generator. Then, we scale the entire construction down inside the disk, leaving the algebraic growth invariant but reducing the length of longest generator. This way, we find arbitrarily large entropy. This construction can be done by a C^0-small perturbation of the metric that is parametric in s.

Step 1: Choice and manipulation of a small disk. For an arbitrary point p, we choose a nearby metric $g_1=g_1(s)$ that is slightly reshaped around p:

We flatten a small disk surrounded by a thin cylindrical annulus. To quantify small, we choose $0 < \epsilon_1$ and $0 < \epsilon_2 \ll \epsilon_1$ in dependence of s: $\epsilon_1(s)$ is a continuous function that is linear for small s and constant $\ll 1$ for $s > s_0$ for some $s_0 \ll 1$.

We then choose g_1 such that

- $g_1 \equiv g_0$ on $Q \backslash B_{4\epsilon_1}(p)$,
- $g_1 = f(r) g_{S^{k-1}} + dr^2$ on $B_{3\epsilon_1}(p)$, where r is a radial coordinate,
- $f(r) \equiv 4$ on the annulus $r \in (2\epsilon_1 - \epsilon_2, 2\epsilon_1 + \epsilon_2)$,
- $f(r) = r^2$ on $B_{\epsilon_1}(p)$,
- $d_{C^0}(g_0, g_1) < s/2$,

where $g_{S^{k-1}}$ is the round metric on the euclidean sphere and where the radii of balls are measured with respect to g_1.

Note that choosing g_1 is very easy starting from coordinates that are orthonormal on T_pQ since we allow C^0-small perturbations. It would be impossible for C^2-small perturbations, as curvature would be an obstruction. The deformation around the annulus is a small deviation from the flat metric as long as ϵ_2 is small in comparison to ϵ_1.

The condition on the annulus is to ensure that the disk $B_{\epsilon_1}(p)$ is surrounded by a totally geodesic codimension 1 sphere, which will help to contain minimizing curves in the interior of the disk.

Step 2: Choice of heads. Let $\iota : L \hookrightarrow B_{\epsilon_1}(0)$ be an embedded codimension 2 submanifold: If $k = 2$, then we choose L to be three points, if $k > 2$ then let $L = L_1 \cup L_2$ have two unknotted components with $L_i \cong S^1 \times S^{k-3}$. Note that in both cases the group growth $\Gamma(\pi_1(B_{\epsilon_1}(p) \backslash \iota L, p)) =: \Gamma$ is positive. There is a length $\lambda(\iota L)$ such that there are generators of $\pi_1(B_{\epsilon_1}(p) \backslash \iota L, p)$ of length at most $\lambda(\iota L)$. To demonstrate the future argument, denote by $\mathcal{P}_{g_1}(\iota, T)$ the set of free homotopy classes of loops in $B_{\epsilon_1}(p) \backslash \iota L$ that are represented by a loop of length $\leqslant T$ and that do neither retract onto ιL nor to $\partial B_{\epsilon_1}(p)$. Then we have

$$\Gamma_T(\#\mathcal{P}_{g_1}(\iota, T)) \geqslant \Gamma/\lambda,$$

since for each free loop we find a (longer) representing based loop and the conjugacy classes of the fundamental group grow as fast as the fundamental group. If we postcompose the embedding ι with a dilation by a factor of t, the algebraic growth Γ will obviously not change but the group will be generated by loops of length $\lambda(t\iota L) = t\lambda(\iota L)$, implying that $\Gamma_T(\#\mathcal{P}_{g_1}(t\iota, T)) \to \infty$ as $t \to 0$.

Step 3: Shaping the necks. The shape of the necks is determined similar to the shape of the base disk. We choose in dependence of s the shape parameters of the neck $(c(s), k(s))$ such that $c(s) < e^{-s/8}$ and $k = 3 + s$. Further, we choose new and even smaller $0 < \epsilon_3$ and $0 < k\epsilon_4 \ll \epsilon_3$. We choose a nearby metric $g_2 = g_2(s)$ that is flattened in an ϵ_3-tube around L, except for an ϵ_4 wide annulus which imitates Example 43. More precisely,

- $g_2 \equiv g_1$ outside the tubular neighborhood $V_{3\epsilon_3}(N)$,
- $g_2 \equiv f(r) g_{SL} + dr^2$ inside $V_{2\epsilon_3}(N)$,
- $f(r) \geqslant f(\epsilon_3 + (3 + \frac{k}{2+(k-2)c})\epsilon_4)$ for all $r \in (\epsilon_3 + \epsilon_4, \epsilon_3 + (3 + \frac{3}{1-c})\epsilon_4)$,

- $c\sqrt{f(r)} \geqslant \sqrt{f(\epsilon_3 + (3 + \frac{k}{2+(k-2)c})\epsilon_4)}$ for all $r \in (\epsilon_3 + \epsilon_4, \epsilon_3 + (2 + \frac{k}{2+(k-2)c})\epsilon_4)$,
- $d_{C^0}(g_2, g_1) < s/2$,

where tubular neighborhoods are taken with respect to the metric g_2 and where g_{SL} is the metric of the normal sphere bundle over L. Note that if $f(r)$ would equal r^2, then the metric would be flat and the assumption would be similar to the choice of the small disk.

Note that the crucial feature of our choice of $c(s)$ is that $\log(1/c^2) < s/2$ because the fourth point forces us to deviate by a factor of c from the cylindrical metric, which forces $d_{C^0}(g_2, g_1) > \log(1/c^2)$ and the fifth point requires $d_{C^0}(g_2, g_1) < s/2$.

This new metric has a retractable neck with sets

$$U = \{r < \epsilon_3 + \epsilon_4\}, V_1 = \{r < \epsilon_3 + 2\epsilon_4\},$$
$$V_2 = \left\{r < \epsilon_3 + (2 + \frac{k}{2 + (k-2)c})\epsilon_4\right\}, W = \left\{r < \epsilon_3 + (3 + \frac{k}{2 + (k-2)c})\epsilon_4\right\}.$$

The inclusion $B_{\epsilon_1}(p)\backslash U \hookrightarrow B_{\epsilon_1}(p)\backslash L$ is obviously a homotopy equivalence. The generators of the fundamental group are possibly a bit longer, but $2\lambda(\iota L)$ suffices if ϵ_3 is small enough.

Step 4: Shrinking for growth. Now, we employ the dilation by t mentioned in Step 2 for the metrics $g_2(s)$ from Step 3. To be more formal, denote by δ_t the dilation by t in the flat model around p and by $g_t(s)$ the metric which coincides with $g_2(s)$ on $Q\backslash\delta_s(B_{2\epsilon_1-\epsilon_2})$ and with $t^2\delta^*_{1/t}g_2$ in a small neighborhood. Note that $t^2\delta^*_{1/t}g_{\mathrm{Euc}} = g_{\mathrm{Euc}}$ and that the scaling leaves ratios intact, so $d_{C^0}(g_t(s), g_1) = d_{C^0}(g_2(s), g_1)$ and in total

$$d_{C^0}(g_t(s), g_0) < d_{C^0}(g_t(s), g_1(s)) + d_{C^0}(g_1(s), g_0) < s.$$

Let ρ be a free homotopy class of loops in the disk with the scaled heads $U_s = \delta_s U$ removed $B_{2\epsilon_1}(p)\backslash U_s$, which neither retracts to ∂U nor to $\partial B_{2\epsilon_1}$. Choosing a length infimizing sequence, we find by Arzelà–Ascoli up to passing to a subsequence a limit loop γ for ρ. This minimizer cannot touch ∂U by construction of a retractable neck. Nor can it touch $\partial B_{2\epsilon_1}$ as otherwise, it would be tangent to a geodesic in the geodesic foliation of $\partial B_{2\epsilon_1}$ that comes from the cylindrical metric on the annulus $r \in (2\epsilon_1 - \epsilon_2, 2\epsilon_1 + \epsilon_2)$ and, thus, would be a geodesic belonging to that foliation, contradicting our assumption on its homotopy class. Thus, each class in $\mathcal{P}_{g_1}(s\iota, T)$ from Step 2 is represented by a geodesic.

As noted in Step 2, the fundamental group $\pi_1(B_{\epsilon_1}(p)\backslash U_s, p)$ is generated by loops of length $< 2t\lambda(\iota L)$. Thus, we may choose for each s a t so small that $\Gamma/(2t\lambda(\iota L)) > e$, where e is the exponential growth required in the statement of the theorem. To describe the necessary choice in dependence of s, note that by Theorem 42 for our choices $c(s) < e^{-s/8}$ and $k = 3 + s$ we obtain that for

$$C < \frac{s+3}{2 + (s+1)e^{-s/8}},$$

we can expect for $d_{C^0}(g_t(s), g) < \log C$ that $h_{\text{top}}(\varphi_g^t) \geqslant \frac{1}{\sqrt{C}}\Gamma(\#\mathcal{P}_{g_t(s)}(T))$. Thus, to enforce $h_{\text{top}}(\varphi_g^t) \geqslant e$ within the $\log \frac{s+3}{2+(s+1)e^{-s/8}}$-balls, we must have a dilation by at least

$$t \sim \text{const}\sqrt{\frac{2+(s+1)e^{-s/8}}{s+3}},$$

where the constant is in dependence of Γ and λ for a specific value. This gives us the required growth of minimizing geodesics which concludes the argument. $\square$

Proof of Corollary 13. The first two points in the corollary are immediately clear. For the third point, we start with the quasi-isometric embedding $\Phi_n : (\mathbb{R}^n, |\cdot|_\infty) \to (\overline{\mathfrak{G}}(T^2), d_{\text{RBM}})$ from [41], where the volume is fixed to 1 and the diameter bound is 100. Note that if Φ_n is quasi-isometric and $\widetilde{\Phi}_n$ is d_{C^0}-close to Φ_n, then also $\widetilde{\Phi}_n$ is quasi-isometric. So, the statement is proved by parametrically performing the above construction. Note that for this only the first step needs to be done parametrically, as from then on the construction is on the small disk which is flat for any starting metric. It is also sufficient to choose one constant but small s and a corresponding constant parameter t. As volume and diameter are C^0-continuous, the perturbed metrics have volume 1 after a rescaling by a factor close to 1 and the diameter still admits the bound of 101 as stated in our corollary. $\square$

Appendix A. Robustness of non-degenerate length spectrum

Here, we prove Proposition 7. The aim is to use only a low amount of technology.

Remark 45. Unfortunately, one cannot say anything about the position of the geodesic that is found by Proposition 7. $\square$

Before we start the proof, let us fix the setup: Let (M, g) be a closed Riemannian manifold. Denote by $\Omega = H^1(S^1, M)$ the Hilbert manifold[6] of closed loops in M. The non-constant critical points of the energy functional $\mathcal{E}_g : \gamma \mapsto \frac{1}{2}\int_0^1 g(\dot\gamma, \dot\gamma)\,\mathrm{d}t$ are exactly the closed geodesics. The negative gradient flow φ_g^t of $\mathcal{E}_g$ has the Palais–Smale property in this space. Denote the sublevel set $\{\gamma \in \Omega \mid \mathcal{E}_g(\gamma) \leqslant a\}$ by Ω_g^a.

That γ is non-constant and non-degenerate means that the connected component of $\operatorname{Crit}\mathcal{E}_g$ containing γ is a circle and Morse–Bott. If all geodesics are non-degenerate, then the energy spectrum is discrete. The following statement describes what happens topologically at a critical energy level.

[6]We use this setting to avoid working in the Fréchet manifold $C^\infty(S^1, M)$. However, by bootstrapping every geodesic ends up being smooth.

Proposition 46. ([9], see also [35]) *Assume that $c \in (a,b)$ is the only critical value in $[a,b]$. Denote $N_1,\dots,N_r$ the components of* $\mathrm{Crit}(\mathcal{E})$ *with $\mathcal{E}(N_i)=c$ and with indices $\lambda_1,\dots,\lambda_r$. Assume they are Morse–Bott. Then*

- *Each manifold N_i carries a well defined vector bundle $\nu^-N_i \subset T\Omega|_{N_i}$ of rank λ_i consisting of negative directions of $d^2\mathcal{E}_g$.*
- *The sublevel set Ω_g^b retracts onto a space homeomorphic to Ω_g^a with the disc bundles $D\nu^-N_i$ disjointly attached to Ω_g^a along their boundaries.*
- *The retraction $r:\Omega_g^b \to \Omega_g^a \bigcup_{\partial D\nu^-N_i} D\nu^-N_i$ can be chosen such that $\mathcal{E}_g \circ r \leqslant \mathcal{E}_g$ and such that $r|_{N_i}=id$ and $r|_{\Omega_g^a}=id$.*

Remark 47. This proposition gives inductive instructions to build a CW-complex homotopy equivalent to Ω. The building blocks are disk bundles, which are cell complexes. The retraction maps inductively provide the attaching maps.

Since we are only interested in the topology, we use the term *topologically non-degenerate* for a curve for which the conclusions of Proposition 46 hold:

Definition 48. We assume that $c \neq 0$ is the only critical value in (a,b). Denote N_i the components of $\mathrm{Crit}(\mathcal{E})$ and assume that they are all isolated circles representing reparametrizations of non-constant closed geodesic γ_i with energy c.

Then, we call γ_i *topologically non-degenerate* if there are vector bundles $\nu^-N_i \subseteq T\Omega|_N$ such that the sublevel set Ω_g^b retracts onto a space homeomorphic to Ω_g^a with the disc bundles $D\nu^-N_i$ attached to Ω_g^a along the boundary via a retraction r with $\mathcal{E}_g \circ r \leqslant \mathcal{E}_g$ and such that $r|_{N_i}=id$ and $r|_{\Omega_g^a}=id$. $\square$

Remark 49. The assumption that the spectral value is isolated is actually too strong for our purpose; It would suffice to demand in Theorem 5 that a topologically non-degenerate γ be isolated in the space of loops. The proof below would then work by localizing the gradient flow. One can do this by multiplying the gradient vector field with a bump function around a neighborhood of N_i that is flow-invariant in the intended energy interval, and that separates γ from other geodesics. The argument would become much more complicated as the resulting flows only locally transport the respective sub-level sets into each other.

We shall use a minimax principle. We use the following formulation from Klingenberg [26]. A *flow-family* $\mathcal{A}$ for $\mathcal{E}_g$ is a collection of subsets of Ω such that $\mathcal{E}_g|_A$ is bounded for all $A \in \mathcal{A}$ and such that $A \in \mathcal{A}$ implies $\varphi_g^t A \in \mathcal{A}$ for $t \geqslant 0$.

Proposition 50. ([26, Theorem 2.1.1]) *Let $\mathcal{A}$ be a flow-family for $\mathcal{E}_g$. Then*

$$\inf_{A\in\mathcal{A}} \sup_A \mathcal{E}_g$$

is a critical value of $\mathcal{E}_g$.

Proof of Proposition 7. We use Proposition 46 to define a suitable flow-family. For simplicity, assume that there is only one critical component. For Proposition 7 it is enough to consider the case $N_1 = N \cong S^1$. The fundamental class of the transverse bundle relative its boundary $[D\nu^- N; \partial D\nu^- N]$ has nonempty intersection with the core N since it has nonempty intersection with any interior point. By extension the same is true for the class $\omega := [\Omega_g^a \bigcup_{\partial D\nu^- N} D\nu^- N; \Omega_g^a]$. Denote by $r^*\omega$ the set of maps $u : (D\nu^- N; \partial D\nu^- N) \to (\Omega_g^b, \Omega_g^a)$ such that $[r \circ u] = \omega$. Then, the set of images of $u \in r^*\omega$ defines a flow-family.

The minimax value for $r^*\omega$ is the critical value c:

$$\inf_{u \in r^*\omega} \max \mathcal{E}_g \circ u \geqslant \inf_{u \in r^*\omega} \max \mathcal{E}_g \circ r \circ u \geqslant \mathcal{E}_g(N) = c.$$

The other inequality is trivial since $\mathcal{E}_g$ restricted to the unstable disk bundle of N has maximum c.

The robustness statement now follows by using the very same retraction r to define a flow family for the perturbed metric $\widetilde{g}$: Let $\varepsilon > 0$ be so small that c is the only critical value of $\mathcal{E}_g$ in $[(1-3\varepsilon)c, (1+3\varepsilon)c]$. Let $\widetilde{g}$ be a metric such that $\|v\|_{\widetilde{g}}^2 \in (1 - \frac{1}{2}\varepsilon, 1 + \frac{1}{2}\varepsilon)\|v\|_g^2$ for all v. Note that for such ε the following chain of inclusions holds

$$\Omega_{\widetilde{g}}^{(1-2\varepsilon)c} \subseteq \Omega_g^{(1-\varepsilon)c} \subseteq \Omega_g^c \subseteq \Omega_{\widetilde{g}}^{(1+\varepsilon)c} \subseteq \Omega_g^{(1+2\varepsilon)c}.$$

Let $r : \Omega_g^{(1+2\varepsilon)c} \to \Omega_g^{(1-\varepsilon)c} \bigcup_{\partial D\nu^- N} D\nu^- N$ be the retraction constructed with φ_g^t and $r^*\omega$ the class described above. Define the subset $\widetilde{\omega} \subset r^*\omega$ by restriction of the target space $u : (D\nu^- N; \partial D\nu^- N) \to (\Omega_{\widetilde{g}}^{(1+\varepsilon)c}, \Omega_{\widetilde{g}}^{(1-2\varepsilon)c})$. The set of images of maps in $\widetilde{\omega}$ is a flow-family for $\varphi_{\widetilde{g}}^t$ since it is defined through sub-level sets of $\mathcal{E}_{\widetilde{g}}$, and it is nonempty since it contains the φ_g^t-unstable disk bundle around N. We have

$$\inf_{u \in \widetilde{\omega}} \max \mathcal{E}_{\widetilde{g}} \circ u \geqslant \inf_{u \in \widetilde{\omega}} \max (1-\varepsilon)\mathcal{E}_g \circ u \geqslant (1-\varepsilon)c.$$

On the other hand for u the φ_g^t-unstable disk bundle around N we have $\max \mathcal{E}_{\widetilde{g}} \circ u \leqslant (1+\varepsilon)c$. Thus, the minimax principle produces some geodesic $\widetilde{\gamma}$ of $\mathcal{E}_{\widetilde{g}}$ with energy $|\mathcal{E}_{\widetilde{g}}(\widetilde{\gamma}) - c| \leqslant \varepsilon c$.

Note that for any u in the flow-family, every path in the image of u is homotopic to a loop in N since the intersection of u and N is nonempty. Thus, also $\widetilde{\gamma}$ is homotopic to the unperturbed geodesic. □

References

[1] Abbondandolo, A., Alves, M. R. R., Saglam, M., Schlenk, F.: Entropy collapse versus entropy rigidity for Reeb and Finsler flows. Preprint arXiv:2103.01144 (2021)

[2] Alves, M.R.R.: Cylindrical contact homology and topological entropy. Geom. Topol. **20**(6), 3519–3569 (2016)

[3] Alves, M.R.R., Meiwes, M.: Braid stability and the Hofer metric. Preprint arXiv:2112.11351 (2021)

[4] Angenent, S.: Parabolic equations for curves on surfaces. I. Curves with p-integrable curvature. Ann. Math. (2) **132**(3), 451–483 (1990)
[5] Angenent, S.: Parabolic equations for curves on surfaces. II. Intersections, blow-up and generalized solutions. Ann. Math. (2) **133**(1), 171–215 (1991)
[6] Angenent, S.: Curve shortening and the topology of closed geodesics on surfaces. Ann. Math. (2) **162**(3), 1187–1241 (2005)
[7] Alves, M.R.R., Pirnapasov, A.: Reeb orbits that force topological entropy. *Ergodic Theory and Dynamical Systems*, pp. 1–44 (2021)
[8] Balacheff, F., Merlin, L.: A curvature-free $\log(2k-1)$ theorem. *Proceedings of the American Mathematical Society* (2020)
[9] Bott, R.: Nondegenerate critical manifolds. Ann. Math. **2**(60), 248–261 (1954)
[10] Bolotin, S.V., Rabinowitz, P.H.: Some geometrical conditions for the existence of chaotic geodesics on a torus. Ergodic Theory Dyn. Syst. **22**(5), 1407–1428 (2002)
[11] Burtscher, A.Y.: Length structures on manifolds with continuous Riemannian metrics. N. Y. J. Math. **21**, 273–296 (2015)
[12] Buser, P.: Geometry and spectra of compact Riemann surfaces. Progress in Mathematics, vol. 106. Birkhäuser Boston Inc., Boston (1992)
[13] Burns, K., Weiss, H.: Spheres with positive curvature and nearly dense orbits for the geodesic flow. Ergodic Theory Dyn. Syst. **22**(2), 329–348 (2002)
[14] Chor, A., Meiwes, M.: Hofer's geometry and topological entropy. Preprint arXiv:2112.04955 (2021)
[15] Contreras, G.: Regularity of topological and metric entropy of hyperbolic flows. Math. Z. **210**(1), 97–111 (1992)
[16] Dahinden, L.: C^0-stability of topological entropy for contactomorphisms. Commun. Cont. Math. (2021). https://doi.org/10.1142/S0219199721500152
[17] de la Harpe, P.: Topics in Geometric Group Theory. Chicago Lectures in Mathematics. University of Chicago Press, Chicago (2000)
[18] Denvir, J., MacKay, R.S.: Consequences of contractible geodesics on surfaces. Trans. Am. Math. Soc. **350**(11), 4553–4568 (1998)
[19] Donnay, V.J.: Transverse homoclinic connections for geodesic flows. In: Hamiltonian dynamical systems (Cincinnati, OH, 1992), volume 63 of IMA Vol. Math. Appl., pages 115–125. Springer, New York, (1995)
[20] Glasmachers, E., Knieper, G.: Minimal geodesic foliation on T^2 in case of vanishing topological entropy. J. Topol. Anal. **3**(4), 511–520 (2011)
[21] Grayson, M.A.: Shortening embedded curves. Ann. Math. (2) **129**(1), 71–111 (1989)
[22] Hartman, P.: Geodesic parallel coordinates in the large. Am. J. Math. **86**, 705–727 (1964)
[23] Hedlund, G.A.: Geodesics on a two-dimensional Riemannian manifold with periodic coefficients. Ann. Math. (2) **33**(4), 719–739 (1932)
[24] Jost, J.: Riemannian Geometry and Geometric Analysis. Universitext, 5th edn. Springer, Berlin (2008)
[25] Katok, A., Knieper, G., Weiss, H.: Formulas for the derivative and critical points of topological entropy for Anosov and geodesic flows. Comm. Math. Phys. **138**(1), 19–31 (1991)

[26] Klingenberg, W.: Lectures on closed geodesics. Grundlehren der Mathematischen Wissenschaften, vol. 230. Springer, Berlin (1978)

[27] Manning, A.: Topological entropy for geodesic flows. Ann. Math. (2) **110**(3), 567–573 (1979)

[28] Mattila, P.: Geometry of sets and measures in Euclidean spaces. volume 44 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (1995).. (**Fractals and rectifiability**)

[29] Meiwes, M.: *Rabinowitz Floer homology, leafwise intersections, and topological entropy*. PhD thesis, (2018)

[30] Milnor, J.: Is entropy effectively computable? https://www.math.iupui.edu/~mmisiure/open/JM1.pdf, (2002)

[31] Mañé, R.: On the topological entropy of geodesic flows. J. Differ. Geom. **45**(1), 74–93 (1997)

[32] Morse, H.M.: A fundamental class of geodesics on any closed surface of genus greater than one. Trans. Am. Math. Soc. **26**(1), 25–60 (1924)

[33] Nabutovsky, A.: Morse landscapes of Riemannian functionals and related problems. In: Proceedings of the International Congress of Mathematicians. Volume II, pages 862–881. Hindustan Book Agency, New Delhi, (2010)

[34] Newhouse, S.E.: Continuity properties of entropy. Ann. Math. (2) **129**(2), 215–235 (1989)

[35] Oancea, A.: Morse theory, closed geodesics, and the homology of free loop spaces. In: *Free loop spaces in geometry and topology*, volume 24 of IRMA Lect. Math. Theor. Phys., pages 67–109. Eur. Math. Soc., Zürich, 2015. With an appendix by Umberto Hryniewicz

[36] Paternain, G.P.: On the topology of manifolds with completely integrable geodesic flows. Ergodic Theory Dyn. Syst. **12**(1), 109–121 (1992)

[37] Petroll, D.: Existenz und Transversalität von homoklinen und heteroklinen Orbits beim geodätischen Fluss. Ph.D. Thesis, Universitat Freiburg, (1996)

[38] Shiohama, K., Shioya, T., Tanaka, M.: The Geometry of Total Curvature on Complete Open Surfaces. Cambridge Tracts in Mathematics, vol. 159. Cambridge University Press, Cambridge (2003)

[39] Schapira, B., Tapie, S.: Regularity of entropy, geodesic currents and entropy at infinity. Ann. Sci. Éc. Norm. Supér. (4) **54**(1), 1–68 (2021)

[40] Sullivan, D.: The density at infinity of a discrete group of hyperbolic motions. Inst. Hautes Études Sci. Publ. Math. **50**, 171–202 (1979)

[41] Stojisavljević, V., Zhang, J.: Persistence Modules, Symplectic Banach-Mazur Distance and Riemannian Metrics. arXiv:1810.11151 (2019)

[42] Usher, M.: Symplectic Banach-Mazur distances between subsets in $\mathbb{C}^n$. J. Topol. Anal. **14**(1), 1–56 (2020)

Marcelo R. R. Alves
Faculty of Science
University of Antwerp
Campus Middelheim, Middelheimlaan 1
2020 Antwerpen
Belgium
e-mail: marcelorralves@gmail.com

Lucas Dahinden
Lucas Dahinden, Mathematisches Institut,
Ruprecht-Karls-Universität Heidelberg
Im Neuenheimer Feld 205
69120 Heidelberg
Germany
e-mail: l.dahinden@gmail.com

Matthias Meiwes and Louis Merlin
Chair for Geometry and Analysis
RWTH Aachen University
Jakobstrasse 2
52064 Aachen
Germany
e-mail: meiwes@mathga.rwth-aachen.de;
louis.merlin@hotmail.fr

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:22
https://doi.org/10.1007/s11784-022-00932-1
Published online April 4, 2022

Journal of Fixed Point Theory and Applications

Bifurcations of balanced configurations for the Newtonian n-body problem in $\mathbb{R}^4$

Luca Asselle, Marco Fenucci and Alessandro Portaluri

Dedicated to Claude Viterbo on the occasion of his sixtieth birthday.

Abstract. For the gravitational n-body problem, the simplest motions are provided by those rigid motions in which each body moves along a Keplerian orbit and the shape of the system is a constant (up to rotations and scalings) configuration featuring suitable properties. While in dimension $d \leq 3$ the configuration must be *central*, in dimension $d \geq 4$ new possibilities arise due to the complexity of the orthogonal group, and indeed there is a wider class of S-*balanced configurations*, containing central ones, which yield simple solutions of the n-body problem. Starting from the recent results in [2], we study the existence of continua of bifurcations branching from a trivial branch of collinear S-balanced configurations and provide an estimate from below on the number of bifurcation instants. In the last part of the paper, by using the continuation method, we explicitly display the bifurcation branches in the case of the three body problem for different choices of the masses.

Mathematics Subject Classification. 37J20 (=bifurcation in finite dimensions), 58J30 (=spectral flows), 70F10 (=N-body problem).

Keywords. n-Body problem, balanced configurations, central configurations, bifurcation of critical points, spectral flow of symmetric matrices.

1. Introduction

The Newtonian n-body problem concerns the motion of n point particles with masses $m_j \in \mathbb{R}^+$ and positions $q_j \in \mathbb{R}^d$, where $j = 1, \dots, n$ and $d \geq 2$, interacting each other according to Newton's law of inverses squares. The particles thus move according to Newton's equations of motion, which in this case read

$$m_j \ddot{q}_j = \frac{\partial U}{\partial q_j} \qquad \text{where} \qquad U(q_1, \dots, q_n) := \sum_{i<j} \frac{m_i m_j}{|q_i - q_j|}.$$

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

Letting M be the $(nd \times nd)$-diagonal MASS MATRIX defined by

$$M := \mathrm{diag}(\underbrace{m_1, \ldots, m_1}_{d\text{-times}}, \ldots, \underbrace{m_n, \ldots, m_n}_{d\text{-times}})$$

the equations of motion can be equivalently written as

$$\ddot{q} = M^{-1} \nabla U(q). \tag{1.1}$$

As the center of mass has an inertial motion, we can fix it without loss of generality at the origin. Among all possible configurations of the system, a crucial role is played by the so-called CENTRAL CONFIGURATIONS (CC for short), namely by those configurations in which $M^{-1}\nabla U(q)$ is parallel to q:

$$M^{-1}\nabla U(q) + \lambda q = 0. \tag{1.2}$$

In other words, the acceleration vector of each particle is pointing towards the origin with magnitude proportional to the distance to the origin. As a straightforward consequence of the homogeneity of the potential we obtain that the proportionality constant λ is actually equal to $-U(q)/\langle Mq, q\rangle$.

Equation (1.2) is a non-linear algebraic equation which is almost impossible to solve explicitly, and despite substantial progresses (starting from the work of—among others—Smale, Conley, Albouy, Chenciner, McCord, Moeckel, Pacella) have been made in the last decades, many basic questions about CC still remain unsolved. Nevertheless, there are several reasons why CC are of interest in the n-body problem and more generally in Celestial Mechanics:

- Every CC defines a HOMOTHETIC SOLUTION of (1.1), namely a solution which preserves its shape for all time while receding from or collapsing into the center of mass.
- Planar CC give rise to a family of periodic motions of (1.1), the so-called RELATIVE EQUILIBRIA, in which the configuration rigidly rotates at a constant angular speed about the center of mass. More generally, any such CC gives rise to a family of HOMOGRAPHIC SOLUTIONS of (1.1) in which each particle traverses an elliptical orbit with eccentricity $e \in (0, 1)$.
- CC control the qualitative behavior of total colliding solutions (and completely parabolic motions) of the n-body problem.

For the n-body problem in $\mathbb{R}^d$, $d \leq 3$, configurations which are not central cannot produce homographic motions at all. If we instead allow dimensions $d \geq 4$, then there is a wider class of so-called "S-balanced configurations" which produces relative equilibria of the n-body problem. These new high dimensional phenomena were first observed by Albouy and Chenciner in [1] (cfr. also [13]) and are due to the higher complexity of the orthogonal group, which allows, for example, to rotate in two mutually orthogonal planes with different angular velocities, thus leading to new ways of balancing the gravitational forces with centrifugal forces. We shall notice that, in contrast with the case $d = 2$, the resulting relative equilibria will be periodic in time only if the angular velocities are rationally dependent, and quasi-periodic otherwise.

We now define S-balanced configurations rigorously in the case $d = 4$, which we will focus for the rest of the paper onto. Thus, fix a positive real number $s > 1$ and consider the (4×4)-diagonal matrix

$$S = \mathrm{diag}(s, s, 1, 1).$$

Any solution of

$$M^{-1}\nabla U(q) + \lambda \widehat{S} q = 0, \quad \widehat{S} := \mathrm{diag}(S, ..., S), \tag{1.3}$$

is called an S-BALANCED CONFIGURATION, SBC for short, and gives rise to a uniformly rotating relative equilibrium solution of Eq. (1.1) in $\mathbb{R}^4$. Clearly, for $s = 1$, we obtain again (1.2). Also, Eq. (1.3) is, for every $s > 1$, invariant under the non-free (actually, not even locally free) diagonal $S^1 \times S^1$-action given by rotations in the $\mathbb{R}^2 \times \{0\}$ and $\{0\} \times \mathbb{R}^2$ planes. In particular, solutions always come in families, namely in S^1-families if they are contained in one of the two planes above and in $S^1 \times S^1$-families otherwise.

Remark 1.1. For $n = 3$, there is a big class of planar non-equilateral and non-collinear isosceles triangles which are SBC but not CC; for further details we refer to [13]. From a physical viewpoint, the larger s is the faster the bodies contained in the plane $\mathbb{R}^2 \times \{0\}$ rotates. Such a rough physical interpretation is paradigmatic of a deep stability issue which we are currently investigating.

In the study of Eq. (1.3) it is quite natural to interpret s as a bifurcation parameter, and hence to try to understand if:

(i) there exist configurations which are SBC (possibly collinear) for every choice of the parameter s (in other words, whether or not there are trivial branches of solutions), and
(ii) how many (if any) bifurcation points one has along such trivial branches.

As far as Question (i) is concerned, we readily see that collinear CC in the plane $\{0\} \times \mathbb{R}^2$ are solutions of (1.3) independently of $s > 1$ and hence define trivial branches of solutions $(\widehat{q}_s)_{s>1}$. Using the variational characterization of SBC (for more details see [2] or the following section) and the fact that along the trivial branches the Morse index jumps at precisely characterized values of the parameter s, we will provide the following answer to Question ii). For a more precise statement we refer to Theorem 4.7.

Theorem A. *For s_1 sufficiently close to 1 and s_2 large enough, there are at least $n!$ bifurcation instants from the trivial branches of solutions $(\widehat{q}_s)_{s\in[s_1,s_2]}$ corresponding to collinear CC in the $\{0\} \times \mathbb{R}^2$-plane.*

We shall notice already at this point that the non-trivial branches emanating from the trivial ones are genuine SBC (that is, not CC). This will be clear from the construction, anyway this also follows from the fact that collinear CC are isolated as central configurations in virtue of Moeckel's 45°-theorem.

As already mentioned, the main idea behind the result above is that, for variational problems in finite dimension (and, under suitable condition, also in infinite dimension), bifurcation instants along some trivial branch are detected by the jump of the Morse index as soon as the trivial branch is degenerate only in finitely many points. From a technical viewpoint, one difficulty to overcome is to rule out the degeneracy due to the $S^1 \times S^1$-symmetry of Equation (1.3). This will be done by means of a reduction argument (see Sect. 2).

In case $n = 3$, we will use numerical methods to provide a rather complete description of the non-trivial branches bifurcating from the trivial branches of collinear CC. Already in such an easy case, we observe some very interesting and rather unexpected phenomena: besides a strong dependence on the choice of the masses (which we recall is not the case for CC, as for any choice of the masses one has precisely 4 CC up to symmetry, namely the three Euler configurations, which are saddle points of U, and Lagrange's equilateral triangle, which is a global minimum of U), we e.g. observe the presence of connections between Lagrange's equilateral triangle and (some of the) Euler configurations through paths of SBC which are for any s local minima of U, as well as of turning points along some of the non-trivial branches at which the Morse index jumps but from which no secondary branches originate. This suggests that for larger values of n extremely interesting new phenomena might occur. We plan to study these aspects further in future work.

We shall also mention that other trivial branches can be constructed from planar CC in the plane $\{0\} \times \mathbb{R}^2$. Since the Morse index jumps also along such branches, we should be able to find other bifurcations instants. However, the problem is here more complicated since the degeneracy due to the symmetry cannot be overcome by reduction, and hence a generalization of the abstract bifurcation result (see Theorem 4.4) to an equivariant setting is needed. We plan to address this issue in future work.

We end up this introduction with a brief summary of the content of this paper: In Sect. 2 we define SBC and discuss their basic properties. In Sect. 3 we briefly recall the definition of the spectral flow in a finite dimensional setting. In Sect. 4 we prove an abstract bifurcation result from the trivial branch of a one parameter $\mathscr{C}^2$-family of functions on a finite-dimensional manifold and then apply it to the study of bifurcations of SBC. Finally, in Sect. 5 we use numerical computations to study the non-trivial branches bifurcating from a trivial branch of collinear CC in the case $n = 3$.

2. S-balanced configurations in the n-body problem

In this section we recall the definition of S-balanced configurations and their basic properties, referring to [2, 13] for the details. For $n \geq 2$, we consider n point-masses $m_1, ..., m_n > 0$, whose positions are denoted by $q_1, ..., q_n \in \mathbb{R}^4$ respectively, and which are supposed to interact with each other according to Newton's law of inverse squares. Setting the MASS MATRIX M to be the diagonal $(4n \times 4n)$-matrix

$$M := \operatorname{diag}(m_1 I_4, ..., m_n I_4),$$

where I_4 is the 4-dimensional identity matrix, we readily see that the equations of motion read

$$M\ddot{q} = \nabla U(q) \tag{2.1}$$

where $q = (q_1, ..., q_n) \in \mathbb{R}^{4n}$ is the CONFIGURATION VECTOR of the n point-masses, ∇ denotes the gradient in $\mathbb{R}^{4n}$, and U is the NEWTONIAN (GRAVITATIONAL) POTENTIAL

$$U(q) := \sum_{i<j} \frac{m_i m_j}{|q_i - q_j|}.$$

The invariance of (2.1) under translations implies in virtue of Noether's theorem that the center of mass

$$\overline{q} := \sum_{i=1}^{n} m_i q_i$$

has an inertial motion, and hence it is not restrictive to fix it at the origin. Therefore, we can without loss of generality suppose that U is defined over the space of COLLISION FREE CONFIGURATIONS WITH CENTER OF MASS AT THE ORIGIN

$$\widehat{\mathbb{X}} := \left\{ q = (q_1, ..., q_n) \in \mathbb{R}^{4n} \;\middle|\; \overline{q} = 0,\ q_i \neq q_j\ \forall i \neq j \right\}.$$

The set over which U is not defined, namely

$$\Delta := \{q \in \mathbb{R}^{4n} \mid \overline{q} = 0\} \setminus \widehat{\mathbb{X}},$$

is called the COLLISION SET. As it is nowadays well-known, Equation (2.1) is extremely hard to solve, and indeed a complete solution is possible only for $n = 2$. Therefore, instead of trying to solve (2.1) explicitly, one can try to look for (simple) solutions with prescribed behavior.

The simplest possible solutions of (2.1) one can think of are those given by *homographic motions*, i.e. rigid motions in which the configuration of the bodies remains constant (up to rotations and scalings) in time. If one makes such an Ansatz in dimension 2 or 3, then one finds that the configuration of the bodies must be CENTRAL, namely a solution of

$$M^{-1}\nabla U(q) + \lambda q = 0. \tag{2.2}$$

In other words, any solution of (2.2) gives, for suitable choice of the initial momentum, rise to a homographic motion. As it turns out, in this case, each body must then move along a Keplerian orbit. In the particular case of the zero angular momentum Keplerian orbit, we retrieve the so called *homothetic motions* in which all masses collapse simultaneously at the origin or recede from total collision. In case of the eccentricity zero Keplerian orbit instead, we retrieve the so called *relative equilibria*, in which the configuration of the bodies rigidly rotates around the origin at uniform speed while keeping its size constant. In dimensions 2 and 3, there are no other possible homographic motions. In dimension 4 instead new possibilities arise due to the higher complexity of the orthogonal group O(4): Indeed, in $\mathbb{R}^4$ it is possible to rotate simultaneously in two mutually orthogonal planes with different angular velocities. This produces a new balance between gravitational and centrifugal forces, thus yielding new periodic or quasi-periodic motions. Thus, for $d = 4$ there is a wider class of configurations, the so called S-BALANCED CONFIGURATIONS, which contains central ones and provides simple solutions to (2.1).

More precisely, fix a positive real number $s > 1$ and consider the (4×4)-diagonal matrix

$$S = \mathrm{diag}(s, s, 1, 1).$$

Any solution of

$$M^{-1}\nabla U(q) + \lambda \widehat{S} q = 0, \quad \widehat{S} := \mathrm{diag}(S, ..., S), \tag{2.3}$$

is called an S-balanced configuration. Clearly, for $s = 1$ we obtain (2.2). It is easily seen that any solution of (2.3) yields a relative equilibrium solution of (2.1),

$$q(t) := \begin{pmatrix} e^{i\sqrt{s}t} & 0 \\ 0 & e^{it} \end{pmatrix} \cdot q,$$

which will then be a periodic solution if $s \in \mathbb{Q}$ and a quasi-periodic solution otherwise. Notice also that (2.3) is invariant under the (diagonal) $S^1 \times S^1$-action on $\widehat{\mathbb{X}}$ given by rotations in the $\mathbb{R}^2 \times \{0\}$ and $\{0\} \times \mathbb{R}^2$ planes, whereas (2.2) is SO(4)-invariant. Both actions are not free (actually, not even locally free).

Taking the scalar product of (2.3) with q and using Euler's theorem, we see that the constant λ appearing in (2.3) must be equal to

$$\frac{U(q)}{I_S(q)},$$

where $I_S(q) := \langle \widehat{S} M q, q\rangle$ is the S-WEIGHTED MOMENT OF INERTIA, and as a direct consequence of the invariance under scalings of (2.3), we see that we can always normalize an S-balanced configuration to satisfy $I_S(q) = 1$. It is therefore natural to introduce the COLLISION FREE CONFIGURATION SPHERE

$$\widehat{\mathbb{S}} := \Big\{ q \in \widehat{\mathbb{X}} \;\Big|\; I_S(q) = 1 \Big\},$$

and to consider only NORMALIZED S-BALANCED CONFIGURATIONS, i.e. solutions of (2.3) which are contained in $\widehat{\mathbb{S}}$. We notice that Equation (2.3) on $\widehat{\mathbb{S}}$ reads

$$M^{-1}\nabla U(q) + U(q)\widehat{S} q = 0. \tag{2.4}$$

To simplify the notation, we will hereafter refer to solutions of (2.4) simply as S-balanced configurations. In other words, all S-balanced configurations will be hereafter assumed to be normalized. We will also use the shorthand notation SBC instead of S-balanced configuration.

Remark 2.1. The interest on S-balanced and central configurations goes far beyond the fact that they yield simple solutions of (2.1). Indeed, their properties turn out to be useful to understand the qualitative behavior of many other classes of solutions to (2.1), as e.g. colliding solutions. For more details we refer to [2].

Remark 2.2. S-balanced configurations have been introduced in the late nineties by Albouy and Chenciner, see [1]. There, and also in [13], the matrix S is supposed to be minus the square of a skew-symmetric matrix. As it turns out, our definition of S-balanced configurations is completely equivalent to that in [1]. Indeed, after replacing the standard basis of $\mathbb{R}^4$ with a (orthonormal) spectral basis of S, we can suppose S to be in diagonal form. Also, the invariance under scalings of the problem implies that we can suppose S to be of the form considered above.

A key feature of SBC is that they admit a variational characterization: Indeed, a configuration vector $q \in \widehat{\mathbb{S}}$ is a SBC if and only if it is a critical point of the restriction of U to $\widehat{\mathbb{S}}$, which with slight abuse of notation will be hereafter denoted also with U.

The Hessian of $U : \widehat{\mathbb{S}} \to \mathbb{R}$ at a critical point q is the quadratic form on $T_q\widehat{\mathbb{S}}$ represented, with respect to the mass-scalar product $\langle M\cdot, \cdot\rangle$, by the $(4n \times 4n)$-matrix

$$H(q) = M^{-1}D^2U(q) + U(q)\widehat{S}.$$

A straightforward computation shows that (i, j)-th entry of $D^2U(q)$ is given by

$$D_{ij} = \frac{m_i m_j}{r_{ij}^3}\big(I_4 - 3u_{ij}u_{ij}^T\big), \quad \text{for } i \neq j, \qquad D_{ii} = -\sum_{j\neq i} D_{ij},$$

where as usual one sets

$$r_{ij} := |q_i - q_j|, \quad u_{ij} := \frac{q_i - q_j}{|q_i - q_j|}.$$

As we already observed, Eq. (2.4) is $S^1 \times S^1$-invariant, and hence SBC always appear in $S^1 \times S^1$-families. In particular, the Hessian $H(q)$ is always degenerate as a quadratic form. Since we do not want to work in a setting where a group of symmetries is acting, we proceed as follows using what in [2] we called the *reduction to (H1) argument*: the submanifold

$$\mathcal{P}_s := \Big\{ q \in \widehat{\mathbb{S}} \;\Big|\; q_k \in \{0\} \times \mathbb{R}^2 \times \{0\},\ \forall k = 1, ..., n \Big\} \subset \widehat{\mathbb{S}}$$

of planar configurations in the plane $\{0\} \times \mathbb{R}^2 \times \{0\}$ is invariant under the gradient flow of U (actually, any submanifold of planar configurations in some coordinate plane does). Ignoring all vanishing components, thus identifying $\{0\} \times \mathbb{R}^2 \times \{0\}$ with $\mathbb{R}^2$, we see that (2.4) on $\mathcal{P}_s$ reads

$$M^{-1}\nabla U(q) + U(q) \cdot \mathrm{diag}\left(\begin{pmatrix} s & 0 \\ 0 & 1 \end{pmatrix}, ..., \begin{pmatrix} s & 0 \\ 0 & 1 \end{pmatrix} \right) q = 0, \tag{2.5}$$

where with slight abuse of notation we denote the mass-matrix on $\mathcal{P}_s$ and the restriction of U to $\mathcal{P}_s$ again with M and U respectively. The main advantage of such a reduction argument is that (2.5) is no longer invariant under the $S^1 \times S^1$-action, but rather only under the action of the discrete group $\mathbb{Z}_2 \times \mathbb{Z}_2$ given by reflections along the main axes. In particular, solutions of (2.5), which are nothing else but planar SBC contained in the plane $\{0\} \times \mathbb{R}^2 \times \{0\} \subset \mathbb{R}^4$, do not a priori come in continuous families, but just in quadruples. We shall also observe that considering planar configurations in $\mathbb{R}^2\times\{(0,0)\}$ or in $\{(0,0)\}\times\mathbb{R}^2$ does not lead to anything interesting, since for such configurations (2.4) reduces to the (normalized) central configurations equation, whereas considering any other plane spanned by $\{v_1, v_2\}$, where $v_1 \in \mathbb{R}^2 \times \{(0,0)\}$ and $v_2 \in \{(0,0)\} \times \mathbb{R}^2$, do not produce different SBC by the $S^1 \times S^1$-invariance of (2.4).

Without further mentioning it, we will hereafter only consider SBC which are contained in $\{0\}\times\mathbb{R}^2\times\{0\}\cong\mathbb{R}^2$ (in other words, we will consider only solutions of (2.5)), and refer to them simply as SBC. Starting point for the results in [2], and for the results of the present paper as well, is a careful study of the inertia indices of collinear SBC, in shorthand notation CSBC. As one readily sees from (2.5), CSBC must be contained in one of the two coordinate axes: we will henceforth call $s-$ CSBC those CSBC which are contained in $\mathbb{R}\times\{0\}\subset\mathbb{R}^2$, and $1-$ CSBC those CSBC which are contained in $\{0\}\times\mathbb{R}\subset\mathbb{R}^2$.

A straightforward computation shows that, after rearranging properly the coordinates of the $s-$CSBC q, we have the following block decomposition of the Hessian matrix

$$H(q) = \begin{pmatrix} -2M^{-1}B(q) & 0 \\ 0 & M^{-1}B(q) \end{pmatrix} + \begin{pmatrix} sU(q)I_n & 0 \\ 0 & U(q)I_n \end{pmatrix},$$

where $B(q)$ is the $(n\times n)$-matrix whose (i,j)-th entry is given by

$$b_{ij}(q) = \frac{m_i m_j}{r_{ij}^3}, \quad b_{ii} = -\sum_{j\neq i} \frac{m_i m_j}{r_{ij}^3}.$$

Similarly, we have the following block decomposition of the Hessian matrix at any $1-$ CSBC $\widehat{q}$:

$$H(\widehat{q}) = \begin{pmatrix} -2M^{-1}B(\widehat{q}) & 0 \\ 0 & M^{-1}B(\widehat{q}) \end{pmatrix} + \begin{pmatrix} U(\widehat{q})I_n & 0 \\ 0 & sU(\widehat{q})I_n \end{pmatrix}.$$

Finally, we shall notice that $1-$ CSBC are actually normalized collinear central configurations, whereas $s-$ CSBC are obtained by scaling normalized collinear central configurations by a factor $1/\sqrt{s}$. Putting these facts together, we proved in [2, Section 2.2] the following result about the inertia indices of CSBC. In what follows we denote by $\iota^0(\widehat{q})$, $\iota^-(\widehat{q})$, $\iota^+(\widehat{q})$ the nullity, Morse index, and Morse coindex respectively of a $1-$ CSBC $\widehat{q}$, with

$$\eta_k(\widehat{q}) < \cdots < \eta_1(\widehat{q}) < \eta_0(\widehat{q}) := -U(\widehat{q}) < 0$$

the distinct eigenvalues of the matrix $M^{-1}B(\widehat{q})$, where $\widehat{q}$ is a fixed $1-$CSBC, and by $\alpha_k(\widehat{q}),...,\alpha_1(\widehat{q}),\alpha_0(\widehat{q}) = 1$ the corresponding multiplicities. Even if the eigenvalues of $M^{-1}B(\widehat{q})$ and their multiplicities depend on the choice of $\widehat{q}$ in general, for the sake of readability we will hereafter drop the dependence on $\widehat{q}$.

Proposition 2.3. *For any $s > 1$, the inertia indices of any $s-$ CSBC q are given by*

$$\iota^0(\widehat{q}) = 0, \quad \iota^+(\widehat{q}) = n-2, \quad \iota^-(\widehat{q}) = n-1.$$

For any $1-$ CSBC $\widehat{q}$ we have:

1. *If* $-\dfrac{\eta_j}{U(\widehat{q})} < s < -\dfrac{\eta_{j+1}}{U(\widehat{q})}$, *for some* $j \in \{0, ..., k-1\}$, *then*

$$\iota^0(\widehat{q}) = 0, \quad \iota^+(\widehat{q}) = n - 2 + \sum_{i=0}^{j} \alpha_i, \quad \iota^-(\widehat{q}) = n - 1 - \sum_{i=0}^{j} \alpha_i.$$

2. *If* $s = -\dfrac{\eta_j}{U(\widehat{q})}$ *for some* $j \in \{1, ..., k\}$, *then*

$$\iota^0(\widehat{q}) = \alpha_j, \quad \iota^+(\widehat{q}) = n - 2 + \sum_{i=0}^{j-1} \alpha_i, \quad \iota^-(\widehat{q}) = n - 1 - \sum_{i=0}^{j} \alpha_i.$$

   In particular, $\widehat{q}$ *is a degenerate critical point of* U.

3. *If* $s > -\dfrac{\eta_k}{U(\widehat{q})}$, *then*

$$\iota^0(\widehat{q}) = 0, \quad \iota^+(\widehat{q}) = 2n - 3, - \quad \iota^-(\widehat{q}) = 0.$$

   In particular, $\widehat{q}$ *is a local minimum of* U.

From the proposition above we can easily deduce several facts: First, CSBC are generically non-degenerate. Second, the inertia indices of $s-$CSBC do not depend on s, whereas those of $1-$ CSBC strongly do. Even more, the Morse index of a $1-$CSBC $\widehat{q}$ jumps at precisely characterized values of s which only depend on the spectrum of the matrix $M^{-1}B(\widehat{q})$ and on the value of the Newtonian potential at $\widehat{q}$. This will enable us in the next section to show the existence of bifurcations of critical points of $\widehat{U}$ starting from $1-$ CSBC.

We shall also notice that in general we have no information about the eigenvalues of the matrix $M^{-1}B(\widehat{q})$ and their multiplicities, besides the fact that $-U(\widehat{q})$ is the largest non-zero eigenvalue and that it is simple. However, it is reasonable to believe that, for generic choice of the masses $m_1, ..., m_n > 0$, all eigenvalues of $M^{-1}B(\widehat{q})$ are simple for any $1-$ CSBC $\widehat{q}$.

Using Proposition 2.3 and the classical Morse inequalities, in [2, Section 4] we gave the following lower bounds on the number of non collinear SBC assuming that all SBC are non-degenerate.

Theorem 2.4. *Assuming that all* SBC *are non-degenerate, the following lower bounds hold:*

1. *If* $s > \max\left\{ -\dfrac{\eta_k(\widehat{q})}{U(\widehat{q})} \;\middle|\; \widehat{q} \text{ is a } 1-\mathrm{CSBC}\right\}$, *then there are at least*

$$3n! - 2(n-1)! - 2$$

   non collinear SBC.

2. *In all other cases, there are at least*

$$n! - 2(n-1)!$$

   non-collinear SBC.

As a corollary, for every $m_1, ..., m_n > 0$ and $s > 1$ such that all SBC *are non-degenerate, we have at least $n! - 2(n-1)!$ relative equilibria of the form*

$$q(t) = \begin{pmatrix} e^{i\sqrt{s}t} & 0 \\ 0 & e^{it} \end{pmatrix} \cdot q, \quad q \text{ is a SBC},$$

for the n-body problem in $\mathbb{R}^4$ which are not induced by central configurations.

The lower bound given in Part 2 can be significantly improved by considering different cases and implementing the asymptotic estimates on the coefficients of the Poincaré polynomial of $\widehat{\mathbb{S}}$ proved in [2, Section 3]. We refrain to do it here to keep the exposition as simple as possible.

A celebrated result of Moeckel, known as the 45°-Theorem, states that for $s = 1$ (which we recall, corresponds to the central configurations case) the manifold of configurations which are collinear along some line is an attractor for the gradient flow of U restricted to $\widehat{\mathbb{S}}$; in particular, collinear central configurations are isolated, and an isolating set is given by the space of configurations for which the "collinearity angle" is less or equal to 45°. This can be seen as a global version of the fact that the Morse index of collinear central configurations is always as large as possible, namely $(d-1)(n-2)$ for the n-body problem in $\mathbb{R}^d$. In [2, Section 5] we generalized the 45°-theorem to SBC, proving (in the setting considered in the present paper) that the manifold of configurations which are collinear along the x-axis is an attractor for the gradient flow of U restricted to $\widehat{\mathbb{S}}$. For the general 45°-theorem for SBC we refer to [2, Theorem 5.6].

Theorem 2.5. *[45°-theorem for* SBC*] The manifold*

$$\Big\{q = (q_1, ..., q_n) \in \mathcal{P} \mid q_k \in \mathbb{R} \times \{0\},\ \forall k = 1, ..., n\Big\} \subset \mathcal{P}$$

is an attractor for the gradient flow of U. More precisely, the collinearity function

$$\theta(q) := \max_{i \neq j} \angle(q_i - q_j, \partial_x)$$

is a Lyapunov function for the gradient vector field of U on the set $\{q \in \mathcal{P} \mid 0 < \theta(q) \leq 45°\}$.

We shall notice that a verbatim generalization of Moeckel's 45°-theorem to SBC is not possible since $1 -$ CSBC are, for suitable values of $s > 1$, local minima of U, and actually there is absolutely no reason why the manifold of configurations which are collinear along some line in $\mathbb{R}^2$ should be invariant under the gradient flow of U. In Sect. 4 we will strengthen these observations by showing that, for increasing value of $s > 1$, we can find families of critical points of U bifurcating from the set of $1 -$ CSBC.

We finish this section recalling that the 45°-theorem for SBC can be used to improve the lower bounds given in Theorem 2.4 on the number of relative equilibria in $\mathbb{R}^4$ assuming non-degeneracy; for more details we refer to [2, Section 6]. We shall notice that here, unlike in Theorem 2.4, we are not able to exclude that such relative equilibria come from central configurations. Nevertheless, the result is still noteworthy since the lower bound that we obtain is larger than the largest known lower bound on the number of planar central configurations, see [10].

Theorem 2.6. *For $s > 1$ fixed, if all* SBC *are non-degenerate there are at least*

$$n!\left(1+\frac{1}{n}+\frac{3}{2}\sum_{j=3}^{n}\frac{1}{j}\right)$$

relative equilibria of the form

$$q(t) = \begin{pmatrix} e^{i\sqrt{s}t} & 0 \\ 0 & e^{it} \end{pmatrix} \cdot q, \quad q \text{ is a SBC},$$

for the n-body problem in $\mathbb{R}^4$.

3. A brief recap on the spectral flow in finite dimension

The spectral flow is an integer-valued homotopy invariant of paths of selfadjoint Fredholm operators introduced by Atiyah, Patodi and Singer in the seventies in connection with the *eta-invariant* and *spectral asymmetry*. In this section, we briefly recall the definition and the basic properties of the spectral flow in a finite dimensional setting. An elementary and self-contained introduction to the spectral flow for bounded selfadjoint Fredholm operators in infinite dimensional real Hilbert spaces can be found in [5], whilst a quick recap and description of the spectral flow in the more general setting of paths of selfadjoint unbounded Fredholm operators having fixed domain appears in the beautiful paper [19]. For further approaches we refer the interested reader to [6–8] and references therein.

Let $(H, \langle\cdot,\cdot\rangle)$ be an Euclidean space and denote by $\mathcal{L}_{sym}(H)$ the vector space of all linear maps $T : H \to H$ that are self-adjoint with respect to $\langle\cdot,\cdot\rangle$. Roughly speaking, the SPECTRAL FLOW $\mathrm{sf}(L_t, t \in [a,b])$ of a continuous path $L : [a,b] \to \mathcal{L}_{sym}(H)$ is the number of negative eigenvalues of L_a that become positive minus the number of positive eigenvalues of L_a that become negative as the parameter t runs from a to b. In other words, the spectral flow measures the net change of eigenvalues crossing 0 and can be interpreted as a sort of generalized signature. This informal description can be made rigorous in very many different ways.

Definition 3.1. Let $L : [a,b] \to \mathcal{L}_{sym}(H)$ be a continuous path of self-adjoint operators having invertible endpoints. We term SPECTRAL FLOW OF L on the interval $[a,b]$ the integer

$$\mathrm{sf}(L_t, t \in [a,b]) := \iota^-(L_a) - \iota^-(L_b)$$

where ι^- denotes the number of negative eigenvalues.

A path of operators having invertible ends will be usually referred to as ADMISSIBLE. Under the non-degeneracy assumption on the endpoints (which is, for different reasons, always assumed throughout the paper) the (RHS) in the equation defining the spectral flow, can be equivalently written as

$$\frac{1}{2}[\mathrm{sgn}(L_a) - \mathrm{sgn}(L_b)],$$

thus pointing out why the spectral flow can be though of as a generalized signature. In this respect we observe that, assuming more regularity on the path L (for instance, that L is at least $\mathscr{C}^1$) it is possible to prove that the local contribution to the spectral flow is provided by the signature of a quadratic form (usually called CROSSING FORM). More precisely, if $t_0 \in (a, b)$ is a CROSSING INSTANT, meaning that $\ker L(t_0) \neq \{0\}$, and if the restriction of the derivative of the path onto $\ker L(t_0)$ is non-degenerate as a quadratic form, then the spectral flow across the instant t_0 can be computed as follows

$$\mathrm{sf}(L_t, t \in [t_0 - \delta, t_0 + \delta]) = \mathrm{sgn}\ L'(t_0)|_{\ker L(t_0)}.$$

Here below we list some properties of the spectral flow that will be used in this paper. In what follows, every path of self-adjoint operators is assumed to be continuous.

1. NORMALIZATION. Let $L : [a, b] \to \mathrm{GL}_{sym}(H)$ be a path of invertible operators. Then
$$\mathrm{sf}(L_t, t \in [a, b]) = 0.$$
2. INVARIANCE UNDER COGREDIENCE. If $L : [a, b] \to \mathcal{L}_{sym}(H)$ is admissible, then for any $M : [a, b] \to \mathrm{GL}_{sym}(H)$ we have
$$\mathrm{sf}(L_t, t \in [a, b]) = \mathrm{sf}(M_t^* L_t M_t, t \in [a, b]).$$
3. CONCATENATION. For $c \in [a, b]$, if $L : [a, b] \to \mathcal{L}_{sym}(H)$ is admissible on both $[a, c]$ and $[c, b]$, then
$$\mathrm{sf}(L_t, t \in [a, b]) = \mathrm{sf}(L_t, t \in [a, c]) + \mathrm{sf}(L_t, t \in [c, b])$$
4. HOMOTOPY INVARIANCE PROPERTY. If $H : [0, 1] \times [a, b] \to \mathcal{L}_{sym}(H)$ is such that the path $t \mapsto H(s, t)$ is admissible for each $s \in [0, 1]$, then
$$\mathrm{sf}(H(s, t), t \in [a, b]) = \mathrm{sf}(H(0, t), t \in [a, b]), \quad \forall s \in [0, 1].$$

We finish this section observing that in Definition 3.1 we required the path L to have invertible endpoints. This assumption can be removed by properly choosing the contribution of the endpoints. In this more general setting, the spectral flow will be homotopy invariant with respect to end-points or, more precisely, end-points stratum homotopy invariant, meaning that the end-points are allowed to vary without changing the nullity.

4. Bifurcations of collinear S-balanced configurations

In this section, we prove an abstract bifurcation result from the trivial branch of a one parameter $\mathscr{C}^2$-family of functions on a finite-dimensional manifold and then apply it to the study of bifurcations S-balanced configurations. We start by introducing the natural geometric framework and some preliminary definitions.

Definition 4.1. A SMOOTH FAMILY OF FINITE-DIMENSIONAL REAL SMOOTH MANIFOLDS $(X_\lambda)_{\lambda\in I}$ parameterized by the interval $I := [a,b] \subset \mathbb{R}$ is a family of manifolds of the form $X_\lambda = p^{-1}(\lambda)$ where $p : X \to I$ is a smooth submersion of a manifold X onto I.

For a smooth family (X_λ) as above, X_λ is a codimensional one submanifold of X for every $\lambda \in I$. For each $x \in X_\lambda$, we have that $T_x X_\lambda = \ker Dp_x$ and

$$T^v X := \{\ker Dp_x | x \in X\}$$

is a vector subbundle of the tangent bundle TX.

A smooth function $F : X \to \mathbb{R}$ defines a smooth family of functions $F_\lambda : X_\lambda \to \mathbb{R}$ by restriction to the fibers of p. We assume that there exists a smooth section $\sigma : I \to X$ of p such that, for every $\lambda \in I$, $\sigma(\lambda)$ is a critical point of F_λ, and in what follows we refer to such a σ as a TRIVIAL BRANCH OF CRITICAL POINTS.

Definition 4.2. We term $\lambda_* \in I$ a BIFURCATION INSTANT FROM THE TRIVIAL BRANCH $\sigma(I)$ if there exists a sequence $\lambda_n \to \lambda_*$ and a sequence $(x_n)_{n\in\mathbb{N}} \subset X$ converging to $\sigma(\lambda_*)$ such that $p(x_n) = \lambda_n$ and each x_n is a critical point for F_{λ_n} not belonging to $\sigma(I)$.

Notation 4.3. We hereafter denote the Hessian of F_λ at the critical point $\sigma(\lambda)$ by h_λ.

The family of Hessians $(h_\lambda)_{\lambda\in I}$ defines a smooth function h on the total space of the pull-back bundle $\mathcal{H} = \sigma^*(T^v X)$ of the vertical tangent bundle. Hence, the function $h : \mathcal{H} \to \mathbb{R}$ defines a family of (generalized) quadratic forms h_λ defined on $T^v X_\lambda$. Using the notation introduced in Sect. 3, we say that the path $\lambda \mapsto h_\lambda$ is admissible if h_a, h_b are non-degenerate. Denoting by $\mathrm{sf}(h_\lambda, \lambda \in [a,b])$ the spectral flow of the path h, we have the following

Theorem 4.4. *If $h = (h_\lambda)_{\lambda\in[a,b]}$ is admissible, then there exists at least one bifurcation instant $\lambda_* \in (a,b)$ of critical points of F from the trivial branch. Moreover, if $\ker h_\lambda \neq \{0\}$ only for finitely many λ, then there are at least*

$$\frac{\left|\,\mathrm{sf}(h_\lambda, \lambda \in [a,b])\right|}{m}$$

distinct bifurcation instants in (a,b) where $m := \max\,\{\dim\ker h_\lambda\}$.

Remark 4.5. The SINGULAR SET $\Sigma(h) := \{\lambda \in I |\ker h_\lambda \neq 0\}$ is finite as soon as the data are analytic, which often occurs in the applications and is indeed the case in the situation considered in the present paper.

Proof. We split the proof into three steps:

Step 1: (Reduction to a fixed Euclidean space) By the vector bundle neighborhood theorem there exists a trivial Euclidean bundle $\mathcal{E} = I \times H$ over $I = [a,b]$ and a fiber preserving smooth map $\psi : \mathcal{E} \to X$ such that $\psi(\lambda, 0) = \sigma(\lambda)$ for every $\lambda \in [a,b]$, and ψ is a diffeomorphism of $\mathcal{E}$ into an open neighborhood $\mathcal{O}$ of $\sigma(I)$ in X.

Let $\widetilde{F} : I \times H \to \mathbb{R}$ be the map defined by composition $\widetilde{F} := F \circ \psi$. So, the map $\widetilde{F}$ defines a smooth one parameter family of functions on H. Since ψ is a fiber preserving diffeomorphism, $u \in H$ is a critical point of $\widetilde{F}_\lambda$ if and only if $x = \psi_\lambda(u)$ is a critical point of F_λ. In particular, 0 is a critical point of $\widetilde{F}_\lambda$ for each $\lambda \in I$.

The Hessian $\widetilde{h}_\lambda$ of $\widetilde{F}_\lambda$ at 0 is given by $\widetilde{h}_\lambda(\zeta) = h_\lambda(D_0\psi_\lambda)([\zeta])$. By the cogredience and normalization properties of the spectral flow, see Section 3, we get that

$$\mathrm{sf}(\widetilde{h}_\lambda, \lambda \in I) = \mathrm{sf}(h_\lambda, \lambda \in I) = \mathrm{sf}(L_\lambda, \lambda \in I) = \iota^-(L_a) - \iota^-(L_b)$$

where $L : I \to L^s(H)$ is a smooth path of self-adjoint operators representing the quadratic form $\widetilde{h}$ with respect to the scalar product of H, namely $\widetilde{h}_\lambda(u) = \langle L_\lambda u, u\rangle$ for every $u \in H$.

Step 2: (Non-vanishing spectral flow implies bifurcation) We assume by contradiction that

$$\mathrm{sf}(L_\lambda, \lambda \in I) = \iota^-(L_a) - \iota^-(L_b) \neq 0$$

and that there are no non-trivial critical points of the path $\lambda \mapsto \widetilde{F}_\lambda$ close to the trivial branch. Then, there exists $\delta > 0$ such that $0 \in H$ is the only critical point of $\widetilde{F}_\lambda$ on B_δ for every $\lambda \in I$, where with $B_\delta \subset H$ we denote the open ball with radius δ around the origin. Without loss of generality we can suppose $\widetilde{F}_\lambda(0) = 0$ for every $\lambda \in I$. For any $\lambda \in I$ and any non-negative integer k, let $C_k(\widetilde{F}_\lambda, 0)$ be the k-th local homology group associated to the isolated critical point 0 of the functional $\widetilde{F}_\lambda$

$$C_k(\widetilde{F}_\lambda, 0) := H_k(\widetilde{F}^0_\lambda \cap B_\delta, \widetilde{F}^0_\lambda \cap B_\delta \setminus \{0\})$$

where, as usually, $H_k(\cdot,\cdot)$ denote the k-th relative singular homology group with integer coefficients, and

$$\widetilde{F}^0_\lambda := \{x \in H | \widetilde{F}_\lambda(x) \leq 0\}$$

denotes the sublevel set of $\widetilde{F}_\lambda$. Since by assumption 0 is an isolated critical point of F_λ for every $\lambda \in I$, for each $k \in \mathbb{N}_0$ the rank of the k-th local homology group is independent of λ. Moreover, the admissibility of the path implies in virtue of the Morse Lemma that

$$C_k(\widetilde{F}_a, 0) = \begin{cases} \mathbb{Z} & \text{if } k = \iota^-(L_a), \\ 0 & \text{otherwise}, \end{cases} \qquad C_k(\widetilde{F}_b, 0) = \begin{cases} \mathbb{Z} & \text{if } k = \iota^-(L_b). \\ 0 & \text{otherwise}. \end{cases}$$

However, since the spectral flow on the interval I does not vanish, we have that $\iota^-(L_a) \neq \iota^-(L_b)$, in contradiction with the fact that the local homology groups do not depend on λ.

Step 3: (Estimate from below on the number of bifurcation instants) To prove this last assertion we proceed as follows. By assumption, there are only finite many instants $a < \lambda_1 < \ldots < \lambda_k < b$ for which the kernel of L_λ is non-zero. Therefore, we can find $\delta > 0$ such that for every $j = 1, \ldots, k$, the intervals $I_j := [\lambda_j - \delta, \lambda_j + \delta]$ are pairwise disjoint and L_λ is non-degenerate at the endpoints of all such intervals. By the additivity property of the spectral flow, we get that

$$\mathrm{sf}(L_\lambda, \lambda \in I) = \sum_{j=1}^{k} \mathrm{sf}(L_\lambda, \lambda \in I_j) = \sum_{j=1}^{k} \left[\iota^-(L_{\lambda_j-\delta}) - \iota^-(L_{\lambda_j+\delta})\right]$$

Since $\dim\ker L_\lambda \le m$ it follows that

$$\left|\iota^-(L_{\lambda_j-\delta}) - \iota^-(L_{\lambda_j+\delta})\right| \le m.$$

Summing up, we immediately get that

$$\left|\,\mathrm{sf}(L_\lambda, \lambda \in I)\right| \le \sum_{j=1}^{k} \left|\,\mathrm{sf}(L_\lambda, \lambda \in I_j)\right| \le dm$$

where d denotes the number of non-vanishing terms in the sum. Since by Step 2 any interval I_j having non-vanishing spectral flow contributes to the total number of bifurcation instants at least by 1, we immediately get that there must exist at least $d = |\,\mathrm{sf}(L_\lambda, \lambda \in I)|/m$ bifurcation points on the interval I. □

Remark 4.6. A similar result holds in infinite dimensional separable Hilbert spaces for a path of (bounded) self-adjoint Fredholm operators even in the strongly indefinite case, meaning that the Fredholm operators are compact perturbations of an invertible operator having an infinite dimensional positive and negative eigenspace. In this case in fact both the Morse index and coindex are infinite. For further details we refer the interested reader to [12,16–18] and references therein for this more general functional analytic setting with application e.g. to semi-Riemannian geodesics and conjugate points.

Theorem 4.4 will be the key ingredient for our bifurcation result for SBC, where the bifurcation parameter is precisely the parameter s appearing in the matrix S. Thus, for every $s > 1$ we consider the function $U : \mathcal{P}_s \to \mathbb{R}$. By the discussion provided in Sect. 2, SBC and in particular CSBC are critical points of U. The latter turn out to be generically non-degenerate, see Proposition 2.3, meaning that the Hessian of U at any CSBC is non-degenerate as a quadratic form for all but finitely many values of s. Before stating the bifurcation theorem for SBC, we shall finally observe that $1-$CSBC, which we recall are nothing else but normalized collinear central configurations along the y-axis, are solutions of (2.5) independently of $s > 1$. In particular, any constant one-parameter family $(\widehat{q}_s = \widehat{q})_{s>1}$, with $\widehat{q}$ a fixed $1-$CSBC, provides a trivial branch of critical points.

Theorem 4.7. *For $s_1, s_2 > 1$, let $J := [s_1, s_2]$ and let $(\widehat{q}_s)_{s\in J}$ be the trivial one-parameter family of* SBC *given by a fixed* $1-$CSBC *$\widehat{q}$, that is $\widehat{q}_s = \widehat{q}$ for every $s \in J$. Setting*

$$\alpha_*(\widehat{q}) := \max\{\alpha_j(\widehat{q}) | j = 1, \dots, k\},$$

for s_1 sufficiently close to 1 and s_2 large enough there exist at least

$$\left\lfloor \frac{n-2}{\alpha_*(\widehat{q})} \right\rfloor$$

bifurcation instants from $\widehat{q}$. As a corollary, for s_1 sufficiently close to 1 *and s_2 large enough there are at least*

$$n! \cdot \left\lfloor \frac{n-2}{\alpha_*} \right\rfloor, \qquad \alpha_* := \max\{\alpha_*(\widehat{q}) \mid \widehat{q} \text{ is a } 1-\mathrm{CSBC}\},$$

bifurcation instants from the trivial families of $1-\mathrm{CSBC}$.

Remark 4.8. Even if we have no information about the α_j's in general, it is reasonable to believe that they are all equal one for generic choice of the masses $m_1, ..., m_n$. In this case, Theorem 4.7 would provide the existence of at least $n!(n-2)$ bifurcation instants, namely at least $n-2$ bifurcation instants for each choice of the ordering of the masses along the y-axis.

Proof. We denote by $\pi : \mathcal{P} \to J$ the (trivial) ellipsoid bundle (topologically, a sphere bundle) over J whose fiber $\pi^{-1}(\{s\}) = \mathcal{P}_s$ is the collision free configuration sphere. Notice that $\mathcal{P}_s$ depends on the parameter s; in particular, for $s = 1$ we obtain a round sphere in the mass metric, which for increasing s is deformed into an ellipsoid having its major axes in the directions corresponding to the eigenvalue 1 of the matrix S.

The Newtonian potential U defines a smooth bundle map $\mathcal{U}$ whose restriction $\mathcal{U}_s$ to each fiber $\pi^{-1}(\{s\})$ is precisely U, and for each $s \in J$ the configuration vector $\widehat{q}_s = \widehat{q}$ is a $1-\mathrm{CSBC}$, hence a critical point of $\mathcal{U}_s$. By Proposition 2.3, we infer that the path $s \mapsto \widehat{q}_s$ is admissible (meaning that the associated path $s \mapsto h_s$ of quadratic forms pointwise defined by the Hessian of U at $\widehat{q}_s$ is admissible) as soon as

$$1 < s_1 < -\frac{\eta_1(\widehat{q})}{U(\widehat{q})}, \qquad s_2 > -\frac{\eta_k(\widehat{q})}{U(\widehat{q})}.$$

Moreover, by setting $j = 0$ in Proposition 2.3, Part 1, we get that the $\iota^-(\widehat{q}_{s_1}) = n-2$, whereas using Item iii) of Proposition 2.3, Part 2, we get that $\iota^-(\widehat{q}_{s_2}) = 0$. In particular, the spectral flow of the path $s \mapsto h_s$ is easily computed to be

$$\mathrm{sf}(h_s, s \in [s_1, s_2]) = \iota^-(h_{s_1}) - \iota^-(h_{s_2}) = n-2$$

where $\iota^-(h_{s_1}) := \iota^-(\xi_{s_1})$ and $\iota^-(h_{s_2}) := \iota^-(\xi_{s_2})$. Theorem 4.4 now implies the claim, observing that the kernel of h_s is non-trivial only in correspondence of the eigenvalues $\eta_j(\widehat{q})$ whose multiplicity is $\alpha_j(\widehat{q})$. □

5. Some explicit examples for $n = 3$

In this section we use numerical computations to study the non-trivial branches bifurcating from a trivial branch of $1-\mathrm{CSBC}$. First, we introduce the continuation method used to compute curves of solutions of a system of nonlinear equations depending on a real parameter, then we describe how to use it to follow the bifurcations branching from a trivial branch of $1-\mathrm{CSBC}$. Finally, we provide some examples for $n = 3$.

5.1. The continuation method

Let $F : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ be a differentiable function, and denote with $q \in \mathbb{R}^n$ the spatial component and with $s \in \mathbb{R}$ the parameter. Take $(q_i, s_i)^T \in \mathbb{R}^n \times \mathbb{R}$ such that $F(q_i, s_i) = 0$: the purpose of a continuation method is to find a zero of F for a different value of the parameter s, starting from the known solution at s_i.

To this end, the pair $(q, s)^T$ is displaced by solving the equation

$$\begin{cases} F(q,s) = 0, \\ \left| \begin{pmatrix} q \\ s \end{pmatrix} - \begin{pmatrix} q_i \\ s_i \end{pmatrix} \right|^2 - \delta^2 = 0, \end{cases} \tag{5.1}$$

where $\delta > 0$ is small. A solution of Equation (5.1) can be computed using the Newton method (see e.g. [20]), thus solving at each step a system of equations given by the matrix

$$\begin{bmatrix} \dfrac{\partial F}{\partial q} & \dfrac{\partial F}{\partial s} \\ 2(q - q_i) & 2(s - s_i) \end{bmatrix}.$$

A first guess $(\bar{q}, \bar{s})$ for the Newton method can be constructed starting from the known solution $(q_i, s_i)^T$, and taking a tangent displacement along the curve of solutions. The tangent line can be approximated by using two consecutive solutions $(q_i, s_i)^T$ and $(q_{i-1}, s_{i-1})^T$, hence the first guess can be taken as

$$\begin{pmatrix} \bar{q} \\ \bar{s} \end{pmatrix} = \begin{pmatrix} q_i \\ s_i \end{pmatrix} + \gamma \begin{pmatrix} q_i - q_{i-1} \\ s_i - s_{i-1} \end{pmatrix}, \qquad \gamma = \frac{\delta}{|(q_i, s_i)^T - (q_{i-1}, s_{i-1})^T|}.$$

If not known, an additional solution $(q_{i-1}, s_{i-1})^T$ can be computed from the known solution $(q_i, s_i)^T$ by simply displacing s_i as $s_{i-1} = s_i + \Delta s$, and then solving the equation $G(q) := F(q, s_{i-1}) = 0$ with the Newton method, constructing a first guess using a displacement of q_i.

In the case of S-balanced configurations, the function F reads

$$F(q, s) = M^{-1} \nabla U(q) + U(q) \widehat{S}(s) q,$$

where $\widehat{S}(s)$ is the block-diagonal matrix

$$\widehat{S}(s) = \left(\begin{pmatrix} s & 0 \\ 0 & 1 \end{pmatrix}, \ldots, \begin{pmatrix} s & 0 \\ 0 & 1 \end{pmatrix} \right).$$

To numerically study the bifurcations branching from a trivial branch of $1-$CSBC in the case of three masses, we use the following scheme. Let m_1, m_2, m_3 be given positive masses, and let an ordering of the three masses along the y-axis be fixed. Then,

1. we compute the corresponding $1-$CSBC (which we recall is nothing else but a normalized collinear central configuration), namely using the Newton method to find the unique zero of the Euler polynomial (see e.g. [14]) corresponding to the chosen ordering of the masses.
2. We compute the eigenvalues of the matrix $M^{-1}B(\widehat{q})$, and the value of the potential $U(\widehat{q})$.

3. We compute the (unique in virtue of Item ii) of Proposition 2.3) value $\tilde{s}$ for which $\widehat{q}$ is a degenerate critical point of U.
4. We displace $\tilde{s}$ by $s = \tilde{s} + \Delta s$, and search for a zero of $F(\,\cdot, s)$ using the Newton method. A first guess is constructed by displacing $\widehat{q}$ along the direction of the kernel of $\partial F/\partial q$.
5. Using the previous two solutions, we start the continuation method with respect to the parameter s.

The same scheme can be used also for $n \geq 4$, modifying the first step for the computation of the corresponding $1-$ CSBC. Notice indeed that also in this case there exists a unique $1-$ CSBC for each ordering of the bodies [11]. We plan to numerically investigate this case in future work.

5.2. Results of the computations

We produced a first example using three unitary equal masses. The corresponding trivial branch of $1-$ CSBC has a bifurcation at $\tilde{s} = 2.4$, from which two non-trivial branches originate. On these branches the parameter s decreases, while the three masses move on isosceles configurations until arriving at an equilateral configuration for $s = 1$. The two branches are symmetrical with respect to the y-axis, with the difference that the pair $\{q_2 - q_1, q_3 - q_1\}$ is positively oriented on one branch and negatively oriented on the other one. Moreover, all solutions found are local minima of the potential U. Figure 1 shows one of the branches in the space (x, y, s); the other branch is not displayed for visibility reasons. Observe that, due to the additional symmetries of the problem, changing the ordering of the masses does not yield qualitatively different behaviors of the bifurcations branches. Also, a branch (actually four branches due to the $\mathbb{Z}_2 \times \mathbb{Z}_2$-symmetry) of saddle points originating at $s = 1$ from Lagrange's equilateral triangle is present but not shown in the figure.

We produced a second example using one unitary mass and two smaller equal masses, say $m_1 = 1$, $m_2 = m_3 = \mu$, $0 < \mu < 1$, for the ordering of the masses given by taking the bigger mass outside the segment joining the two smaller ones. Also here we have a unique bifurcation point $\tilde{s}$ from which two non-trivial (again, symmetrical with respect to the y-axis) branches originate. Along each non-trivial branch the parameter s decreases at first until reaching a turning point s_{turn}, where the three masses are placed on the vertices of a symmetric with respect to the y-axis isosceles triangle. After the turning point, the parameter s increases up to the bifurcation value $\tilde{s}$, and the corresponding branch of SBC reaches for $s \uparrow \tilde{s}$ the $1-$ CSBC corresponding to the ordering of the masses in which the two smaller masses are swapped. In the (x, y) plane, the masses m_2 and m_3 appear to rotate around a common point, while the mass m_1 appears to oscillate. The oscillation becomes larger and larger as the value μ approaches 1, and the isosceles configuration at the turning point tends to an equilateral one, which is reached in the limit $\mu \uparrow 1$. On the other hand, the oscillations are very small for $\mu \ll 1$, making the position of m_1 almost constant. Also here, all configurations along the branches are local minima of the potential U. In correspondence of the turning

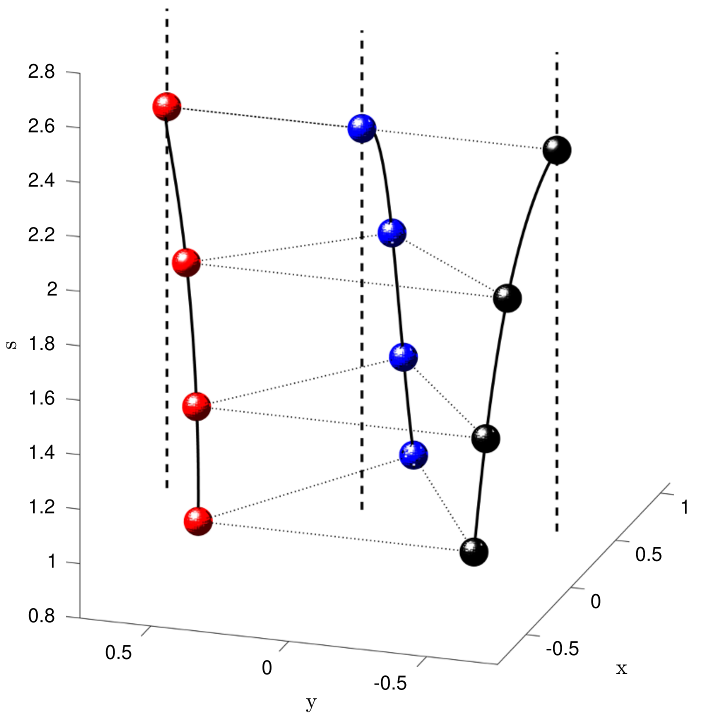


FIGURE 1. The branches of SBC in the case of three equal masses in the space (x, y, s). The three dashed vertical lines represent the trivial branch of $1-$CSBC, while the black and the red thick curves represent the two non-trivial branches originating from the bifurcation point

point s_{turn} we also have two additional branches originating from the non-trivial branches, namely a first one along which the parameter s decreases to 1 reaching in the limit $s \downarrow 1$ the equilateral configuration, and a second one along which the parameter s increases to infinity reaching in the limit $s \to +\infty$ a limit configuration. The critical points along the first secondary branch are local minima of U, whereas along the latter one we have saddle points. The non-trivial branches in the space (x, y, s) corresponding to the cases $\mu = 0.99$ and $\mu = 0.01$ respectively are depicted in Fig. 2. In the limit $\mu \uparrow 1$ the non-trivial branches "tend" to the non-trivial branches depicted in Fig. 1.

In case the unitary mass is placed between the two smaller ones, the behavior is similar to the three equal masses case: the parameter s decreases down to 1, where the masses reach an equilateral configuration.

A third example has been produced using two unitary masses and a smaller one, say $m_1 = m_2 = 1$, $m_3 = \mu$, $0 < \mu < 1$, for the ordering of the masses in which the smaller mass lies outside the segment joining the two bigger ones. Two non-trivial branches, symmetrical with respect to the

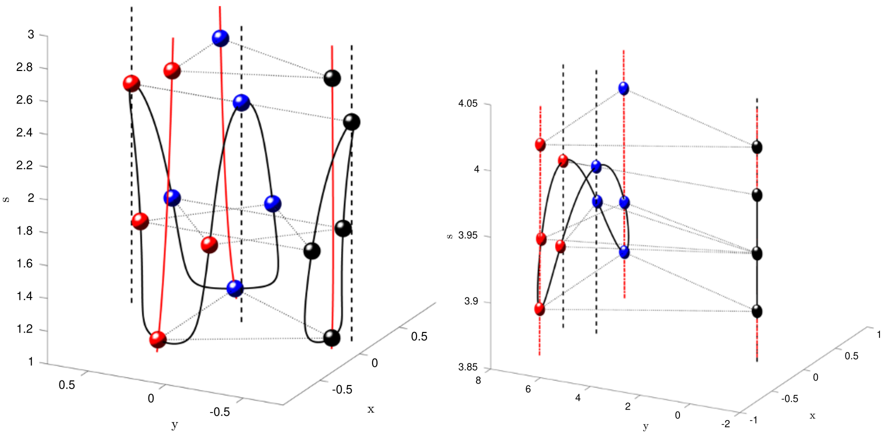


FIGURE 2. The branches of SBC in the case $m_1 = 1$, $m_2 = m_3 = \mu$, $0 < \mu < 1$, in the space (x, y, s). The left panel refers to the case $\mu = 0.99$, while the right one to $\mu = 0.01$. The black particle indicates the mass m_1, while the blue and red particles refer to m_2 and m_3 respectively. As in Fig. 1, the dashed vertical lines represent the trivial branch of $1-$CSBC, whereas the thick black curve represents a branch originating from the bifurcation point. Another branch, in which the masses follow the same curve in the opposite direction, is also present but not shown. The secondary branches originating in correspondence of the turning point are drawn in red

y-axis, originate from the (unique) bifurcation point $\tilde{s}$: on these branches the parameter s initially decreases, and the masses are placed on scalene configurations corresponding to local minima of U. A turning point s_{turn} is reached during the continuation, and the differential of F is singular at s_{turn}. For $\mu \sim 1$, the turning point s_{turn} is close to $s = 1$, and again the non-trivial branches approach the non-trivial branches depicted in Fig. 1 in the limit $\mu \uparrow 1$. On the other hand, the turning point becomes closer and closer to the bifurcation value $\tilde{s}$ for $\mu \to 0$. After the turning point, the parameter s increases to $+\infty$, with the masses still placed on scalene configurations which in this case, however, correspond to saddle points of U. Therefore, we have a jump on the Morse index in correspondence of the turning point s_{turn}. We also looked for secondary branches at the turning point s_{turn}. To construct an initial guess for the Newton method, we displaced the configuration at s_{turn} in two different manners: 1) along the direction of the kernel of $\partial F/\partial q$, and 2) displacing q_1, q_2, q_3 along random directions $v_1, v_2, v_3 \in S^1$, respectively. In both cases, we did not find any other branch. Observe that this does not contradict Theorem 4.4, since the considered branch is not trivial. We plan to examine the occurrence of bifurcation points on non-trivial branches in future work. Following the non-trivial branch after the turning point, we see that the configurations are asymptotic as s grows to infinity to a limit configuration.

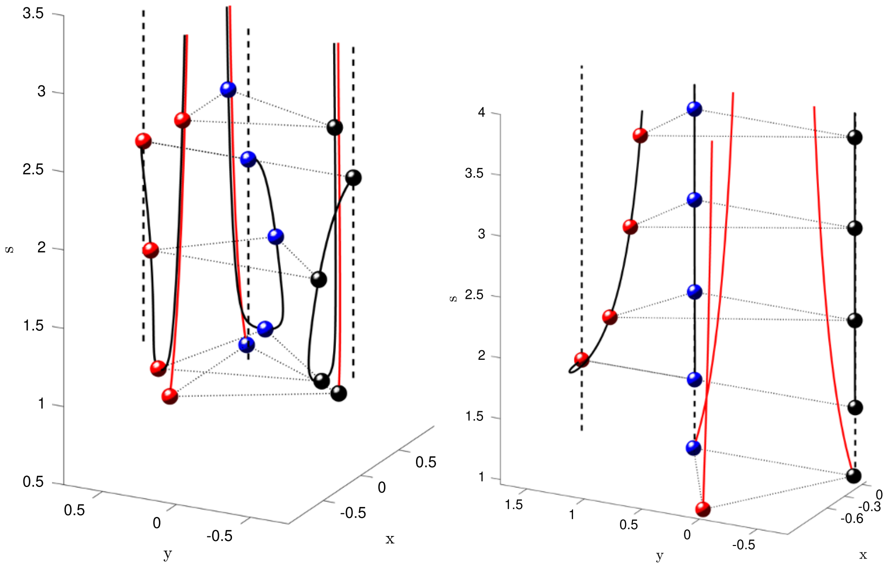


FIGURE 3. The branches of SBC in the case $m_1 = m_2 = 1$, $m_3 = \mu$, $0 < \mu < 1$, in the space (x, y, s). The left panel refers to the case $\mu = 0.99$, while the right panel to the case $\mu = 0.01$. The black particle indicates the mass m_1, while the blue and the red particles refer to m_2 and m_3, respectively. As in Figs. 1 and 2, the dashed vertical line represents the trivial branch, while the thick black curve represent one branch originating from the bifurcation point. Another branch, symmetrical with respect to the y-axis, is also present but not shown. The branch of saddle points originating from the equilateral triangle is drawn in red. The black and red branches are asymptotic for $s \to +\infty$ to different asymptotic configurations

As the left panel in Fig. 3 might suggest, the asymptotic configuration seems to be the limit also of a branch of saddle points originating at $s = 1$ from the equilateral triangle. This is in fact not the case, as it can be easily deduced from the right panel. The branches in the space (x, y, s) corresponding to $\mu = 0.99$ and $\mu = 0.01$ respectively are depicted in Fig. 3.

In the case that the mass m_3 lies between m_1 and m_2, the result is similar to the three equal masses case, with the bifurcation parameter s decreasing to 1, where an equilateral configuration is reached.

Further examples have been produced using three different masses by means of an additional parameter, but no qualitatively different behavior has been observed. More precisely, for fixed parameters $\mu, \lambda < 1$, we considered masses $(m_1, m_2, m_3) = (1, \lambda, \mu)$. As it turns out, the qualitative behavior of the corresponding non-trivial branches only depends on λ: For $\lambda < \min\{1, \mu\}$ we are in situation analogous to the one depicted in Fig. 1, with two branches originating from Euler's configuration on which s decreases to 1 reaching in the limit $s \downarrow 1$ Lagrange's equilateral triangle along minima of U. For $\lambda = \min\{1, \mu\}$, we have the same qualitative behavior as in Fig. 2: along each non-trivial branch the parameter s decreases at first until reaching a turning point. After the turning point, the parameter s increases up to the bifurcation value $\tilde{s}$, and the corresponding branch of SBC reaches for $s \uparrow \tilde{s}$ the $1-$ CSBC corresponding to the ordering of the masses in which the two smaller masses are swapped. In correspondence of the turning point two additional branches originate, namely one along which s decreases reaching for $s = 1$ Lagrange's equilateral triangle on minima of U, and one along which s increases to infinity on saddle points of U. Finally, in the case $\lambda > \min\{\mu, 1\}$ we have the same qualitative behavior as in Fig. 3: The parameter s decreases until a turning point is reached, after which s grows indefinitely and the configuration tends to an asymptotic configuration.

Videos showing how the position of the masses in the (x, y)-plane changes along a non-trivial branch, for the cases discussed in this section, can be found at [4].

5.3. Final comments

We are now in position to make some final comments on the numerical implementations discussed in the previous subsection and on their consequences on the dynamics of the n-body problem in $\mathbb{R}^4$. Before doing that, we shall recall that the bifurcations of SBC we found can be seen as planar SBC in $\mathbb{R}^4$ contained in the plane $\{0\} \times \mathbb{R}^2 \times \{0\} \subset \mathbb{R}^4$, and as such define relative equilibria for the n-body problem in $\mathbb{R}^4$ by

$$\mathbf{q}(t) = \begin{pmatrix} e^{i\sqrt{s}t} & 0 \\ 0 & e^{it} \end{pmatrix} \cdot \begin{pmatrix} 0 \\ x \\ y \\ 0 \end{pmatrix},$$

where $(0, x, y, 0)^T = q$ is any planar SBC contained in $\{0\} \times \mathbb{R}^2 \times \{0\}$.

In Figure 1 we observe for the case of three equal masses a connection between Lagrange's equilateral triangle q_{Lagr} and Euler's collinear configuration q_{Eul} through a branch of SBC which are local minima of U. This yields a continuum of relative equilibria $(\mathbf{q}_s)_{s\in[1,\tilde{s}]}$ for the 3-body problem in $\mathbb{R}^4$

$$\mathbf{q}_s(t) = \begin{pmatrix} e^{i\sqrt{s}t} & 0 \\ 0 & e^{it} \end{pmatrix} \cdot \begin{pmatrix} 0 \\ x_s \\ y_s \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ x_1 \\ y_1 \\ 0 \end{pmatrix} = q_{\text{Lagr}}, \quad \begin{pmatrix} 0 \\ x_{\tilde{s}} \\ y_{\tilde{s}} \\ 0 \end{pmatrix} = q_{\text{Eul}},$$

which is precisely the one found abstractly in [2, Corollary 2.15]. An illustration of $\mathbf{q}_1$ can be found in [15, Figure 9]. For $s > \tilde{s}$ such a continuum can be extended simply by taking $q_s = q_{\text{Eul}}$ for all $s > \tilde{s}$. Finally, for $s \in [1, \tilde{s}) \cap \mathbb{Q}$ or $s > \tilde{s}$, the relative equilibrium $\mathbf{q}_s$ is periodic, and it is quasi-periodic otherwise.

A similar connection can be observed also for arbitrary masses, namely between Lagrange's equilateral triangle and Euler's collinear configuration in which the smaller mass lies between the other two masses. Such connections cannot be found within the class of central configurations, as Euler's configurations are isolated in virtue of Moeckel's 45°-theorem.

As the implementations suggest, changes in the qualitative behavior of the non-trivial branches are to be expected especially when deforming the masses parameters by passing through a configuration of the masses that forces additional symmetries of the problem, namely a configurations of the masses in which at least two of the masses are equal. We plan to investigate this aspect further for larger values of n.

Another interesting group of open questions concerns the stability of relative equilibria generated by a SBC. As already observed, to a given planar CC it is possible to associate a relative equilibrium solution in which all bodies rigidly rotate around their center of mass. In rotating coordinates, this relative equilibrium reduces to an equilibrium (actually, to the CC originating the orbit through a circle action). So, it is quite natural to ask whether there is or not a relation between the linear stability of the relative equilibrium (as a periodic orbit) and the Morse index of the associated CC (in the rotating frame). Several results in this direction are nowadays available in the literature: For instance, in [9] the authors provide a sufficient condition for the linear instability of a relative equilibrium originated by a non-degenerate CC in the plane in terms of the Morse index, which in [3] is generalized to a broader class of singular operators as well as to the case of relative equilibria generated by a possibly degenerate critical point. In this latter case, besides the Morse index, a key role is played by the Jordan normal form associated to the Floquet multiplier 1.

On the other hand, nothing seems to be known about the linear (and spectral) stability for higher-dimensional relative equilibria, see e.g. [13]. In higher-dimension, the situation is indeed much more involved: For relative equilibria which are periodic in time (that is, for $s \in \mathbb{Q}$) we expect, besides Morse index and Jordan normal form, also the parameter s to play a key role in the characterization of the stability properties. For relative equilibria which are quasi-periodic in time (that is, for $s \in \mathbb{R}\backslash\mathbb{Q}$) a similar characterization relating the KAM stability of the relative equilibrium to the inertia indices of the associated SBC should be possible as well. All these questions will be addressed in future work.

Acknowledgements

Luca Asselle is partially supported by the DFG-grant 380257369 "Morse theoretical methods in Hamiltonian dynamics". Marco Fenucci is partially supported by the MSCA-ITN Stardust-R, Grant Agreement n. 813644 under the H2020 research and innovation program.

Funding Open Access funding enabled and organized by Projekt DEAL.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Albouy, A.: Chenciner, Alain Le problème des n corps et les distances mutuelles. (French) [The n-body problem and mutual distances] Invent. Math. **131**(1), 151–184 (1998)

[2] Asselle, L.: Portaluri, Alessandro Morse theory for S-balanced configurations in the Newtonian n-body problem. J. Dyn. Diff. Equ. (2021). https://doi.org/10.1007/s10884-021-10036-y

[3] Barutello, V.L., Jadanza, R.D., Portaluri, A.: Linear instability of relative equilibria for n-body problems in the plane. J. Differ. Equ. **257**(6), 1773–1813 (2014)

[4] Fenucci, Marco Personal Webpage. http://adams.dm.unipi.it/~fenucci/research/sbalanced-3b.html

[5] Fitzpatrick, P.M., Pejsachowicz, J., Recht, L.: Spectral flow and bifurcation of critical points of strongly-indefinite functionals. I. General theory. J. Funct. Anal. **162**(1), 52–95 (1999)

[6] Hu, X.: Portaluri, Alessandro Index theory for heteroclinic orbits of Hamiltonian systems. Calc. Var. Partial Differ. Equ. **56**(6), Paper No. 167, 24 pp (2017)

[7] Hu, X.: Portaluri, Alessandro Bifurcation of heteroclinic orbits via an index theory. Math. Z. **292**(1–2), 705–723 (2019)

[8] Hu, X.: Portaluri, Alessandro; Yang, Ran A dihedral Bott-type iteration formula and stability of symmetric periodic orbits. Calc. Var. Partial Differ. Equ. **59**(2), Paper No. 51, 40 pp (2020)

[9] Hu, X., Sun, S.: Stability of relative equilibria and Morse index of central configurations. C. R. Math. Acad. Sci. Paris **347**(21–22), 1309–1312 (2009)

[10] McCord, C.: Planar central configuration estimates in the N-body problem. Ergodic Theory Dyn. Syst. **16**, 1059–1070 (1996)

[11] Moulton, Forest Ray The Straight Line Solutions of the Problem of n Bodies. Ann. Math. **12**, 1–17 (1910)

[12] Musso, M., Pejsachowicz, J.: Portaluri, Alessandro Morse index and bifurcation of p-geodesics on semi Riemannian manifolds. ESAIM Control Optim. Calc. Var. **13**(3), 598–621 (2007)

[13] Moeckel, R.: Central configurations. Lecture notes given at the Centre de Recerca Matemàtica (CRM), Barcelona, January 27–31 (2014)

[14] Moeckel, R.: Lectures on Celestial Mechanics (especially on Central configurations). Preprint available at http://www-users.math.umn.edu/~rmoeckel/notes/Notes.html

[15] Moeckel, R.: Central configurations. http://www.scholarpedia.org/article/Central_configurations

[16] Piccione, P., Portaluri, A., Tausk, D.V.: Spectral flow, Maslov index and bifurcation of semi-Riemannian geodesics. Ann. Global Anal. Geom. **25**(2), 121–149 (2004)

[17] Portaluri, A.: Waterstraat, Nils Bifurcation results for critical points of families of functionals. Differ. Integral Equ. **27**(3–4), 369–386 (2014)

[18] Portaluri, A.: Waterstraat, Nils A K-theoretical invariant and bifurcation for homoclinics of Hamiltonian systems. J. Fixed Point Theory Appl. **19**(1), 833–851 (2017)

[19] Robbin, J.: Salamon, Dietmar the spectral flow and the Maslov index. Bull. Lond. Math. Soc. **27**(1), 1–33 (1995)

[20] Stoer, J.: Bulirsch, Roland Introduction to Numerical Analysis. Springer (2002)

Luca Asselle
Ruhr-Universität Bochum
Universitätsstraße 150
Bochum 44801
Germany
e-mail: luca.asselle@rub.de

Marco Fenucci
Department of Astronomy, Faculty of Mathematics
University of Belgrade
Studentski trg 16
Belgrade 11000
Serbia
e-mail: marco_fenucci@matf.bg.ac.rs

Alessandro Portaluri
Università degli Studi di Torino
Largo Paolo Braccini, 2
Grugliasco 10095 Turin
Italy
e-mail: alessandro.portaluri@unito.it

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:44
https://doi.org/10.1007/s11784-022-00963-8
Published online May 13, 2022

Journal of Fixed Point Theory and Applications

Relative Hofer–Zehnder capacity and positive symplectic homology

Gabriele Benedetti and Jungsoo Kang

Dedicated to Claude Viterbo on the occasion of his 60th birthday.

Abstract. We study the relationship between a homological capacity $c_{\mathrm{SH}^+}(W)$ for Liouville domains W defined using positive symplectic homology and the existence of periodic orbits for Hamiltonian systems on W: if the positive symplectic homology of W is non-zero, then the capacity yields a finite upper bound to the π_1-sensitive Hofer–Zehnder capacity of W relative to its skeleton and a certain class of Hamiltonian diffeomorphisms of W has infinitely many non-trivial contractible periodic points. En passant, we give an upper bound for the spectral capacity of W in terms of the homological capacity $c_{\mathrm{SH}}(W)$ defined using the full symplectic homology. Applications of these statements to cotangent bundles are discussed and use a result by Abbondandolo and Mazzucchelli in the appendix, where the monotonicity of systoles of convex Riemannian two-spheres in $\mathbb{R}^3$ is proved.

Mathematics Subject Classification. 37J46, 53D40.

Keywords. Floer theory, symplectic homology, Liouville domains, periodic orbits of Hamiltonian systems.

1. Periodic orbits for Hamiltonian systems

In this paper, we prove new existence and multiplicity results for periodic orbits of Hamiltonian systems on Liouville domains using positive symplectic homology. We present our results in Sect. 2, and, to put them into context, we give in this first section a brief (and incomplete) account of previous work on the existence of periodic orbits for Hamiltonian systems that will be relevant to our work. For more details, we recommend the excellent surveys [14,23,24].

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.
With an appendix by Alberto Abbondandolo and Marco Mazzucchelli.

1.1. From calculus of variation to pseudoholomorphic curves

From the end of the 1970s to the beginning of the 1990s, a series of tremendous advances has been made in understanding the existence of periodic solutions for smooth Hamiltonian systems in $\mathbb{R}^{2n}$ with standard symplectic form $\omega_{\mathbb{R}^{2n}}$ using variational techniques. Classically, two settings have been considered.

In the non-autonomous setting, one may consider a smooth Hamiltonian function H which is periodic in time (say with period one) and compactly supported, and study the set $\mathcal{P}_m(H)$ of periodic orbits of the Hamiltonian vector field of H with some integer period $m \in \mathbb{N}$. This is equivalent to looking at periodic points of the Hamiltonian diffeomorphism obtained as the time-one map of the Hamiltonian flow of H. The set $\mathcal{P}_m(H)$ surely contains trivial orbits, namely those which are constant and contained in the zero-set of the Hamiltonian, and the question is if there are non-trivial elements in this set. In 1992, Viterbo proved the following remarkable result.

Theorem 1.1. (Viterbo [63]) *Let $\varphi\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ be a compactly supported Hamiltonian diffeomorphism different from the identity. Then, φ admits a non-trivial one-periodic point and infinitely many distinct non-trivial periodic points.*

In the autonomous setting, the Hamiltonian function H is coercive and independent of time. Thus, it is meaningful to study the set of periodic orbits having arbitrary real period and lying on a regular energy level Σ of H, which is compact by coercivity. Let $j_0\colon \Sigma \hookrightarrow \mathbb{R}^{2n}$ be the inclusion of Σ, and denote the set of periodic orbits of H on Σ by $\mathcal{P}(\Sigma)$. This notation is justified since periodic orbits of H on Σ correspond to closed characteristics of Σ, i.e., embedded circles in Σ which are everywhere tangent to the characteristic distribution $\ker(j_0^*\omega_{\mathbb{R}^{2n}})$. In 1987 Viterbo [62] proved the existence of a closed characteristic for a large class of hypersurfaces Σ in $\mathbb{R}^{2n}$, namely those of contact type (see [36] for the definition), thus confirming the Weinstein conjecture in $\mathbb{R}^{2n}$ [70]. This considerably extends previous results by Weinstein [69] and Rabinowitz [49] for hypersurfaces being convex or starshaped, two properties which are not invariant by symplectomorphisms.

Soon after, it was understood that Viterbo's result is a manifestation of more general phenomena which go under the name of nearby existence and almost existence. In order to explain these phenomena, we consider a thickening of $j_0\colon \Sigma \to \mathbb{R}^{2n}$, namely an embedding $j\colon (-\varepsilon_0, \varepsilon_0) \times \Sigma \to \mathbb{R}^{2n}$ such that $j(0,\cdot) = j_0$. We denote $\Sigma_\varepsilon = j(\{\varepsilon\} \times \Sigma)$ for $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$. For a thickening of j_0, the nearby existence theorem [35] ensures that for a sequence of values $\varepsilon_k \to 0$, there holds $\mathcal{P}(\Sigma_{\varepsilon_k}) \neq \varnothing$. On the other hand, the almost existence theorem [55] yields the stronger assertion that $\mathcal{P}(\Sigma_\varepsilon) \neq \varnothing$ for almost every $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$. Viterbo's result follows from either of these two theorems noticing that a contact, or more generally stable [36], hypersurface admits a thickening such that the characteristic foliation of Σ is diffeomorphic to the one of Σ_ϵ for all $\varepsilon \in (-\varepsilon_0, \varepsilon_0)$. A classical argument by Hofer and Zehnder [36] proves the almost existence theorem by showing the finiteness of the Hofer–Zehnder capacity for bounded domains of $\mathbb{R}^{2n}$ of which we now recall the definition. If $U \subset \mathbb{R}^{2n}$ is a domain possibly with boundary, let $\mathcal{H}(U)$ be the set of Hamiltonians $H\colon U \to (-\infty, 0]$ that vanish outside a compact set of $U\backslash\partial U$ and achieve their minimum on an open set of U. The Hofer–Zehnder capacity of U is defined by

$$c_{\mathrm{HZ}}(U) := \sup_{H\in\mathcal{H}(U)} \big\{ -\min H \,\big|\, \mathcal{P}_{\leq 1}(H) = \mathrm{Crit}(H)\big\} \in (0,\infty],$$

where we used the notation $\mathrm{Crit}(H)$ for the set of critical points of H and $\mathcal{P}_{\leq t}(H)$ for the set of periodic orbits of the Hamiltonian vector field of H with period at most t.

Much of research in Hamiltonian dynamics in the last 30 years has been driven by considering the two settings described above for general symplectic manifolds (M,ω). For instance, one can ask if the nearby or almost existence theorems hold for a given hypersurface $\Sigma \subset M$. To this end, the Hofer–Zehnder capacity, whose definition extends verbatim to open subsets of M, still plays a central role as the almost existence theorem holds for hypersurfaces contained in a domain $U \subset M$ of finite Hofer–Zehnder capacity [46]. It is also sometimes useful to consider a refined quantity, called the π_1-sensitive Hofer–Zehnder capacity, given by

$$c^o_{\mathrm{HZ}}(U) := \sup_{H\in\mathcal{H}(U)} \big\{ -\min H \,\big|\, \mathcal{P}^o_{\leq 1}(H) = \mathrm{Crit}(H)\big\} \in (0,\infty],$$

where $\mathcal{P}^o_{\leq 1}(H)$ is the set of elements in $\mathcal{P}_{\leq 1}(H)$ which are contractible in M. Clearly $c_{\mathrm{HZ}}(U) \leq c^o_{\mathrm{HZ}}(U)$ and finiteness of $c^o_{\mathrm{HZ}}(U)$ implies the almost existence theorem with the additional information that the periodic orbits we find are contractible in M. However, verifying finiteness of $c_{\mathrm{HZ}}(U)$ or $c^o_{\mathrm{HZ}}(U)$ is usually highly non-trivial and there are also many examples where these quantities are infinite [61].

The first relevant class of manifolds one encounters outside $\mathbb{R}^{2n}$ is given by cotangent bundles T^*Q over closed manifolds Q endowed with the standard symplectic form ω_{T^*Q}. Besides their importance as phase spaces in classical mechanics, they give local models for Lagrangian submanifolds of abstract symplectic manifolds by the Weinstein neighborhood theorem. This implies that if Q admits a Lagrangian embedding into $\mathbb{R}^{2n}$ (for instance Q is a torus), then any bounded subset $U \subset T^*Q$ can be symplectically embedded into a bounded subset of $\mathbb{R}^{2n}$ and hence the almost existence theorem holds for U, see for instance [37, Proposition 1.9]. For general manifolds Q, Hofer and Viterbo [33] could carry over the variational setting employed in $\mathbb{R}^{2n}$ by exploiting the fact that the fibers of T^*Q are linear spaces to prove the following remarkable result (see the work of Asselle and Starostka [8] for a simplified approach).

Theorem 1.2. (Hofer and Viterbo [33]) *The nearby existence theorem holds for hypersurfaces $\Sigma \subset T^*Q$ which bound a compact region containing the zero-section in its interior.*

To go beyond cotangent bundles, one leaves the classical variational approach and enters the theory of pseudoholomorphic curves initiated by Gromov in [27]. For closed symplectic manifolds, the non-autonomous setting is governed by the Arnold conjecture [4], giving a lower bound on the number of fixed points of Hamiltonian diffeomorphisms, and by the Conley conjecture, asserting that for a large class of symplectic manifolds every Hamiltonian diffeomorphism has infinitely many periodic points. These conjectures led to a tremendous development in the field and have been settled by now in very general forms, see, e.g., [17–19,32,39,40,48] for the Arnold conjecture and [23,25] for the Conley conjecture.

Pseudoholomorphic curves have also been employed in the autonomous setting starting from the pioneering work of Hofer and Viterbo [34], where they are used to give an upper bound on c^o_{HZ} of certain symplectic manifolds. Their approach has been further developed using Gromov–Witten invariants in [41–44,60].

For both the non-autonomous and the autonomous setting on general symplectic manifolds, a key role is played by the notion of displaceability and of spectral invariants. Recall that a domain U is displaceable in M if there exists a compactly supported Hamiltonian diffeomorphism ϕ^1_K of M such that $\phi^1_K(U)\cap U=\varnothing$. Quantitatively speaking, one can define the displacement energy of U by

$$e(U) := \inf\left\{\|K\| \mid \phi^1_K(U)\cap U=\varnothing\right\},$$

where $\|K\|$ is the Hofer norm of K, see [36]. Finiteness of $e(U)$ is equivalent to displaceability of U. If M is either closed or convex at infinity, one defines the spectral capacity of U by

$$c_\sigma(U) := \sup\{\sigma(H)\},$$

where the supremum is taken over all Hamiltonian functions compactly supported in $U\backslash\partial U$ and $\sigma(H)$ represents the spectral invariant of H associated with the fundamental class in Floer homology [21,47,53,63]. The chain of inequalities

$$c^o_{\mathrm{HZ}}(U)\le c_\sigma(U)\le e(U) \tag{1.1}$$

holds by [57,60] generalizing work contained in [15,21,36,53,54,63] (see also [29] for an extension of this argument to some open, non-convex manifolds). One significant implication of finiteness of $c_\sigma(U)$ is that every Hamiltonian diffeomorphism φ compactly supported in $U\backslash\partial U$ has finite spectral norm and thus carries infinitely many non-trivial contractible periodic points. Since a bounded domain in $\mathbb{R}^{2n}$ is displaceable, the latter inequality in (1.1) generalizes Theorem 1.1.

Thus, displaceability gives a remarkable criterion to show that $c_\sigma(U)$ and hence $c^o_{\mathrm{HZ}}(U)$ are finite. In many cases of interest, however, either it is difficult to prove that a subset is displaceable or it can be shown that many subsets are not displaceable even by topological reasons (see the notion of stable displacement for a fix of this second difficulty [54]). A precious tool to study periodic orbits in these cases is represented by symplectic homology, see, e.g., [13]. It was originally defined for $\mathbb{R}^{2n}$ by Floer and Hofer [16] and for more general manifolds by Cieliebak et al. [12], and further developed by Viterbo in [66]. In the next subsection, we discuss how symplectic homology can help us in finding periodic orbits on Liouville domains, an important class of symplectic manifolds.

1.2. Capacities and symplectic homology for Liouville domains

A Liouville domain (W, λ) is a compact manifold W with boundary ∂W such that the exterior derivative $\mathrm{d}\lambda$ of the one-form λ is symplectic and the Liouville vector field Y on W characterized by $\mathrm{d}\lambda(Y, \cdot) = \lambda$ points outwards along ∂W. The one-form $\lambda|_{\partial W}$ restricted to ∂W is a contact form and we denote by $\mathrm{spec}(R_\lambda)$ the set of periods of periodic orbits of the Reeb vector field R_λ of $(\partial W, \lambda|_{\partial W})$. It is a nowhere dense closed subset of $\mathbb{R}$.

We define the skeleton W_{sk} of the Liouville domain to be the set of α-limits of the flow ϕ_Y^t of Y on W, i.e.,

$$W_{\mathrm{sk}} := \bigcap_{t<0} \phi_Y^t(W).$$

The complement $W \backslash W_{\mathrm{sk}}$ is symplectomorphic to the negative half of the symplectization of $(\partial W, \lambda|_{\partial W})$ via the map

$$\Psi\colon \partial W \times (0,1] \to W, \quad (x,r) \mapsto \phi_Y^{\log r}(x),$$

where $\Psi^*\lambda = r\lambda|_{\partial W}$, and we can complete $(W, \mathrm{d}\lambda)$ by attaching a cylindrical end

$$\widehat{W} := W \cup \big(\partial W \times [1,\infty)\big) \tag{1.2}$$

using Ψ and by setting $\hat{\lambda} = r\lambda|_{\partial W}$ on the cylindrical end.

The simplest examples of Liouville domains are starshaped domains in $\mathbb{R}^{2n}$. In this case, Y is the radial vector field and the skeleton is a single point. Other examples are fiberwise starshaped domains in the cotangent bundle T^*Q over a closed manifold Q with the canonical one-form. In this case, Y is the fiberwise radial vector field and the skeleton is the zero-section. Moreover, $\mathbb{R}^{2n}$ and T^*Q are exactly the completions of the two examples we just described.

The symplectic homology $\mathrm{SH}(W;\alpha)$ of (W, λ) in the free-homotopy class $\alpha \in [S^1, W]$ is, roughly speaking, generated by periodic orbits of the Reeb vector field R_λ and, when $\alpha = \mathbb{1}$ is the class of contractible loops, also by a Morse complex for $(W, \partial W)$. We refer to Sect. 3 for the precise definition of symplectic homology. One can also construct the positive symplectic homology $\mathrm{SH}^+(W;\alpha)$ which is generated just by the periodic Reeb orbits and not by the Morse complex, so that $\mathrm{SH}^+(W;\alpha) = \mathrm{SH}(W;\alpha)$ for $\alpha \neq \mathbb{1}$.

Symplectic homology carries a natural filtration given by periods of Reeb orbits. Using this filtration, several kinds of symplectic capacities can be constructed. One of them is $c_{\mathrm{SH}}(W) \in (0,\infty]$ which reads off the minimal filtration level such that the Morse complex of $(W, \partial W)$ is annihilated in $\mathrm{SH}(W;\mathbb{1})$, see (3.13), [66, Section 5.3], [31], and [28, Proposition 3.5]. It is finite if and only if $\mathrm{SH}(W;\mathbb{1})$ vanishes. We have the inequalities

$$c_{\mathrm{HZ}}^o(W) \le c_{\mathrm{SH}}(W) \le e(W), \tag{1.3}$$

where the displacement energy is taken with respect to the completion of W. The former inequality is due to Irie [37]. The latter is due to Borman–McLean [9, Theorem 1.5(ii)], Kang [38, Corollary A1] and Ginzburg–Shon [28, Corollary 3.9], see also [28, Corollary 3.10] for an inequality involving the stable displacement energy, and [56, Theorem 2.2] for a similar inequality in wrapped Floer homology. We complement (1.3) and (1.1) by the following theorem which relies on Irie's idea in [37].

Theorem 1.3. *For a Liouville domain* (W, λ), *there holds* $c_\sigma(W) \le c_{\mathrm{SH}}(W)$, *and thus*

$$c^o_{\mathrm{HZ}}(W) \le c_\sigma(W) \le c_{\mathrm{SH}}(W) \le e(W).$$

In particular, if $\mathrm{SH}(W; \mathbb{1}) = 0$, *then every Hamiltonian diffeomorphism supported in* $W \backslash \partial W$ *admits a non-trivial fixed point and infinitely many distinct non-trivial periodic points.*

We see that vanishing of $\mathrm{SH}(W; \mathbb{1})$, which is the case, e.g., if W is displaceable in its completion, immediately implies the almost existence theorem for contractible orbits and the statement of Theorem 1.1 for W. This line of attack is however not suited to study finiteness of the Hofer–Zehnder capacity for bounded fiberwise starshaped domains W in cotangent bundles T^*Q, since $\mathrm{SH}(W; \mathbb{1})$ does not vanish in this case. Indeed by [1,5–7,59,64], there exists Viterbo's isomorphism

$$\mathrm{H}(\mathcal{L}_\alpha Q) \cong \mathrm{SH}(W; \alpha) \tag{1.4}$$

over $\mathbb{Z}_2$-coefficients (which is sufficient for our purpose), where $\mathcal{L}_\alpha Q$ denotes the free-loop space of Q in the class $\alpha \in [S^1, Q] \cong [S^1, W]$. However, symplectic homology can still be effectively used to prove the almost existence theorem on cotangent bundles, and we describe two instances how this has been done.

(a) For a large class of closed manifolds Q including those for which the Hurewicz map $\pi_2(Q) \to \mathrm{H}_2(Q; \mathbb{Z})$ is non-zero, Albers, Frauenfelder, and Oancea [2] generalized an idea by Ritter in the simply connected case [50, Corollary 8] and showed that the symplectic homology of $W \subset T^*Q$ twisted by some local coefficients vanishes. The capacity $c_{\mathrm{SH}}(W)$ using this twisted version still gives an upper bound for $c^o_{\mathrm{HZ}}(W)$.

    On the other hand, Frauenfelder and Pajitnov [20] considered the S^1-equivariant version $\mathrm{H}^{S^1}(\mathcal{L}_\mathbb{1} Q) \cong \mathrm{SH}^{S^1}(W; \mathbb{1})$ of isomorphism (1.4) with $\alpha = \mathbb{1}$ and rational coefficients. Generalizing an approach due to Viterbo [65], they observe that when Q belongs to the class of rationally inessential manifolds, for instance when Q is simply connected, then $\mathrm{H}^{S^1}(\mathcal{L}_\mathbb{1} Q)$ with rational coefficients vanishes by Goodwillie's theorem and gives an upper bound on $c^o_{\mathrm{HZ}}(W)$ by a finite capacity coming from $\mathrm{SH}^{S^1}(W; \mathbb{1})$ similarly as above.

(b) The total homology $\bigoplus_\alpha \mathrm{SH}(W; \alpha)$ admits a ring structure with unit e. Irie [37] proved that $c_{\mathrm{HZ}}(W)$ is finite if there exists a free-homotopy class α different from $\mathbb{1}$ such that $e = x * y$ for some $x \in \mathrm{SH}(W; \alpha)$ and $y \in \mathrm{SH}(W; \alpha^{-1})$. Moreover, using the fact that $*$ corresponds to the Chas–Sullivan product in $\bigoplus_\alpha \mathrm{H}_*(\mathcal{L}_\alpha Q)$ through the Viterbo isomorphism, he showed that such condition is fulfilled when the evaluation map $\mathcal{L}_\alpha Q \to Q$, $x \mapsto x(0)$ has a section.

Like the inequality $c^o_{\mathrm{HZ}}(W) \leq c_{\mathrm{SH}}(W)$, also the inequality $c_\sigma(W) \leq c_{\mathrm{SH}}(W)$ established in Theorem 1.3 holds when one twists the coefficients as in [2]. We obtain, therefore, the following result in the spirit of Theorem 1.1.

Corollary 1.4. *Let Q be a closed manifold such that the Hurewicz map $\pi_2(Q) \to \mathrm{H}_2(Q;\mathbb{Z})$ is non-zero. Then, every compactly supported Hamiltonian diffeomorphism on T^*Q has a non-trivial one-periodic point and infinitely many distinct non-trivial periodic points.* □

The approach coming from inequality (1.3) and from (a) above are based on the vanishing of symplectic homology. One can ask if knowledge on the positive symplectic homology $\mathrm{SH}^+(W;\alpha)$ can be translated into a bound for some kind of Hofer–Zehnder capacity. For $\alpha \neq \mathbb{1}$, this idea has been explored by Biran et al. in [10]. Inspired by the work of Gatien and Lalonde [26], they introduced a relative capacity as follows. Let U be a symplectic domain possibly with boundary, and let $Z \subset U \backslash \partial U$ be a compact subset. Let $\mathcal{H}_{\mathrm{BPS}}(U,Z)$ be the set of smooth Hamiltonians $H\colon S^1 \times U \to \mathbb{R}$ such that H vanishes outside a compact subset of $U\backslash\partial U$ and $\max_{S^1\times Z} H$ is negative. Then, the relative capacity with class $\alpha \in [S^1,U]$, $\alpha \neq \mathbb{1}$ introduced in [10] is given by

$$c_{\mathrm{BPS}}(U,Z;\alpha) := \sup_{H\in\mathcal{H}_{\mathrm{BPS}}(U,Z)} \Big\{ -\max_{S^1\times Z} H \;\Big|\; \mathcal{P}_1(H;\alpha) = \varnothing \Big\} \in (0,\infty],$$

where $\mathcal{P}_1(H;\alpha)$ is the set of elements of $\mathcal{P}_1(H)$ in the class α. By definition, if $c_{\mathrm{BPS}}(U,Z;\alpha)$ is finite, we can infer the existence of a one-periodic orbit in the class α for every $H \in \mathcal{H}_{\mathrm{BPS}}(U,Z)$ with $-\max_{S^1\times Z} H > c_{\mathrm{BPS}}(U,Z;\alpha)$. Moreover, much as in the case of the Hofer–Zehnder capacity, finiteness of c_{BPS} implies the almost existence theorem for hypersurfaces $\Sigma \subset U\backslash\partial U$ bounding a compact region containing Z in its interior.

For the unit-disc cotangent bundle D^*Q of Q with Finsler metric F, there holds

$$\ell_\alpha = c_{\mathrm{BPS}}(D^*Q, Q;\alpha), \quad \alpha \neq \mathbb{1} \tag{1.5}$$

where ℓ_α is the minimal F-length of a closed curve in Q in the class $\alpha \in [S^1,Q] \cong [S^1, D^*Q]$. For Riemannian metrics, this is established by Biran et al. in [10] for $Q = \mathbb{T}^n$ and by Weber [68] for all closed manifolds Q. For general Finsler metrics, this is proved by Gong and Xue in [30]. As a result, the almost existence theorem for non-contractible orbits holds for compact hypersurfaces which bound a compact region containing the zero-section in its interior. This strengthens the nearby existence Theorem 1.2 of Hofer and Viterbo for non-simply connected manifolds Q.

For Liouville domains W and any class $\alpha \in [S^1,W]$, we can now use $\mathrm{SH}^+(W;\alpha)$ to define $c_{\mathrm{SH}^+}(W;\alpha) \in (0,\infty]$ as the minimal filtration level at which a non-zero class of $\mathrm{SH}^+(W;\alpha)$ appears, see (3.14). Thus, the finiteness of $c_{\mathrm{SH}^+}(W;\alpha)$ is equivalent to the non-vanishing of $\mathrm{SH}^+(W;\alpha)$. For contractible loops, we use the notation $c_{\mathrm{SH}^+}(W) := c_{\mathrm{SH}^+}(W;\mathbb{1})$ and we have $c_{\mathrm{SH}^+}(W) \leq c_{\mathrm{SH}}(W)$ as observed in Lemma 3.3. For non-contractible loops, Weber proved in [68] (although the result is not explicitly stated) that

$$c_{\mathrm{BPS}}(W, W_{\mathrm{sk}}; \alpha) \le c_{\mathrm{SH}^+}(W; \alpha), \quad \alpha \ne \mathbb{1}. \tag{1.6}$$

For cotangent bundles the filtered version of Viterbo's isomorphism (1.4) yields

$$\ell_\alpha = c_{\mathrm{SH}^+}(D^*Q; \alpha), \quad \alpha \ne \mathbb{1},$$

so that ℓ_α, $c_{\mathrm{BPS}}(D^*Q, Q; \alpha)$ and $c_{\mathrm{SH}^+}(D^*Q; \alpha)$ all coincide.

2. Statement of main results

The present paper originates as an attempt to study a counterpart to the inequality (1.6) for $\alpha = \mathbb{1}$. To this purpose, we need to replace the Biran–Polterovich–Salamon capacity with another relative Hofer–Zehnder capacity which we now define following the work of Ginzburg and Gürel [22]. For a symplectic manifold U possibly with boundary and a compact subset Z of $U \backslash \partial U$, we define the set $\mathcal{H}(U, Z)$ of smooth Hamiltonians $H\colon U \to \mathbb{R}$ such that H vanishes outside a compact subset of $U \backslash \partial U$ and $H = \min H < 0$ on a neighborhood of Z. Then, the relative Hofer–Zehnder capacity is given by

$$c_{\mathrm{HZ}}(U, Z) := \sup_{H \in \mathcal{H}(U,Z)} \big\{ -\min H \bigm| \mathcal{P}_{\le 1}(H) = \mathrm{Crit}(H) \big\} \in (0, \infty].$$

Considering only orbits in the class $\alpha \in [S^1, U]$, we can also define $c_{\mathrm{HZ}}(U, Z; \alpha)$. By definition,

$$c_{\mathrm{HZ}}(U, Z; \alpha) \le c_{\mathrm{BPS}}(U, Z; \alpha), \quad \alpha \ne \mathbb{1}. \tag{2.1}$$

Moreover, for Liouville domains W, one can easily see

$$\min \mathrm{spec}(R_\lambda; \alpha) \le c_{\mathrm{HZ}}(W, W_{\mathrm{sk}}; \alpha) \quad \forall \alpha \in [S^1, W] \tag{2.2}$$

where $\mathrm{spec}(R_\lambda; \alpha)$ is the set of periods of Reeb orbits of $(\partial W, \lambda|_{\partial W})$ with the class α in W.

From now on, we focus on contractible orbits and write $c^o_{\mathrm{HZ}}(U, Z) := c_{\mathrm{HZ}}(U, Z; \mathbb{1})$. In order to deal with time-dependent Hamiltonians as well, we introduce a slightly different version of $c^o_{\mathrm{HZ}}(U, Z)$ when the symplectic form $\omega = \mathrm{d}\lambda$ is exact on U. We consider the set $\widetilde{\mathcal{H}}(U, Z)$ consisting of smooth time-dependent Hamiltonians $H\colon S^1 \times U \to \mathbb{R}$ such that H vanishes outside a compact subset of $S^1 \times (U \backslash \partial U)$ and $H = \min H < 0$ on a neighborhood of $S^1 \times Z$. The action of p-periodic loops $\gamma\colon \mathbb{R}/p\mathbb{Z} \to U$ with respect to $H\colon S^1 \times U \to \mathbb{R}$ is given by

$$\mathcal{A}_H(\gamma) = \int_0^p \gamma^*\lambda - \int_0^p H(t, \gamma(t))\, \mathrm{d}t. \tag{2.3}$$

For all $a \in (0, \infty]$, we define

$$\widetilde{c}^{\,o}_{\mathrm{HZ}}(U, Z, a) := \sup_{H \in \widetilde{\mathcal{H}}(U,Z)} \big\{ -\min H \bigm| \forall x \in \mathcal{P}^o_1(H), \\ \mathcal{A}_H(x) \notin (-\min H, -\min H + a] \big\} \in (0, \infty].$$

By definition, there holds

$$c^{o}_{\mathrm{HZ}}(U,Z) \le \widetilde{c}^{o}_{\mathrm{HZ}}(U,Z,a) \quad \forall a \in (0,\infty]. \tag{2.4}$$

Moreover, every $H \in \widetilde{\mathcal{H}}(U,Z)$ with $-\min H > \widetilde{c}^{o}_{\mathrm{HZ}}(U,Z,a)$ has a (non-constant) contractible one-periodic orbit x with

$$\mathcal{A}_H(x) \in (-\min H, -\min H + a]. \tag{2.5}$$

Building on these two facts, we obtain the following implications of this newly defined capacity to periodic orbits in the autonomous and non-autonomous setting. To this end, recall that a free-homotopy class $\alpha \in [S^1, U]$ is called torsion if $\alpha^p = \mathbb{1}$ for some $p \in \mathbb{N}$.

Proposition 2.1. *Let $(U, \mathrm{d}\lambda)$ be an exact symplectic manifold possibly with boundary and let Z be a compact subset of $U \backslash \partial U$.*

- (a) *If $\widetilde{c}^{o}_{\mathrm{HZ}}(U,Z,a) < \infty$ for some $a \in (0,\infty]$, then the almost existence theorem for contractible orbits holds for every hypersurface in $U\backslash\partial U$ bounding a compact region containing Z in its interior.*
- (b) *Assume that $\widetilde{c}^{o}_{\mathrm{HZ}}(U,Z,a) < \infty$ for some $a < \infty$. If $H \in \widetilde{\mathcal{H}}(U,Z)$ has only finitely many one-periodic orbits with torsion free-homotopy classes and with action greater than $-\min H$, then for every sufficiently large prime number p, there exists a contractible, non-iterated, non-constant, p-periodic orbit of H which has action greater than $-p\min H$. In particular every $H \in \widetilde{\mathcal{H}}(U,Z)$ carries infinitely many distinct contractible, non-constant periodic orbits.*

Proof. Suppose that $\widetilde{c}^{o}_{\mathrm{HZ}}(U,Z,a)$ is finite for some $a \in (0,\infty]$. Then, by (2.4), $c^{o}_{\mathrm{HZ}}(U,Z)$ is also finite, and the almost existence theorem for contractible orbits follows from [22, Theorem 2.14]. This proves (a).

To show (b), we assume that a is finite. Let $H \in \widetilde{\mathcal{H}}(U,Z)$ be as in the statement, and let $\gamma_1,\dots,\gamma_m$ be all one-periodic orbits of H with torsion free-homotopy classes and with action greater than $-\min H$. We choose $\epsilon > 0$ such that

$$\mathcal{A}_H(\gamma_i) \ge -\min H + \epsilon \quad \forall i \in \{1,\dots,m\}.$$

For $p \in \mathbb{N}$, we define the p-th iteration $H^{\natural p}\colon S^1 \times W \to \mathbb{R}$ of H by $H^{\natural p}(t,z) := pH(pt,z)$ so that the Hamiltonian flow of $H^{\natural p}$ and that of H are related by $\phi^t_{H^{\natural p}} = \phi^{pt}_H$. Thus, one-periodic orbits of $H^{\natural p}$ can be viewed as p-periodic orbits of H. If $\gamma\colon S^1 \to W$ is a one-periodic orbit of H, then its p-th iteration $\gamma^p\colon S^1 \to W$ defined by $\gamma^p(t) := \gamma(pt)$ is a one-periodic orbit of $H^{\natural p}$. Moreover, there obviously holds $\mathcal{A}_{H^{\natural p}}(\gamma^p) = p\mathcal{A}_H(\gamma)$. Let now p be a prime number so large that

$$\epsilon p > a, \quad -\min H^{\natural p} = -p\min H > \widetilde{c}^{o}_{\mathrm{HZ}}(U,Z,a).$$

By the definition of $\widetilde{c}^{o}_{\mathrm{HZ}}(U,Z,a)$, $H^{\natural p}$ has a contractible one-periodic orbit γ_{new} such that

$$-\min H^{\natural p} < \mathcal{A}_{H^{\natural p}}(\gamma_{\mathrm{new}}) \le -\min H^{\natural p} + a.$$

The first inequality shows that γ_{new} is non-constant. The latter one yields that γ_{new} is non-iterated. Indeed if γ_{new} is iterated, then it has to be the p-th iteration of γ_i for some $1 \le i \le m$ and this is absurd since

$$\mathcal{A}_{H^{\natural p}}(\gamma_i^p) = p\mathcal{A}_H(\gamma_i) \ge p(-\min H + \epsilon) > -\min H^{\natural p} + a, \quad \forall i = 1, \dots, m.$$

This finishes the proof of (b). □

We are now ready to state our main result. It says that the π_1-sensitive Hofer–Zehnder capacity of a Liouville domain relative to its skeleton can be bounded by the capacity obtained from positive symplectic homology in the contractible class.

Theorem 2.2. *For every Liouville domain W, there holds*

$$\tilde{c}^o_{\mathrm{HZ}}(W, W_{\mathrm{sk}}, a) \le c_{\mathrm{SH}^+}(W) \quad \forall a \in \big[c_{\mathrm{SH}^+}(W), \infty\big].$$

Hence if $c_{\mathrm{SH}^+}(W)$ is finite, or equivalently $\mathrm{SH}^+(W; \mathbb{1})$ is non-zero, then the same conclusion as in Proposition 2.1 holds for $(U, Z) = (W, W_{\mathrm{sk}})$.

Remark 2.3. The hypothesis $\mathrm{SH}^+(W; \mathbb{1}) \ne 0$ in Theorem 2.2 is indispensable. For example, if W is a fiberwise starshaped domain in T^*S^1, then $\mathrm{SH}^+(W; \mathbb{1}) = 0$ and none of (a) and (b) in Proposition 2.1 is true.

Remark 2.4. For $\delta > 0$, let

$$W^\delta := \phi_Y^{\log \delta}(W). \tag{2.6}$$

Then, the proof of Theorem 2.2 actually shows that for any $\delta \in (0, 1]$,

$$\tilde{c}^o_{\mathrm{HZ}}(W, W^\delta, a) \le (1 - \delta)c_{\mathrm{SH}^+}(W) \quad \forall a \in \big[(1 - \delta)c_{\mathrm{SH}^+}(W), \infty\big]$$

and this subsumes Theorem 2.2 since $W_{\mathrm{sk}} = \bigcap_{\delta>0} W^\delta$.

Theorem 2.2 will follow from Proposition 4.1, which provides a lower bound on the number of contractible one-periodic orbits of $H \in \widetilde{\mathcal{H}}(W, W_{\mathrm{sk}})$ with action in a certain interval in terms of positive symplectic homology. Combining Theorem 2.2 with the isomorphism (1.4), we immediately obtain the following corollary in T^*Q.

Corollary 2.5. *Let Q be a closed manifold such that $\mathrm{H}(\mathcal{L}_{\mathbb{1}}Q, Q)$ is non-zero.*

(a) *The almost existence theorem for contractible orbits holds for every hypersurface of T^*Q bounding a compact region containing the zero-section.*
(b) *Every compactly supported smooth Hamiltonian $H\colon S^1 \times T^*Q \to \mathbb{R}$ with $H = \min H < 0$ on $S^1 \times U$, where U is a neighborhood of the zero-section, has infinitely many distinct non-constant, contractible periodic orbits with action greater than $-\min H$.* □

Remark 2.6. Let Q be simply connected. By the theory of minimal models of Sullivan [58,67], the group $\mathrm{H}(\mathcal{L}_{\mathbb{1}}Q, Q)$ is infinite dimensional and, in particular, non-zero. As observed by Thomas Rot in a MathOverflow post [51], using the (relative) Hurewicz theorem and the long exact sequence of the pair $(\mathcal{L}_{\mathbb{1}}Q, Q)$ in homology and homotopy, one can show that $\mathrm{H}_{k-1}(\mathcal{L}_{\mathbb{1}}Q, Q) \ne 0$, if k is the smallest integer such that $\pi_k(Q) \neq 0$. When $\pi_1(Q) \cong \mathbb{Z}$ conditions on the homotopy groups of Q ensuring $\mathrm{H}(\mathcal{L}_{\mathbb{1}}Q, Q) \neq 0$ are given in [3, Corollary 1.9].

Remark 2.7. The corollary finds application also to exact twisted cotangent bundles. Let $\mathrm{d}\theta$ be an exact two-form on a closed manifold Q and consider the twisted cotangent bundle $(T^*Q, \omega_{T^*Q} + \pi^*(\mathrm{d}\theta))$, where ω_{T^*Q} is the canonical symplectic form on T^*Q and π is the foot-point projection $\pi\colon T^*Q \to Q$. If $\mathrm{H}(\mathcal{L}_{\mathbb{1}}Q, Q) \neq 0$, then the statements in Corollary 2.5 hold by replacing the zero-section with the graph of the one-form θ.

If D^*Q is the unit-disc cotangent bundle of a Riemannian metric g, the exact value of $c_{\mathrm{SH}^+}(D^*Q)$ and of $\tilde{c}^o_{\mathrm{HZ}}(D^*Q, Q, a)$ for $a \geq c_{\mathrm{SH}^+}(D^*Q)$ can be computed via Viterbo isomorphism if we know the homology of $\mathcal{L}_{\mathbb{1}}Q$ filtered by the square root of the g-energy of loops sufficiently well. For instance, applying a theorem of Ziller [71] and Lemma A.2 contained in the appendix of the present paper written by Abbondandolo and Mazzucchelli, we get the following statement.

Corollary 2.8. *Let Q be a closed manifold endowed with a Riemannian metric g. Let D^*Q be the unit-disc cotangent bundle of g and denote by $\ell_{\mathbb{1}}$ the length of the shortest non-constant contractible geodesic for the metric. If (Q, g) is a compact, non-aspherical homogeneous space (for instance a compact rank one symmetric space) or (Q, g) is a two-sphere with positive curvature, then there holds*

$$\ell_{\mathbb{1}} = \tilde{c}^o_{\mathrm{HZ}}(D^*Q, Q, a) = c_{\mathrm{SH}^+}(D^*Q), \quad \forall a \in \big[c_{\mathrm{SH}^+}(D^*Q), \infty\big].$$

Finally, we can adapt [22, Theorem 3.2] to obtain weaker statements on the existence of periodic orbits for H belonging to a class larger than $\widetilde{\mathcal{H}}(W, W_{\mathrm{sk}})$ of functions that are allowed to be time-dependent also on W_{sk}. The definition of the Floer homology HF and the canonical map $\iota^{\epsilon,\infty}_{-\infty}$ is given in Sect. 3.

Theorem 2.9. *Let (W, λ) be a Liouville domain, and let $H\colon S^1 \times W \to \mathbb{R}$ be a Hamiltonian which is supported in $S^1 \times (W \backslash \partial W)$ and satisfies* $\max_{S^1 \times W_{\mathrm{sk}}} H < 0$.

(a) *For every small $a > 0$, there holds*

$$\mathrm{rk}\, \mathrm{HF}^{(a,\infty)}(H) \geq \mathrm{rk}\, \left[\iota^{\epsilon,\infty}_{-\infty}\colon \mathrm{H}(W, \partial W) \to \mathrm{SH}(W; \mathbb{1})\right].$$

In particular, if $\mathrm{SH}(W; \mathbb{1})$ is non-zero, then H has a contractible one-periodic orbit with positive action.

(b) *Assume in addition that $\max_{S^1 \times W} H = 0$. For every small $a > 0$, there exists a surjective homomorphism*

$$\mathrm{HF}^{(a,\infty)}(H) \longrightarrow \mathrm{H}(W, \partial W).$$

In particular, H has a contractible one-periodic orbit with positive action.

Organization of the paper

In Sect. 3, we recall the precise definition of Hamiltonian Floer homology, of symplectic homology, of the associated capacities. At the end of the section, a proof of Theorem 1.3 is given. In Sect. 4, we prove Theorem 2.2, Theorem 2.9, and Corollary 2.8. The appendix, written by Abbondandolo and Mazzucchelli, shows the monotonicity of the systoles for convex Riemannian two-spheres in $\mathbb{R}^3$. In doing so, they prove Lemma A.2 which is needed in Corollary 2.8.

3. Floer homologies and symplectic capacities

In this section, we define the capacities given by symplectic homology and prove Theorem 1.3. Prior to this, we give a concise construction of Floer homology and refer to [10,12,22,52,66,68] for details.

3.1. Hamiltonian Floer homology

For a Liouville domain (W,λ), let $(\widehat{W},\hat{\lambda})$ be its completion defined in (1.2) and $W^\delta \subset \widehat{W}$ for $\delta > 0$ be the subset given in (2.6). We consider a smooth Hamiltonian $H\colon S^1 \times \widehat{W} \to \mathbb{R}$ with

$$H(t,r,x) = \tau r + \eta \quad (t,r,x) \in S^1 \times (\widehat{W}\backslash W^\delta) = S^1 \times (\delta,\infty) \times \partial W \tag{3.1}$$

for some $\delta > 0$, $\tau \in (0,\infty)\backslash\mathrm{spec}(R_\lambda)$ and $\eta \in \mathbb{R}$. The constant τ is called the slope of H. The action spectrum $\mathrm{spec}(H)$ is the set of action values of all critical points of $\mathcal{A}_H$. This is a compact, nowhere dense subset of $\mathbb{R}$. Let $a \le b$ be numbers in $\overline{\mathbb{R}} := \mathbb{R}\cup\{-\infty,\infty\}$ not belonging to $\mathrm{spec}(H)$. We denote by $\mathcal{P}_1^{(a,b)}(H;\alpha)$ the set of one-periodic orbits of H with free-homotopy class α and with action in (a,b). Suppose that all elements in $\mathcal{P}_1^{(a,b)}(H;\alpha)$ are nondegenerate. Then, this is a finite set due to $\tau \notin \mathrm{spec}(R_\lambda)$. The Floer chain group is

$$\mathrm{CF}^{(a,b)}(H;\alpha) := \bigoplus_{x\in\mathcal{P}_1^{(a,b)}(H;\alpha)} \mathbb{Z}_2\langle x\rangle.$$

Let J be a smooth S^1-family of almost complex structures on $\widehat{W}$ with the property that $\mathrm{d}\hat{\lambda}(\cdot,J(t,u)\cdot)$ is an inner product on $T_u\widehat{W}$ for all $(t,u) \in S^1\times\widehat{W}$ and satisfying

$$J^*\hat{\lambda} = \mathrm{d}r, \quad \text{on } \{r \ge \delta_0\} \tag{3.2}$$

for some $\delta_0 > 0$ such that all element in $\mathcal{P}_1^{(a,b)}(H;\alpha)$ are contained in the interior of W^{δ_0}. For $x,y \in \mathcal{P}_1(H;\alpha)$, we denote by $\mathcal{M}(x,y)$ the moduli space of Floer cylinders connecting x and y, namely smooth solutions $u\colon \mathbb{R}\times S^1 \to \widehat{W}$ of

$$\partial_s u - J(t,u)\big(\partial_t u - X_H(t,u)\big) = 0 \quad \lim_{s\to-\infty} u(s,\cdot) = x, \quad \lim_{s\to+\infty} u(s,\cdot) = y, \tag{3.3}$$

where $(s,t) \in \mathbb{R} \times S^1$ and X_H is the Hamiltonian vector field of H defined via the equation $\mathrm{d}\hat{\lambda}(X_H, \cdot) = -\mathrm{d}H$. Unless $x = y$, there is a free $\mathbb{R}$-action on $\mathcal{M}(x,y)$ by translating solutions in the s-direction. We define $n(x,y)$ as the parity of $\mathcal{M}(x,y)/\mathbb{R}$ if $x \neq y$ and it is a finite set. Otherwise we set $n(x,y) = 0$. The differential $\partial\colon \mathrm{CF}^{(a,b)}(H;\alpha) \to \mathrm{CF}^{(a,b)}(H;\alpha)$ is defined by the linear extension of the formula

$$\partial x := \sum_{y \in \mathcal{P}_1^{(a,b)}(H;\alpha)} n(x,y)y.$$

For a generic choice of J, we indeed have $\partial \circ \partial = 0$ and denote the Floer homology of H with action-window $(a,b) \subset \mathbb{R}$ and with free-homotopy class $\alpha \in [S^1, \widehat{W}]$ by

$$\mathrm{HF}^{(a,b)}(H;\alpha) := \mathrm{H}(\mathrm{CF}^{(a,b)}(H;\alpha), \partial).$$

Simplifying the notation we denote $\mathrm{HF}(H;\alpha) = \mathrm{HF}^{(-\infty,\infty)}(H;\alpha)$. As the notation indicates, a different choice of J produces an isomorphic homology via a continuation homomorphism.

Given $a < b < c$ not belonging to $\mathrm{spec}(H)$, the exact sequence of chain complexes

$$0 \longrightarrow \mathrm{CF}^{(a,b)}(H;\alpha) \longrightarrow \mathrm{CF}^{(a,c)}(H;\alpha) \longrightarrow \mathrm{CF}^{(b,c)}(H;\alpha) \longrightarrow 0$$

induced by natural inclusion and projection gives rise to the long exact sequence

$$\cdots \xrightarrow{\delta} \mathrm{HF}^{(a,b)}(H;\alpha) \xrightarrow{\iota} \mathrm{HF}^{(a,c)}(H;\alpha) \xrightarrow{\pi} \mathrm{HF}^{(b,c)}(H;\alpha) \xrightarrow{\delta} \mathrm{HF}^{(a,b)}(H;\alpha) \xrightarrow{\iota} \cdots. \tag{3.4}$$

Let $H, K\colon S^1 \times \widehat{W} \to \mathbb{R}$ be two smooth Hamiltonians with the aforementioned properties and $H \leq K$. We choose a smooth monotone homotopy H_s, $s \in \mathbb{R}$ from H to K, namely

$$H_s = H \quad \forall s \leq -1, \quad H_s = K \quad \forall s \geq 1, \quad \partial_s H_s \geq 0 \quad \forall s \in \mathbb{R},$$

and H_s has a constant slope, see (3.1), for every $s \in \mathbb{R}$. Consider the moduli space of solutions of (3.3) with H replaced by H_s and define the continuation homomorphism

$$\Phi = \Phi_{H,K}^{(a,b)}\colon \mathrm{HF}^{(a,b)}(H;\alpha) \longrightarrow \mathrm{HF}^{(a,b)}(K;\alpha) \tag{3.5}$$

in an analogous way to defining the differential. Another choice of monotone homotopy produces the same continuation homomorphism. Moreover, the map Φ induces a commuting map from the exact sequence (3.4) for H to that for K. If we consider another smooth Hamiltonian $G\colon S^1 \times \widehat{W} \to \mathbb{R}$ satisfying $K \leq G$, then we have continuation homomorphisms $\Phi_{K,G}^{(a,b)}$ and $\Phi_{H,G}^{(a,b)}$ and there holds $\Phi_{H,G}^{(a,b)} = \Phi_{K,G}^{(a,b)} \circ \Phi_{H,K}^{(a,b)}$. In the case that H and K have the same slope and $(a,b) = (-\infty,\infty)$, continuation homomorphisms $\Phi_{H,K} = \Phi_{H,K}^{(-\infty,\infty)}$ and $\Phi_{K,H}$ are still defined and satisfy $\Phi_{H,K} \circ \Phi_{K,H} = \Phi_{K,K} = \mathrm{Id}$ and $\Phi_{K,H} \circ \Phi_{H,K} = \Phi_{H,H} = \mathrm{Id}$. Thus, $\mathrm{HF}(H;\alpha) \cong \mathrm{HF}(K;\alpha)$ for H and K with the same slope.

In fact the above construction extends to smooth Hamiltonians $H: S^1 \times \widehat{W} \to \mathbb{R}$ such that elements $\mathcal{P}_1^{(a,b)}(H;\alpha)$ are not necessarily nondegenerate. The nondegeneracy condition can be achieved by a small compact perturbation K of H. Moreover, the Floer homology $\mathrm{HF}^{(a,b)}(K;\alpha)$ is independent of the choice of a small perturbation up to continuation isomorphisms. To be precise, if G is another small compact perturbation of H, then we have continuation homomorphisms $\Phi_{K,G}^{(a,b)}$ and $\Phi_{G,K}^{(a,b)}$ which are inverse to each other. Here it is crucial that $a,b \notin \mathrm{spec}(H)$, and K and G have the same slope. Thus, we set $\mathrm{HF}^{(a,b)}(H;\alpha) := \mathrm{HF}^{(a,b)}(K;\alpha)$.

Next we define the Floer homology of $H: S^1 \times W \to \mathbb{R}$ with support in $S^1 \times (W \backslash \partial W)$. We choose $\delta_1 \in (0,1)$ such that $H = 0$ on $W \backslash W^{\delta_1}$. Then, we smoothly extend H to $\widehat{H}: S^1 \times \widehat{W} \to \mathbb{R}$ to satisfy

- $\widehat{H} = H$ on $W \backslash W^{\delta_2}$ for some $\delta_2 \in (\delta_1, 1)$;
- $\widehat{H}(t,r,x) = h(r)$ on $(t,r,x) \in \widehat{W} \backslash W^{\delta_1} = S^1 \times (\delta_1, \infty) \times \partial W$ where $h: (\delta_1, \infty) \to \mathbb{R}$ is a smooth convex function;
- $h'(r) = \epsilon$ for some $0 < \epsilon < \min \mathrm{spec}(R_\lambda)$ on $\widehat{W} \backslash W$.

Then, we define the Floer homology of H as that of $\widehat{H}$:

$$\mathrm{HF}^{(a,b)}(H;\alpha) := \mathrm{HF}^{(a,b)}(\widehat{H};\alpha) \tag{3.6}$$

where $a,b \in \overline{\mathbb{R}} \backslash \mathrm{spec}(H)$ as usual. Due to the choice of slope ϵ, $\mathcal{P}_1(H) = \mathcal{P}_1(\widehat{H})$ and furthermore the definition (3.6) is independent of the choice of $\widehat{H}$.

Finally, we make an action computation that will be repeatedly used. Let $H: S^1 \times W \to \mathbb{R}$ be such that there exist $\delta > 0$ and a smooth function $h: (\delta, \infty) \to \mathbb{R}$ with the property that $H = h$ on $S^1 \times (W \backslash W^\delta) = S^1 \times (\delta, \infty) \times \partial W$. In this case, the action of one-periodic orbits x of H located on ∂W^r for $r \in (\delta, \infty)$ is explicitly computed as

$$\mathcal{A}_H(x) = rh'(r) - h(r), \tag{3.7}$$

which is minus the y-intercept of the tangent line of h at r.

3.2. Symplectic homology

Let $\mathcal{H}^{a,b}$ for $a \le b$ in $\overline{\mathbb{R}}$ be the set of smooth Hamiltonians $H: S^1 \times \widehat{W} \to \mathbb{R}$ satisfying $a,b \notin \mathrm{spec}(H)$, $H|_{S^1 \times W} < 0$, and (3.1) for some $\delta \ge 1$, $\tau \in (0,\infty) \backslash \mathrm{spec}(R_\lambda)$, and $\eta \in \mathbb{R}$. We endow $\mathcal{H}^{a,b}$ with the partial relation $\le$ given by the pointwise inequality so that for every $H, K \in \mathcal{H}^{a,b}$ with $H \le K$, we have the continuation homomorphism defined in (3.5). Floer homology groups of elements in $\mathcal{H}^{a,b}$ together with continuation homomorphisms form a direct system, and the direct limit is called the symplectic homology of W:

$$\mathrm{SH}^{(a,b)}(W;\alpha) := \varinjlim_{H \in \mathcal{H}^{a,b}} \mathrm{HF}^{(a,b)}(H;\alpha).$$

We remark that the symplectic homology changes only when the action-window crosses $\mathrm{spec}(R_\lambda)$, i.e., $\mathrm{SH}^{(a,b)}(W;\alpha) \cong \mathrm{SH}^{(a',b')}(W;\alpha)$ if $(a,b) \cap \mathrm{spec}(R_\lambda) = (a',b') \cap \mathrm{spec}(R_\lambda)$. Let ϵ denote a constant such that $0 < \epsilon <$

$\min \operatorname{spec}(R_\lambda)$. Thus, $(-\infty,\epsilon)\cap\operatorname{spec}(R_\lambda)=\varnothing$, and there holds

$$\mathrm{SH}^{(-\infty,\epsilon)}(W;\alpha)\cong\begin{cases}\mathrm{H}(W,\partial W) & \alpha=\mathbb{1}\,,\\ 0 & \alpha\neq\mathbb{1}\,,\end{cases} \tag{3.8}$$

where $\mathrm{H}(W,\partial W)$ is the relative homology of the pair $(W,\partial W)$. We denote

$$\mathrm{SH}(W;\alpha):=\mathrm{SH}^{(-\infty,+\infty)}(W;\alpha),\quad \mathrm{SH}^{+}(W;\alpha)=\mathrm{SH}^{(\epsilon,\infty)}(W;\alpha).$$

For any $a\leq b\leq c$ in $\overline{\mathbb{R}}\backslash\operatorname{spec}(R_\lambda)$, the exact sequence (3.4) leads to the exact sequence

$$\cdots\xrightarrow{\delta}\mathrm{SH}^{(a,b)}(W;\alpha)\xrightarrow{\iota}\mathrm{SH}^{(a,c)}(W;\alpha)\xrightarrow{\pi}\mathrm{SH}^{(b,c)}(W;\alpha)\xrightarrow{\delta}\mathrm{SH}^{(a,b)}(W;\alpha)\xrightarrow{\iota}\cdots. \tag{3.9}$$

We decorate ι to indicate involved action-windows as follows:

$$\iota_a^{b,c}:\mathrm{SH}^{(a,b)}(W;\alpha)\longrightarrow\mathrm{SH}^{(a,c)}(W;\alpha). \tag{3.10}$$

This map is functorial in the sense that $\iota_a^{c,d}\circ\iota_a^{b,c}=\iota_a^{b,d}$ holds for any $d\geq c$. Indeed, the map ι in (3.4) defined for each H has such a property and is compatible with the continuation homomorphism in (3.5). Thus, the desired functorial property for symplectic homology follows. Applying the exact sequence (3.9) to $(a,b,c)=(-\infty,\epsilon,\infty)$, we deduce

$$\mathrm{SH}^{+}(W;\alpha)\cong\mathrm{SH}(W;\alpha),\quad \alpha\neq\mathbb{1}$$

and

$$\begin{aligned}\dim\mathrm{SH}(W;\mathbb{1})=\infty\quad&\Longleftrightarrow\quad\dim\mathrm{SH}^{+}(W;\mathbb{1})=\infty,\\ \mathrm{SH}(W;\mathbb{1})=0\quad&\Longrightarrow\quad\mathrm{SH}^{+}(W;\mathbb{1})\cong\mathrm{H}(W;\partial W).\end{aligned}$$

Symplectic homologies with finite action-window and the homomorphisms in (3.10) can be interpreted as Floer homologies of suitably chosen Hamiltonians and continuation homomorphisms between them. For $a\in(0,\infty)\backslash\operatorname{spec}(R_\lambda)$, we consider the set $\mathcal{G}_a$ of smooth functions $g_a\colon\widehat{W}\to\mathbb{R}$ such that there are positive numbers ϵ',δ,c with $\delta<1$ depending on g_a with

- $g_a=-\epsilon'$ on $W^{1-\delta}$;
- on $\widehat{W}\backslash W^{1-\delta}$, the function g_a depends only on r and there holds $g_a''(r)\geq 0$;
- $g_a=a(r-1)-c$ on $\widehat{W}\backslash W$.

A non-constant one-periodic orbit of g_a sits in ∂W^r for $r>1-\delta$ such that $g'(r)\in\operatorname{spec}(R_\lambda)$, and corresponds to a closed Reeb orbit of $(\partial W,\lambda|_{\partial W})$ with period $g'(r)$. If we consider the piecewise linear function

$$\bar{g}_a\colon\widehat{W}\to\mathbb{R},\quad \bar{g}_a|_W=0,\quad \bar{g}_a|_{\widehat{W}\backslash W}=a(r-1),$$

then choosing ϵ',δ,c small enough, the function g_a can be arbitrarily C^0-close to $\bar{g}_a$ on $\widehat{W}$ and C^∞-close to $\bar{g}_a$ away from ∂W, and furthermore the action of a non-constant one-periodic orbit of g_a can be arbitrarily close to the period of the corresponding Reeb orbit by (3.7).

Lemma 3.1. *Let ϵ, a, b be real numbers such that*

$$0 < \epsilon < \min \operatorname{spec}(R_\lambda) < a < b, \quad a, b \notin \operatorname{spec}(R_\lambda).$$

There exist $g_a \in \mathcal{G}_a$ and $g_b \in \mathcal{G}_b$ which can be taken arbitrarily C^0-close to $\bar{g}_a$ and $\bar{g}_b$ respectively such that the following diagram commutes

$$\begin{tikzcd}
\mathrm{HF}^{(\epsilon,b)}(g_b;\alpha) \arrow[r, "\cong", "\phi"'] & \mathrm{SH}^{(\epsilon,b)}(W;\alpha) \\
\mathrm{HF}^{(\epsilon,b)}(g_a;\alpha) \arrow[u, "\Phi"] \arrow[ur, bend right, "\phi"'] & \\
\mathrm{HF}^{(\epsilon,a)}(g_a;\alpha) \arrow[u, "\cong", "\iota"'] \arrow[r, "\cong", "\phi"'] & \mathrm{SH}^{(\epsilon,a)}(W;\alpha). \arrow[uu, "\iota_\epsilon^{a,b}"']
\end{tikzcd} \tag{3.11}$$

Here Φ is a continuation homomorphism, ι is a homomorphism from (3.4), $\iota_\epsilon^{a,b}$ is from (3.10), and the maps ϕ are homomorphisms in the direct system.

Proof. For any increasing sequence $(a_i)_{i\in\mathbb{N}} \subset (0,\infty)\backslash\operatorname{spec}(R_\lambda)$ with $a_1 = a$, we choose a sequence of functions $g_{a_i} \in \mathcal{G}_{a_i}$ such that (g_{a_i}) is cofinal in $\mathcal{H}^{\epsilon,a}$ with $g_{i+1} \geq g_i$ and for each $i \in \mathbb{N}$ there is a monotone homotopy $(g^s)_{s\in[0,1]}$ from g_{a_i} to $g_{a_{i+1}}$ with the property that for all $s \in [0,1]$, the function g^s has no one-periodic orbit x with $\mathcal{A}_{g^s}(x) \in \{\epsilon, a\}$. Then, the continuation homomorphism induced by (g^s),

$$\Phi\colon \mathrm{HF}^{(\epsilon,a)}(g_{a_i};\alpha) \xrightarrow{\cong} \mathrm{HF}^{(\epsilon,a)}(g_{a_{i+1}};\alpha)$$

is an isomorphism, see for instance [68, Lemma 2.8], and thus the lower horizontal arrow in (3.11) is an isomorphism. The same argument shows that the upper horizontal map is also an isomorphism. Moreover, the map ι is an isomorphism since $g_{a_1} = g_a$ does not have one-periodic orbits with action greater than a. Finally, the commutativity of the diagram follows from the definitions of the involved homomorphisms. □

Remark 3.2. The statement of Lemma 3.1 holds *mutatis mutandis* with ϵ replaced by $-\infty$ and $a < b$ any pair of positive numbers not in $\operatorname{spec}(R_\lambda)$.

3.3. Capacities from Floer and symplectic homology

We now define the capacities mentioned in Sect. 1. To define the spectral invariant, we take a function $f\colon \widehat{W} \to \mathbb{R}$ in $\mathcal{G}_\epsilon$ where $0 < \epsilon < \min \operatorname{spec}(R_\lambda)$. As discussed above, we have

$$\mathrm{HF}(f;\mathbb{1}) \cong \mathrm{SH}^{(-\infty,\epsilon)}(W;\mathbb{1}) \cong \mathrm{H}(W,\partial W).$$

We denote by $e_f \in \mathrm{HF}(f;\mathbb{1})$ the homology class corresponding to the fundamental class in $\mathrm{H}(W,\partial W)$ through the above isomorphism. Let $H\colon S^1 \times W \to \mathbb{R}$ be a smooth Hamiltonian supported in $S^1 \times (W\backslash\partial W)$ whose Floer homology is defined as in (3.6). For $a \in \mathbb{R}\backslash\operatorname{spec}(H)$, we consider the chain of homomorphisms

$$\mathrm{HF}(f;\mathbb{1}) \xrightarrow{\Phi_{f,H}} \mathrm{HF}(H;\mathbb{1}) \xrightarrow{\pi_a} \mathrm{HF}^{(a,\infty)}(H;\mathbb{1}), \tag{3.12}$$

where $\Phi_{f,H}$ is a continuation homomorphism, which is in fact an isomorphism since f and H have the same slope. The map π_a is from (3.4). The spectral invariant of H is defined by

$$\sigma(H) := \inf\{a \mid \pi_a \circ \Phi_{f,H}(e_f) = 0\}.$$

The spectral capacity $c_\sigma(W)$ of W is defined by the supremum of $\sigma(H)$ over all smooth Hamiltonians $H\colon S^1 \times W \to \mathbb{R}$ supported in $S^1 \times (W\backslash\partial W)$.

Due to (3.8), we can view the homomorphism $\iota_{-\infty}^{\epsilon,c}$ defined in (3.10) as a map

$$\iota_{-\infty}^{\epsilon,c}\colon \mathrm{H}(W,\partial W) \longrightarrow \mathrm{SH}^{(-\infty,c)}(W;\mathbb{1}).$$

We consider the number

$$c_{\mathrm{SH}}(W) := \inf\left\{c > 0 \;\middle|\; \iota_{-\infty}^{\epsilon,c} = 0\right\} \in (0,\infty]. \tag{3.13}$$

We note that due to functoriality $\iota_{-\infty}^{\epsilon,c} = 0$ for any $c > c_{\mathrm{SH}}(W)$. It is known that $\mathrm{SH}(W;\mathbb{1})$ admits a ring structure with unit given by the image of the fundamental class of $\mathrm{H}(W,\partial W)$ under $\iota_{-\infty}^{\epsilon,\infty}$, thus the quantity $c_{\mathrm{SH}}(W)$ is finite if and only if $\mathrm{SH}(W;\mathbb{1})$ vanishes. Using positive symplectic homology we can also define the quantity

$$c_{\mathrm{SH}^+}(W;\alpha) := \inf\{c > 0 \mid \iota_\epsilon^{c,\infty}\colon \mathrm{SH}^{(\epsilon,c)}(W;\alpha) \to \mathrm{SH}^+(W;\alpha) \text{ is non-zero}\} \in (0,\infty]. \tag{3.14}$$

It is finite if and only if $\mathrm{SH}^+(W;\alpha) \neq 0$. This is equivalent to $\mathrm{SH}(W;\alpha) \neq 0$ for $\alpha \neq \mathbb{1}$ and to $\mathrm{SH}(W;\mathbb{1}) \ncong \mathrm{H}(W,\partial W)$ for $\alpha = \mathbb{1}$ due to (3.9). We use the notation $c_{\mathrm{SH}^+}(W) = c_{\mathrm{SH}^+}(W;\mathbb{1})$.

Lemma 3.3. *There holds* $c_{\mathrm{SH}}(W) \geq c_{\mathrm{SH}^+}(W)$. *Moreover, the equality holds if* $\mathrm{SH}(W;\mathbb{1}) = 0$ *and* $\mathrm{rk}\,\mathrm{H}(W,\partial W) = 1$.

Proof. This is an immediate consequence of the commutative diagram induced by (3.9):

$$\begin{tikzcd}
\cdots \arrow[r] & \mathrm{SH}^{(\epsilon,c)}(W;\mathbb{1}) \arrow[r, "\delta_1"] \arrow[d, "\iota_\epsilon^{c,\infty}"] & \mathrm{H}(W,\partial W) \arrow[r, "\iota_{-\infty}^{\epsilon,c}"] \arrow[d, "\iota_{-\infty}^{\epsilon,\epsilon}=\mathrm{id}"] & \mathrm{SH}^{(-\infty,c)}(W;\mathbb{1}) \arrow[r] \arrow[d, "\iota_{-\infty}^{c,\infty}"] & \cdots \\
\cdots \arrow[r] & \mathrm{SH}^+(W;\mathbb{1}) \arrow[r, "\delta_2"] & \mathrm{H}(W,\partial W) \arrow[r, "\iota_{-\infty}^{\epsilon,\infty}"] & \mathrm{SH}(W;\mathbb{1}) \arrow[r] & \cdots
\end{tikzcd}$$

For any $c > c_{\mathrm{SH}}(W)$, we have $\iota_{-\infty}^{\epsilon,c} = 0$. Thus, δ_1 and also $\iota_\epsilon^{c,\infty}$ are non-zero. This shows that $c_{\mathrm{SH}^+}(W) \leq c$, and hence $c_{\mathrm{SH}}(W) \geq c_{\mathrm{SH}^+}(W)$.

Suppose $\mathrm{SH}(W;\mathbb{1}) = 0$ and $\mathrm{rk}\,\mathrm{H}(W,\partial W) = 1$. Then, $\iota_{-\infty}^{\epsilon,c} = 0$ if and only if $\delta_1 \neq 0$, and this is equivalent also to $\iota_\epsilon^{c,\infty} \neq 0$ since δ_2 is an isomorphism. □

Before showing the announced results about $c_{\mathrm{SH}^+}(W)$ in Sect. 4, we prove Theorem 1.3, which asserts $c_\sigma(W) \leq c_{\mathrm{SH}}(W)$.

Proof of Theorem 1.3

It suffices to show that $\sigma(H) \leq c_{\mathrm{SH}}(W)$ for every smooth Hamiltonian $H\colon S^1 \times W \to \mathbb{R}$ with support in $S^1 \times (W \backslash \partial W)$ when $c_{\mathrm{SH}}(W)$ is finite. For any $a > c_{\mathrm{SH}}(W)$ not belonging to $\mathrm{spec}(R_\lambda) \cup \mathrm{spec}(H)$, which is a closed nowhere dense set, we extend H to a smooth function $\widetilde{H}\colon S^1 \times \widehat{W} \to \mathbb{R}$ in the same manner as in defining $\widehat{H}$ in (3.6) but with $h'(r) = \epsilon$ replaced by $h'(r) = a$. The homomorphisms in (3.12) can be completed to a commutative diagram

$$\begin{tikzcd}
 & \mathrm{HF}(\widetilde{H};\mathbb{1}) \arrow[r, "\widetilde{\pi}_a"] & \mathrm{HF}^{(a,\infty)}(\widetilde{H};\mathbb{1}) \\
\mathrm{HF}(f;\mathbb{1}) \arrow[ur, "\Phi_{f,\widetilde{H}}"] \arrow[dr, "\Phi_{f,H}"'] & & \\
 & \mathrm{HF}(H;\mathbb{1}) \arrow[uu] \arrow[r, "\pi_a"] & \mathrm{HF}^{(a,\infty)}(H;\mathbb{1}) \arrow[uu, "\cong"', "\Phi"]
\end{tikzcd}$$

where vertical arrows are continuation homomorphisms, and Φ is even an isomorphism since H and $\widetilde{H}$ have the same 1-periodic orbits in the action-window (a,∞). We claim

$$\Phi_{f,\widetilde{H}}(e_f) = 0.$$

Once the claim is verified, the diagram shows that $\pi_a \circ \Phi_{f,H}(e_f) = 0$ which implies $\sigma(H) \leq a$ and hence $\sigma(H) \leq c_{\mathrm{SH}}(W)$ as we wanted. The claim now is a consequence of the fact that by Lemma 3.1 and Remark 3.2, there is a commutative diagram

$$\begin{tikzcd}
\mathrm{HF}(\widetilde{H};\mathbb{1}) \arrow[r, "\Phi_{\widetilde{H},g}", "\cong"'] & \mathrm{HF}(g;\mathbb{1}) \arrow[r, "\cong"'] & \mathrm{SH}^{(-\infty,a)}(W;\mathbb{1}) \\
 & \mathrm{HF}(f;\mathbb{1}) \arrow[ul, "\Phi_{f,\widetilde{H}}"] \arrow[u, "\Phi_{f,g}"] \arrow[r, "\cong"'] & \mathrm{SH}^{(-\infty,\epsilon)}(W;\mathbb{1}) \arrow[u, "\iota^{\epsilon,a}_{-\infty}"']
\end{tikzcd} \tag{3.15}$$

for some function $g \in \mathcal{G}_a$, where $\mathrm{HF}(g;\mathbb{1}) = \mathrm{HF}^{(-\infty,a)}(g;\mathbb{1})$ and $\mathrm{HF}(f;\mathbb{1}) = \mathrm{HF}^{(-\infty,\epsilon)}(f;\mathbb{1})$ since g and f have no one-periodic orbits outside the action-windows $(-\infty,a)$ and $(-\infty,\epsilon)$, respectively, by (3.7). Notice that the triangular diagram is commutative since the maps involved are continuation maps and that the horizontal arrow $\Phi_{\widetilde{H},g}$ is an isomorphism since $\widetilde{H}$ and g have the same slope. Now, $\iota^{\epsilon,a}_{-\infty} = 0$ since $a > c_{\mathrm{SH}}(W)$, and therefore, the claim $\Phi_{f,\widetilde{H}}(e_f) = 0$ follows by commutativity of the diagram. $\square$

4. Proofs of the main results

In this section, we will be working exclusively with contractible loops. Therefore, we will omit the symbol $\mathbb{1}$ from the notation to make formulas more readable and write for instance $\mathrm{HF}(H)$ and $\mathrm{SH}(W)$ for $\mathrm{HF}(H;\mathbb{1})$ and $\mathrm{SH}(W;\mathbb{1})$ respectively. We start by proving the following fundamental result, which is an adaptation of [22, Proposition 5.2] to our setting.

Proposition 4.1. *Let $H \in \widetilde{\mathcal{H}}(W, W_{\mathrm{sk}})$ and let $a \in (0, -\min H)\backslash \mathrm{spec}(R_\lambda)$. We assume that all elements of the set*

$$\Gamma := \{x \in \mathcal{P}_1^o(H) \mid -\min H < \mathcal{A}_H(x) < -\min H + a\}$$

are nondegenerate. Then, there holds

$$\#\Gamma \geq \mathrm{rk}\left[\iota_\epsilon^{a,\infty} \colon \mathrm{SH}^{(\epsilon,a)}(W) \to \mathrm{SH}^+(W)\right],$$

where $0 < \epsilon < \min \mathrm{spec}(R_\lambda)$ as usual.

Proof. Let us consider H as in the statement. There is no loss of generality in assuming $-\min H + a \notin \mathrm{spec}(H)$. Indeed since $\mathrm{spec}(R_\lambda)$ is closed, for $a' < a$ sufficiently close to a, we have $a' \notin \mathrm{spec}(R_\lambda)$, $\mathrm{rk}\, \iota_\epsilon^{a',\infty} = \mathrm{rk}\, \iota_\epsilon^{a,\infty}$. We choose $\delta > 0$ such that

$$H|_{W^\delta} = -\min H, \quad H|_{\widehat{W}\backslash W^{1-\delta}} = 0.$$

Let $\widehat{H} \colon \widehat{W} \to \mathbb{R}$ be a smooth function such that $\widehat{H} = H$ on $W^{1-\delta/2}$ and $\widehat{H} = \widehat{h}$ on $\widehat{W}\backslash W^{1-\delta/2}$, where $\widehat{h}$ is a smooth function depending only on r such that

$$\widehat{h}'' \geq 0, \quad \widehat{h}|_{\widehat{W}\backslash W} = a(r-1) + c$$

for some $c > 0$ small enough. All one-periodic orbits of $\widehat{H}$ that are not one-periodic orbits of H have action less than a by (3.7):

$$\{x \in \mathcal{P}_1(\widehat{H}) \mid \mathcal{A}_{\widehat{H}}(x) > -\min H\} = \{x \in \mathcal{P}_1(H) \mid \mathcal{A}_H(x) > -\min H\}.$$

In order to relate the Floer homology of H with the positive symplectic homology of W, we introduce two auxiliary functions. First, we choose a smooth function $k_b \colon \widehat{W} \to \mathbb{R}$ which is obtained by smoothening the piecewise linear function that is equal to $\min H$ on W^η for $\eta < \delta$ and to $b(r-\eta) + \min H$ for $b \in \mathbb{R}\backslash \mathrm{spec}(R_\lambda)$ on $W\backslash W^\eta$. The function k_b depends only on r on $\widehat{W}\backslash W_{\mathrm{sk}}$ and $k_b''(r) \geq 0$, see Fig. 1. Taking b large enough, we have $k_b \geq \widehat{H}$. The constant one-periodic orbits of k_b have action equal to $-\min H$. We also take ϵ and η small enough so that the following action estimate holds by (3.7):

$$-\min H + \epsilon < \mathcal{A}_{k_b}(x) < -\min H + a \quad \forall x \in \mathcal{P}_1(k_b)\backslash \mathrm{Crit}\, k_b. \tag{4.1}$$

Similarly, we take $f_a \colon \widehat{W} \to \mathbb{R}$ to be a convex, smooth approximation of the piecewise linear function which is equal to $\min H$ on W and equal to $a(r-1) + \min H$ on $\widehat{W}\backslash W$, see Fig. 1. We have $f_a \leq \widehat{H}$. All constant one-periodic orbits of f_a have action $-\min H$ and

$$-\min H + \epsilon < \mathcal{A}_{f_a}(x) < -\min H + a \quad \forall x \in \mathcal{P}_1(f_a)\backslash \mathrm{Crit}\, f_a. \tag{4.2}$$

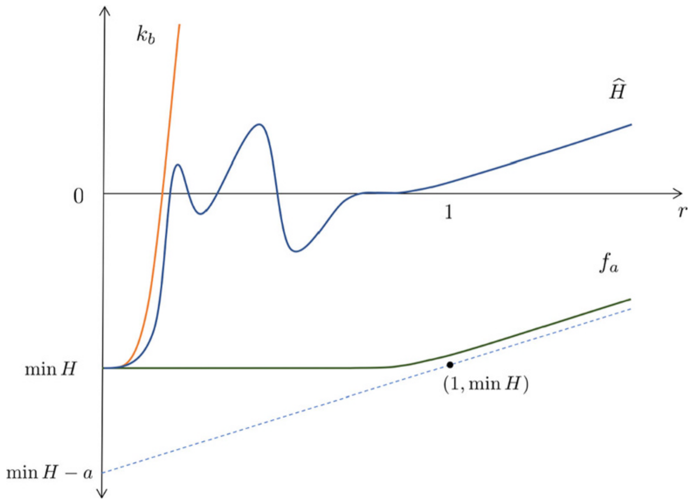


FIGURE 1. The Hamiltonians $\widehat{H}$, k_b and f_a

We claim that there exists a commutative diagram

$$\begin{tikzcd}
& \mathrm{HF}^{(-\min H+\epsilon/2,-\min H+a)}(k_b) \arrow[r, "\cong"] & \mathrm{SH}^{(\epsilon,b)}(W) \\
\mathrm{HF}^{(-\min H+\epsilon/2,-\min H+a)}(\widehat{H}) \arrow[ur, "\Phi_2"] & & \\
& \mathrm{HF}^{(-\min H+\epsilon/2,-\min H+a)}(f_a) \arrow[ul, "\Phi_1"] \arrow[uu, "\Phi_3"] \arrow[r, "\cong"] & \mathrm{SH}^{(\epsilon,a)}(W), \arrow[uu, "\iota_\epsilon^{a,b}"']
\end{tikzcd} \tag{4.3}$$

where $\epsilon > 0$ is chosen so that $-\min H+\epsilon/2 \notin \operatorname{spec}(\widehat{H})$. The maps Φ_1, Φ_2, and Φ_3 are continuation homomorphisms induced by monotone homotopies and satisfy $\Phi_2 \circ \Phi_1 = \Phi_3$. Once the diagram (4.3) is established, the proposition follows from

$$\begin{aligned}
\#\Gamma = \#\big\{x \in \mathcal{P}_1^o(\widehat{H}) \mid -\min H < \mathcal{A}_{\widehat{H}}(x) < -\min H + a\big\} \\
\geq \operatorname{rk} \mathrm{HF}^{(-\min H+\epsilon/2,-\min H+a)}(\widehat{H}) \\
\geq \operatorname{rk}\Big[\iota_\epsilon^{a,b}\colon \mathrm{SH}^{(\epsilon,a)}(W) \to \mathrm{SH}^{(\epsilon,b)}(W)\Big] \\
\geq \operatorname{rk}\Big[\iota_\epsilon^{a,\infty}\colon \mathrm{SH}^{(\epsilon,a)}(W) \to \mathrm{SH}^{+}(W)\Big],
\end{aligned}$$

where the second inequality is due to (4.3), and the last inequality holds by the identity $\iota_\epsilon^{a,\infty} = \iota_\epsilon^{b,\infty} \circ \iota_\epsilon^{a,b}$.

Let us now define the horizontal isomorphisms in (4.3) and show that the rectangular diagram commutes. To this purpose, we define

$$\widetilde{k}_b := k_b - \min H - \epsilon/2, \quad \widetilde{f}_a := f_a - \min H - \epsilon/2.$$

We also consider a smooth family of functions $\widetilde{k}_b^s\colon \widehat{W} \to \mathbb{R}$ for $s \in [0, 1-\eta]$ such that

- $\widetilde{k}_b^s = -\epsilon/2$ on $W^{\eta+s}$,
- $\widetilde{k}_b^s(r) := \widetilde{k}_b^s(r-s)$ for $r > \delta + s$.

We note that $\widetilde{k}_b^0 = \widetilde{k}_b$ and $\widetilde{k}_b^{1-\eta} \geq \widetilde{f}_a$. We define the rectangular diagram in (4.3) as composition of the diagrams:

$$\begin{tikzcd}
\mathrm{HF}^{(-\min H+\epsilon/2,-\min H+a)}(k_b) \arrow[r, "\cong", "\Phi_4"'] & \mathrm{HF}^{(\epsilon,\infty)}(\widetilde{k}_b) & \mathrm{HF}^{(\epsilon,\infty)}(\widetilde{k}_b^{1-\eta}) \arrow[l, "\cong"', "\Phi_7"] \arrow[r, "\cong"] & \mathrm{SH}^{(\epsilon,b)}(W) \\
\mathrm{HF}^{(-\min H+\epsilon/2,-\min H+a)}(f_a) \arrow[u, "\Phi_3"] \arrow[r, "\cong", "\Phi_5"'] & \mathrm{HF}^{(\epsilon,\infty)}(\widetilde{f}_a) \arrow[u, "\Phi_6"] \arrow[r, "\mathrm{id}"'] & \mathrm{HF}^{(\epsilon,\infty)}(\widetilde{f}_a) \arrow[u, "\Phi_8"] \arrow[r, "\cong"] & \mathrm{SH}^{(\epsilon,a)}(W) \arrow[u, "\iota_\epsilon^{a,b}"']
\end{tikzcd} \tag{4.4}$$

Since k_b and f_a do not have one-periodic orbits with action greater than $-\min H + a$,

$$\mathrm{HF}^{(-\min H+\epsilon/2,-\min H+a)}(k_b) = \mathrm{HF}^{(-\min H+\epsilon/2,\infty)}(k_b),$$
$$\mathrm{HF}^{(-\min H+\epsilon/2,-\min H+a)}(f_a) = \mathrm{HF}^{(-\min H+\epsilon/2,\infty)}(f_a).$$

The maps Φ_4 and Φ_5 are canonical isomorphisms: The functions k_b (resp. f_a) and $\widetilde{k}_b$ (resp. $\widetilde{f}_a$) have the same one-periodic orbits with action shifted by $-\min H - \epsilon/2$ and the same Floer cylinders. These can also be understood as continuation maps of monotone homotopies. The map Φ_3 is a continuation homomorphism, and Φ_6 equals Φ_3 with action shifted by $-\min H - \epsilon/2$. Thus, leftmost rectangle readily commutes. The monotone homotopy $\widetilde{k}_b^s$ between $\widetilde{k}_b^{1-\eta}$ and $\widetilde{k}_b^0 = \widetilde{k}_b$ has no one-periodic orbit with action equal to ϵ for all s. Therefore, the continuation homomorphism Φ_7 induced by $\widetilde{k}_b^s$ is an isomorphism. The map Φ_8 is also a continuation homomorphism induced by a monotone homotopy, and the rectangle in the middle commutes since all maps are continuation homomorphisms. Finally, the rightmost rectangle follows from (3.11) since $\widetilde{f}_a$ and $\widetilde{k}_b^{1-\eta}$ can be taken as g_a and g_b respectively given in Lemma 3.1 and, again by action reasons,

$$\mathrm{HF}^{(\epsilon,\infty)}(\widetilde{k}_b^{1-\eta}) = \mathrm{HF}^{(\epsilon,b)}(\widetilde{k}_b^{1-\eta}), \quad \mathrm{HF}^{(\epsilon,\infty)}(\widetilde{f}_a) = \mathrm{HF}^{(\epsilon,a)}(\widetilde{f}_a).$$

□

4.1. Proof of Theorem 2.2

The statement is void if $c_{\mathrm{SH}^+}(W) = \infty$. Thus, we suppose $c_{\mathrm{SH}^+}(W) < \infty$. It is enough to show

$$\tilde{c}^o_{\mathrm{HZ}}\big(W, W_{\mathrm{sk}}, c_{\mathrm{SH}^+}(W)\big) \leq c_{\mathrm{SH}^+}(W).$$

We assume by contradiction that there is $H \in \widetilde{\mathcal{H}}(W, W_{\mathrm{sk}})$ such that

$$-\min H > c_{\mathrm{SH}^+}(W), \quad \mathcal{A}_H(x) \notin (-\min H, -\min H + c_{\mathrm{SH}^+}(W)] \quad \forall x \in \mathcal{P}_1^o(H).$$

Since $\mathrm{spec}(H)$ is closed and $\mathrm{spec}(R_\lambda)$ is nowhere dense, there exists $a \in (c_{\mathrm{SH}^+}(W), -\min H)$, $a \notin \mathrm{spec}(R_\lambda)$ such that

$$\mathcal{A}_H(x) \notin (-\min H, -\min H + a) \quad \forall x \in \mathcal{P}_1^o(H).$$

This contradicts Proposition 4.1, and thus the theorem is proved. $\square$

4.2. Proof of Theorem 2.9(a)

Let $H\colon S^1\times W\to\mathbb{R}$ be a smooth Hamiltonian with support inside $S^1\times(W\backslash\partial W)$ and such that $H|_{S^1\times W_{\mathrm{sk}}}<0$. We extend H smoothly to $\widehat{H}\colon S^1\times\widehat{W}\to\mathbb{R}$ as in the definition of $\mathrm{HF}^{(a,b)}(H)=\mathrm{HF}^{(a,b)}(\widehat{H})$ in (3.6). We consider two piecewise linear functions

$$\bar f\colon\widehat{W}\longrightarrow\mathbb{R},\quad \bar f|_W=-c,\quad \bar f|_{\widehat{W}\backslash W}=\epsilon(r-1)-c$$
$$\bar k\colon\widehat{W}\longrightarrow\mathbb{R},\quad \bar k|_{W^\delta}=-d,\quad \bar k|_{\widehat{W}\backslash W^\delta}=b(r-\delta)-d$$

for some positive numbers $c,d,b,\delta>0$ with $b\notin\mathrm{spec}(R_\lambda)$ and for some $0<\epsilon<\min\mathrm{spec}(R_\lambda)$. Smoothening $\bar f\colon\widehat{W}\to\mathbb{R}$ near ∂W and $\bar k\colon\widehat{W}\to\mathbb{R}$ near ∂W^δ, we obtain smooth functions $f\colon\widehat{W}\to\mathbb{R}$ and $k\colon\widehat{W}\to\mathbb{R}$ respectively, both of which depend only on r on $\widehat{W}\backslash W_{\mathrm{sk}}$ and convex. The assumption on H ensures that for large $b,c>0$ and for small $d,\delta>0$, we have

$$f(z)\le\widehat{H}(t,z)\le k(z)\quad\forall(t,z)\in S^1\times\widehat{W}.$$

Then, for any $a\in(0,d)\backslash\mathrm{spec}(H)$, we have the commutative diagram

$$\begin{tikzcd}
 & \mathrm{HF}^{(a,\infty)}(k) \arrow[r, "\cong"] & \mathrm{SH}^{(-\infty,b)}(W) \\
\mathrm{HF}^{(a,\infty)}(\widehat{H}) \arrow[ur, "\Phi_2"] & & \\
 & \mathrm{HF}^{(a,\infty)}(f) \arrow[r, "\cong"] \arrow[uu, "\Phi_3"] \arrow[ul, "\Phi_1"'] & \mathrm{SH}^{(-\infty,\epsilon)}(W), \arrow[uu, "\iota_{-\infty}^{\epsilon,b}"']
\end{tikzcd} \tag{4.5}$$

where the Φ's are continuation homomorphisms induced by monotone homotopies. The rectangular diagram is obtained as the rightmost one in (4.4) noticing that $\mathrm{HF}^{(a,\infty)}(k)=\mathrm{HF}(k)$ and $\mathrm{HF}^{(a,\infty)}(f)=\mathrm{HF}(f)$ as k and f do not have one-periodic orbits with action less than a by (3.7). The diagram in (4.5) readily yields

$$\mathrm{rk}\,\mathrm{HF}^{(a,\infty)}(H)=\mathrm{rk}\,\mathrm{HF}^{(a,\infty)}(\widehat{H})\ge\mathrm{rk}\,\Phi_3=\mathrm{rk}\,\iota_{-\infty}^{\epsilon,b}=\mathrm{rk}\,\iota_{-\infty}^{\epsilon,\infty}$$

where the last equality holds for large $b>0$. Since $\mathrm{SH}^{(-\infty,\epsilon)}(W)\cong\mathrm{H}(W,\partial W)$ by (3.8), the proof is complete. $\square$

4.3. Proof of Theorem 2.9(b)

Let $H\colon S^1\times W\to(-\infty,0]$ be a smooth Hamiltonian supported in $S^1\times(W\backslash\partial W)$ such that $H|_{S^1\times W_{\mathrm{sk}}}<0$. We extend H to $\widehat{H}\colon S^1\times\widehat{W}\to\mathbb{R}$ as in the proof of Theorem 2.9(a). Let $k\colon\widehat{W}\to\mathbb{R}$ be a smooth function obtained also as before by smoothening a piecewise linear function

$$\bar k\colon\widehat{W}\longrightarrow\mathbb{R},\quad \bar k|_{W^\delta}=-d,\quad \bar k|_{\widehat{W}\backslash W^\delta}=\epsilon(r-\delta)-d$$

for $d,\delta>0$ and $0<\epsilon<\min\mathrm{spec}(R_\lambda)$. We take $d,\delta>0$ small enough and choose a sufficiently large $c>0$ to satisfy

$$(k-c)(z)\le\widehat{H}(t,z)\le k(z)\quad\forall(t,z)\in S^1\times\widehat{W}.$$

Then, for any $a \in (0,d)\backslash \operatorname{spec}(H)$, the following diagram commutes:

$$\begin{tikzcd} & \mathrm{HF}^{(a,\infty)}(k). \\ \mathrm{HF}^{(a,\infty)}(\widehat{H}) \arrow[ur, "\Phi_2"] & \\ & \mathrm{HF}^{(a,\infty)}(k-c) \arrow[ul, "\Phi_1"] \arrow[uu, "\cong"', "\Phi_3"] \end{tikzcd}$$

where Φ's are continuation homomorphisms induced by monotone homotopies. Moreover, Φ_3 is an isomorphism since k and $k-c$ have the same slope and possess no one-periodic orbits with action less than a. Hence we conclude that Φ_2 is a surjective homomorphism, and this finishes the proof since $\mathrm{HF}^{(a,\infty)}(k) = \mathrm{HF}(k) \cong \mathrm{H}(W;\partial W)$. □

4.4. Proof of Corollary 2.8

Let Q be a closed manifold endowed with a metric g and denote by D^*Q the associated unit-disc cotangent bundle. The foot-point projection $\pi\colon T^*Q \to Q$ gives a bijection between periodic Reeb orbits on $\partial(D^*Q)$ with respect to the canonical one-form λ_{T^*Q} and closed geodesics on Q where the periods of Reeb orbits correspond to the length of geodesics. In particular if $\ell_{\mathbb{1}}$ denotes the length of the shortest non-constant, contractible closed geodesic, then

$$\min \operatorname{spec}(R_\lambda, \mathbb{1}) = \ell_{\mathbb{1}}.$$

Let $\epsilon > 0$ be a positive number smaller than this common value. Consider the square root of the energy functional on the loop space of contractible loops

$$\mathcal{E}\colon \mathcal{L}_{\mathbb{1}}Q \to \mathbb{R}, \quad \mathcal{E}(x) = \Big(\int_0^1 \|\dot{x}(t)\|_g^2 \,\mathrm{d}t\Big)^{1/2},$$

where $\|\cdot\|_g$ is the norm induced by g. The functional $\mathcal{E}$ coincides with the length on the set of geodesics and yields a filtration $\mathrm{H}^{(\epsilon,a)}(\mathcal{L}_{\mathbb{1}}Q)$ of the singular homology of the loop space for $a \geq \epsilon$ together with inclusion homomorphisms

$$\jmath_\epsilon^{a,b}\colon \mathrm{H}^{(\epsilon,a)}(\mathcal{L}_{\mathbb{1}}Q) \to \mathrm{H}^{(\epsilon,b)}(\mathcal{L}_{\mathbb{1}}Q)$$

for $\epsilon \leq a \leq b$. By the action filtration version of Viterbo isomorphism [68, Theorem 2.9], we have the commutative diagram

$$\begin{tikzcd} \mathrm{SH}^{(\epsilon,b)}(D^*Q) \arrow[r, "\cong"] & \mathrm{H}^{(\epsilon,b)}(\mathcal{L}_{\mathbb{1}}Q) \\ \mathrm{SH}^{(\epsilon,a)}(D^*Q) \arrow[r, "\cong"] \arrow[u, "\iota_\epsilon^{a,b}"] & \mathrm{H}^{(\epsilon,a)}(\mathcal{L}_{\mathbb{1}}Q). \arrow[u, "\jmath_\epsilon^{a,b}"'] \end{tikzcd} \tag{4.6}$$

By definition of ϵ, we have $\mathrm{SH}^+(D^*Q) = \mathrm{SH}^{(\epsilon,\infty)}(D^*Q)$ and $\mathrm{H}(\mathcal{L}_{\mathbb{1}}Q, Q) = \mathrm{H}^{(\epsilon,\infty)}(\mathcal{L}_{\mathbb{1}}Q)$. Thus,

$$c_{\mathrm{SH}^+}(D^*Q) = c(\mathcal{E}) := \inf\{a > 0 \mid \jmath_\epsilon^{a,\infty} \neq 0\}.$$

Since $\ell_{\mathbb{1}} \leq c_{\mathrm{SH}^+}(D^*Q)$ by constructing a suitable radial Hamiltonian, we are left to show $\ell_{\mathbb{1}} \geq c(\mathcal{E})$ in the two cases mentioned in the statement.

- Let (Q,g) be a closed, non-aspherical homogeneous space. Since Q is non-aspherical, $\ell_{\mathbb{1}}$ is finite by the classical Lusternik–Fet theorem and the set of closed, non-constant, contractible geodesics with length $\ell_{\mathbb{1}}$ is non-empty. By [71, Theorem 5], the map
$$\jmath_{\epsilon}^{\ell_{\mathbb{1}}+\epsilon,\infty}\colon \mathrm{H}^{(\epsilon,\ell_{\mathbb{1}}+\epsilon)}(\mathcal{L}_{\mathbb{1}}Q) \to \mathrm{H}(\mathcal{L}_{\mathbb{1}}Q,Q)$$
is non-zero for small ϵ small. Thus, $\ell_{\mathbb{1}}+\epsilon \geq c(\mathcal{E})$ and the result follows letting ϵ to 0.
- Let (Q,g) be a two-sphere with strictly positive Gaussian curvature. Abbondandolo and Mazzucchelli show in Lemma A.2 below that there is a continuous path $u\colon [-1,1] \to \{\mathcal{E} \leq \ell_{\mathbb{1}}\}$ with $u(-1),u(1) \in Q$ representing a non-trivial element in $\mathrm{H}_1(\mathcal{L}_{\mathbb{1}}Q,Q)$. Thus, $\jmath_{\epsilon}^{\ell_{\mathbb{1}}+\epsilon,\infty} \neq 0$ for all ϵ sufficiently small and we conclude that $\ell_{\mathbb{1}} \geq c(\mathcal{E})$.

Acknowledgements

G. Benedetti acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the projects: 281869850 (RTG 2229); 390900948 (EXC-2181/1); 281071066 (SFB/TRR 191). J. Kang was supported by Research Resettlement Fund for the new faculty of Seoul National University and by TJ Park Science Fellowship of POSCO TJ Park Foundation. We are grateful to Thomas Rot for discussions around Remark 2.6, to Pierre-Alexandre Mailhot and Egor Shelukhin and to Viktor Ginzburg and Kei Irie for their feedback on this paper.

Appendix: The monotonicity of the systole of convex Riemannian two-spheres (by Alberto Abbondandolo[1] and Marco Mazzucchelli)[2]

Throughout this appendix, the notion of convexity must be understood in the differentiable sense: A compact three-ball $B \subset \mathbb{R}^3$ with smooth boundary is strictly convex when there exists a smooth function $F\colon \mathbb{R}^3 \to [0,\infty)$ with positive definite Hessian at every point and such that $\partial B = F^{-1}(1)$. Equivalently, the boundary sphere $M = \partial B$, which will always be equipped with the Riemannian metric g that is the restriction of the ambient Euclidean metric, has strictly positive Gaussian curvature. The systole $\mathrm{sys}(M) > 0$ is the length of the shortest closed geodesic of (M,g). The main result of this appendix answers in dimension 3 a question that was posed to us by Yaron Ostrover:

Theorem A.1. *Let $B_1 \subseteq B_2$ be two compact strictly convex three-balls in $\mathbb{R}^3$ with smooth boundary. Then,* $\mathrm{sys}(\partial B_1) \leq \mathrm{sys}(\partial B_2)$.

[1] Ruhr Universität Bochum, Fakultät für Mathematik, alberto.abbondandolo@rub.de.
[2] CNRS, École Normale Supérieure de Lyon, UMPA, marco.mazzucchelli@ens-lyon.fr.

The main ingredient of the proof is the observation that the systole of positively curved Riemannian two-spheres coincides with the classical Birkhoff min–max, as we will now prove. Let (M,g) be a Riemannian two-sphere. We denote the energy functional on the $W^{1,2}$ free-loop space by

$$E\colon \Lambda M = W^{1,2}(S^1,M) \to [0,\infty), \quad E(\zeta) = \int_{S^1} \|\dot\zeta(t)\|_g^2 \mathrm{d}t.$$

Here and in the following, we denote by $S^1 = \mathbb{R}/\mathbb{Z}$ the 1-periodic circle. We consider the unit sphere $S^2 \subset \mathbb{R}^3$. For each $z \in [-1,1]$, we denote by $\gamma_z\colon S^1 \to S^2$ the parallel at latitude z, parametrized as

$$\gamma_z(t) = \left(\sqrt{1-z^2}\cos(2\pi t), \sqrt{1-z^2}\sin(2\pi t), z\right).$$

For each continuous map $u\colon [-1,1] \to \Lambda M$ such that $E(u(0)) = E(u(1)) = 0$, there exists a unique continuous map $\tilde u\colon S^2 \to M$ such that $u(z) = \tilde u \circ \gamma_z$ for each $z \in [-1,1]$. We denote by $\mathcal{U}$ the space of such maps u whose associated $\tilde u$ has degree 1. The Birkhoff min–max value

$$\mathrm{bir}(M,g) = \inf_{u\in\mathcal{U}} \max_{z\in[-1,1]} E(u(z))^{1/2}$$

is the length of some closed geodesic of (M,g).

Lemma A.2. *On every positively curved closed Riemannian two-sphere (M,g), we have*

$$\mathrm{bir}(M,g) = \mathrm{sys}(M,g).$$

Proof. Let $\gamma\colon S^1 \to M$ be a shortest closed geodesic of (M,g) parametrized with constant speed, so that $E(\gamma) = L(\gamma)^2 = \mathrm{sys}(M,g)^2$. A theorem of Calabi–Cao [11] implies that γ is simple, that is, an embedding $\gamma\colon S^1 \hookrightarrow M$. We fix an orientation on M, and consider the corresponding complex structure of (M,g). Namely, for every non-zero $v \in T_xM$, the tangent vector $Jv \in T_xM$ is obtained by rotating v in the positive direction of an angle $\pi/2$. We consider the vector field $\nu(t) = J\dot\gamma(t)$ orthogonal to $\dot\gamma(t)$. Notice that ν is a parallel vector field, since the complex structure J is parallel. If K_g denotes the Gaussian curvature of (M,g), we have

$$\mathrm{d}^2E(\gamma)[\nu,\nu] = \int_{S^1} \left(\|\nabla_t\nu\|_g^2 - K_g\|\dot\gamma\|_g^2\|\nu\|_g^2\right)\mathrm{d}t = -\int_{S^1} K_g\|\dot\gamma\|_g^4\mathrm{d}t < 0. \tag{A.1}$$

We now consider Morse's finite-dimensional approximation of the free loop space (see, e.g., [45]). We fix a positive integer k that is large enough so that $d(\zeta(t_0),\zeta(t_1)) < \mathrm{injrad}(M,g)$ for all $\zeta \in \Lambda M$ with $E(\zeta) \le E(\gamma) = \mathrm{sys}(M,g)^2$ and for all $t_0,t_1 \in \mathbb{R}$ with $|t_1 - t_0| < 1/k$. Here, d denotes the Riemannian distance on (M,g). We consider the open finite-dimensional manifold

$$\Lambda_k M = \left\{\boldsymbol{x} = (x_0,\dots,x_{k-1}) \in M\times\cdots\times M \;\middle|\; d(x_i,x_{i+1}) < \mathrm{injrad}(M,g) \;\; \forall i \in \mathbb{Z}_k\right\}.$$

Such a manifold admits an embedding

$$\iota\colon \Lambda_k M \hookrightarrow \Lambda M, \quad \iota(\boldsymbol{x}) = \gamma_{\boldsymbol{x}},$$

where each restriction $\gamma_{\boldsymbol{x}}|_{[i/k,(i+1)/k]}$ is the shortest geodesic parametrized with constant speed joining x_i and x_{i+1}. We denote the restricted energy functional by

$$E_k = E \circ \iota \colon \Lambda_k M \to [0,\infty), \quad E_k(\boldsymbol{x}) = k \sum_{i\in\mathbb{Z}_k} d(x_i, x_{i+1})^2.$$

Let $\boldsymbol{x} := \iota^{-1}(\gamma)$. We consider the tangent vector $\boldsymbol{v} := (v_0, \dots, v_{k-1}) \in T_{\boldsymbol{x}}(\Lambda_k M)$ such that $v_i = \nu(i/k)$ for all $i \in \mathbb{Z}_k$. Inequality (A.1) readily implies that $\mathrm{d}\iota(\boldsymbol{x})\boldsymbol{v}$ lies in the negative cone of the Hessian $\mathrm{d}^2E(\gamma)$, since

$$\begin{aligned} \mathrm{d}^2E_k(\boldsymbol{x})[\boldsymbol{v},\boldsymbol{v}] &= \tfrac{\mathrm{d}^2}{\mathrm{d}z^2}\big|_{z=0} E(\iota(\exp_{\boldsymbol{x}}(z\boldsymbol{v})) \\ &\le \tfrac{\mathrm{d}^2}{\mathrm{d}z^2}\big|_{z=0} E(\exp_{\gamma(\cdot)}(z\nu(\cdot))) \\ &= \mathrm{d}^2E(\gamma)[\nu,\nu] \\ &< 0. \end{aligned} \tag{A.2}$$

Here, the exponential map in $\Lambda_k M$ is the one associated with the natural Riemannian metric $g \oplus \cdots \oplus g$.

The complement $M\backslash\gamma$ has two connected components B_+ and B_-, each one diffeomorphic to a two-ball. The vector field ν points into one of them, say B_+. We define the continuous map

$$w\colon [-1/3, 1/3] \to \Lambda_k M, \quad w(z) = \exp_{\boldsymbol{x}}(z\epsilon\boldsymbol{v}).$$

Notice that $w(0) = \boldsymbol{x}$. We fix $\epsilon > 0$ small enough so that, for all $z \in (0, 1/3]$, the loop $\iota(w(\pm z))$ is entirely contained in the open ball $B_\pm$, and by Eq. (A.2), we have

$$E_k(w(z)) < E_k(w(0)) = \mathrm{sys}(M,g)^2, \quad \forall z \in [-1/3,1/3]\backslash\{0\}.$$

We now consider the open subspaces $U_+, U_- \subset \Lambda_k M$ given by

$$U_\pm = \Lambda_k M \cap (B_\pm \times \cdots \times B_\pm).$$

We have $w(\pm 1/3) \in U_\pm$. The flow ϕ_s of the anti-gradient $-\nabla E_k$ is complete in positive time s in the sublevel set $E_k^{-1}([0, \mathrm{sys}(M,g)^2])$. We claim that

$$\phi_s(w(\pm 1/3)) \in U_\pm, \quad \forall s \ge 0.$$

Indeed, assume by contradiction that there exists $s_0 > 0$ such that $\phi_{s_0}(w(\pm 1/3)) \in \partial U_\pm$, and take s_0 to be the minimal such time. If $\boldsymbol{y} := \phi_{s_0}(w(\pm 1/3))$, the components of the anti-gradient vector $\boldsymbol{z} := -\nabla E_k(\boldsymbol{y})$ are given by

$$z_i = 2(\dot\gamma_{\boldsymbol{y}}(\tfrac{i}{k}^+) - \dot\gamma_{\boldsymbol{y}}(\tfrac{i}{k}^-)), \quad \forall i \in \mathbb{Z}_k.$$

Since $\boldsymbol{y} \in \partial U_\pm$, at least one of its components y_i must belong to $\partial B_\pm$. Assume that all the y_i's belong to $\partial B_\pm$, and therefore, they are of the form $y_i = \gamma(t_i)$ for some $t_i \in S^1$. In this case, we have $z_i = \lambda_i\dot\gamma(t_i)$ for some $\lambda_i \in \mathbb{R}$; but this is impossible, since it would imply that all the components of $\phi_s(w(\pm 1/3))$ belong to $\partial B_\pm$ for all $s \in \mathbb{R}$, and thus that $\phi_s(w(\pm 1/3))$ belong to $\partial U_\pm$ for all $s \in \mathbb{R}$. Therefore, at least one component $y_i \in \partial B_\pm$ is adjacent to a component in the interior $y_{i-1} \in B_\pm$. However, this implies that the vector

z_i points inside $B_\pm$, and therefore, $\phi_{s_0-\delta}(w(\pm 1/3)) \notin U_\pm$ for all $\delta > 0$ small enough, contradicting the minimality of s_0.

We set $\delta := \min\{\mathrm{injrad}(M,g), \mathrm{sys}(M)/(4k)\}$. Since $E_k(\phi_s(w(\pm 1/3))) < \mathrm{sys}(M,g)^2$ for all $s \geq 0$, and since $\mathrm{sys}(M,g)^2$ is the smallest positive critical value of E_k, we can fix a large enough $s > 0$ such that $E_k(\phi_s(w(\pm 1/3))) < \delta^2$. We extend w to a map $w\colon [-2/3, 2/3] \to \Lambda_k M$ by setting

$$w(\pm z) = \phi_{(3z-1)s}(w(\pm 1/3)), \quad \forall z \in [1/3, 2/3].$$

Notice that $w(\pm z) \in U_\pm$ for all $z \in (0, 2/3]$, and $E_k(w(\pm 2/3)) < \delta^2$. We set

$$\boldsymbol{y}^\pm = (y_0^\pm, \dots, y_{k-1}^\pm) := w(\pm 2/3).$$

For each $r \in [0,1]$, we define $\boldsymbol{y}^\pm(r) = (y_0^\pm(r), \dots, y_{k-1}^\pm(r))$ by

$$y_i^\pm(r) := \exp_{y_0^\pm}((1-r)\exp_{y_0^\pm}^{-1}(y_i^\pm)).$$

Notice that $\boldsymbol{y}^\pm(0) = \boldsymbol{y}^\pm$, $\boldsymbol{y}^\pm(r) \in U_\pm$, and

$$E_k(\boldsymbol{y}^\pm(r)) = k \sum_{i \in \mathbb{Z}_k} d(y_i^\pm(r), y_{i+1}^\pm(r))^2 < 4k^2\delta^2 \leq \mathrm{sys}(M,g)^2, \quad \forall r \in [0,1],$$
$$E_k(\boldsymbol{y}^\pm(1)) = 0.$$

We extend w to a continuous map $w\colon [-1,1] \to \Lambda_k M$ by setting

$$w(\pm z) = \boldsymbol{y}^\pm(3z-2), \quad \forall z \in [2/3, 1].$$

Finally, we define $u := \iota \circ w\colon [-1,1] \to \Lambda M$. Notice that the associated continuous map $\tilde{u}\colon S^2 \to M$ has degree 1; indeed, the preimage $u^{-1}(\gamma(t))$ is a singleton for every $t \in S^1$, and the restriction of u to a neighborhood of $u^{-1}(\gamma)$ is a homeomorphism onto its image. Therefore, $u \in \mathcal{U}$, and

$$\mathrm{bir}(M,g) \leq \max_{z \in [-1,1]} E(u(z))^{1/2} = E(u(0))^{1/2} = \mathrm{sys}(M,g).$$

On the other hand, $\mathrm{bir}(M,g)^2$ is a positive critical value of E, and therefore,

$$\mathrm{bir}(M,g) \geq \mathrm{sys}(M,g).$$

□

Proof of Theorem A.1. We set $M_i := \partial B_i$, $i = 1, 2$. Since the regions $B_1 \subset B_2$ are strictly convex, for each $x \in M_2$, there exists a unique $\pi(x) \in M_1$ such that

$$\|x - \pi(x)\| = \min_{y \in M_1} \|x - y\|.$$

The map $\pi\colon M_2 \to M_1$ is a 1-Lipschitz homeomorphism with respect to the Riemannian metrics g_i on M_i that are restriction of the ambient Euclidean metric. In particular, for every $W^{1,2}$ curve $\gamma_2\colon S^1 \to M_2$, if we denote by $\gamma_1\colon = \pi \circ \gamma_2$ its image in M_1, we have

$$\int_{S_1} \|\dot\gamma_2(t)\|^2 \mathrm{d}t \geq \int_{S_1} \|\dot\gamma_1(t)\|^2 \mathrm{d}t$$

We denote by $\mathcal{U}_1$ and $\mathcal{U}_2$ the family of maps involved in the definition of the Birkhoff min–max values of M_1 and M_2 respectively. Notice that $\pi \circ u \in \mathcal{U}_1$

for all $u \in \mathcal{U}_2$. Therefore, if we denote the energy of $W^{1,2}$ loops $\gamma\colon S^1 \to \mathbb{R}^3$ by

$$E(\gamma) = \int_{S^1} \|\dot\gamma(t)\|^2 \,\mathrm{d}t,$$

we have

$$\mathrm{bir}(M_2) = \inf_{u\in\mathcal{U}_2} \max_{z\in[-1,1]} E(u(z))^{1/2} \geq \inf_{u\in\mathcal{U}_2} \max_{z\in[-1,1]} E(\pi\circ u(z))^{1/2} \geq \mathrm{bir}(M_1).$$

This, together with Lemma A.2, implies that $\mathrm{sys}(M_2) \geq \mathrm{sys}(M_1)$. $\square$

References

[1] Abouzaid, M.: Symplectic cohomology and Viterbo's theorem. In: Latschev, J., Oancea, A. (eds.) Free Loop Spaces in Geometry and Topology. IRMA Lectures in Mathematics and Theoretical Physics, vol. 24, pp. 271–485. European Mathematical Society, Zürich (2015)

[2] Albers, P., Frauenfelder, U., Oancea, A.: Local systems on the free loop space and finiteness of the Hofer–Zehnder capacity. Math. Ann. **367**(3–4), 1403–1428 (2017)

[3] Abreu, M., Gutt, J., Kang, J., Macarini, L.: Two closed orbits for non-degenerate Reeb flows. Math. Proc. Camb. Philos. Soc. **170**(3), 625–660 (2021)

[4] Arnol'd, V.I.: The first steps of symplectic topology. Uspekhi Mat. Nauk **41**(6)(252), 3–18 (1986)

[5] Abbondandolo, A., Schwarz, M.: On the Floer homology of cotangent bundles. Commun. Pure Appl. Math. **59**(2), 254–316 (2006)

[6] Abbondandolo, A., Schwarz, M.: Corrigendum: On the Floer homology of cotangent bundles. Commun. Pure Appl. Math. **67**(4), 670–691 (2014)

[7] Abbondandolo, A., Schwarz, M.: The role of the Legendre transform in the study of the Floer complex of cotangent bundles. Commun. Pure Appl. Math. **68**(11), 1885–1945 (2015)

[8] Asselle, L., Starostka, M.: The Palais–Smale condition for the Hamiltonian action on a mixed regularity space of loops in cotangent bundles and applications. Calc. Var. Partial Differ. Equ. **59**(4), Paper No. 113, 28 (2020)

[9] Borman, M.S., McLean, M.: Bounding Lagrangian widths via geodesic paths. Compos. Math. **150**(12), 2143–2183 (2014)

[10] Biran, P., Polterovich, L., Salamon, D.: Propagation in Hamiltonian dynamics and relative symplectic homology. Duke Math. J. **119**(1), 65–118 (2003)

[11] Calabi, E., Cao, J.G.: Simple closed geodesics on convex surfaces. J. Differ. Geom. **36**(3), 517–549 (1992)

[12] Cieliebak, K., Floer, A., Hofer, H.: Symplectic homology. II. A general construction. Math. Z **218**(1), 103–122 (1995)

[13] Cieliebak, K., Ginzburg, V.L., Kerman, E.: Symplectic homology and periodic orbits near symplectic submanifolds. Comment. Math. Helv. **79**(3), 554–581 (2004)

[14] Cieliebak, K. Hasselblatt, B., Hofer, H., Latschev, J., Schlenk, F.: Quantitative symplectic geometry. In: Hasselblatt, B. (ed.) Dynamics, Ergodic Theory, and Geometry. Mathematical Sciences Research Institute Publications, vol. 54, pp 1–44. Cambridge University Press, Cambridge (2007)

[15] Frauenfelder, U., Ginzburg, V.L., Schlenk, F.: Energy capacity inequalities via an action selector. In: Entov, M., Pinchover, Y., Sageev, M. (eds.) Geometry, Spectral Theory, Groups, and Dynamics. Contemporary Mathematics, vol. 387, pp 129–152. American Mathematical Society, Providence (2005)

[16] Floer, A., Hofer, H.: Symplectic homology. I. Open sets in $\mathbf{C}^n$. Math. Z. **215**(1), 37–88 (1994)

[17] Floer, A.: Morse theory for Lagrangian intersections. J. Differ. Geom. **28**(3), 513–547 (1988)

[18] Floer, A.: Symplectic fixed points and holomorphic spheres. Commun. Math. Phys. **120**(4), 575–611 (1989)

[19] Fukaya, K., Ono, K.: Arnold conjecture and Gromov–Witten invariant. Topology **38**(5), 933–1048 (1999)

[20] Frauenfelder, U., Pajitnov, A.: Finiteness of π_1-sensitive Hofer–Zehnder capacity and equivariant loop space homology. J. Fixed Point Theory Appl. **19**(1), 3–15 (2017)

[21] Frauenfelder, U., Schlenk, F.: Hamiltonian dynamics on convex symplectic manifolds. Isr. J. Math. **159**, 1–56 (2007)

[22] Ginzburg, V.L., Gürel, B.Z.: Relative Hofer–Zehnder capacity and periodic orbits in twisted cotangent bundles. Duke Math. J. **123**(1), 1–47 (2004)

[23] Ginzburg, V.L., Gürel, B.Z.: The Conley conjecture and beyond. Arnold Math. J. **1**(3), 299–337 (2015)

[24] Ginzburg, V.L.: Marsden, J.E., Ratiu, T. S.: The Weinstein conjecture and theorems of nearby and almost existence. In: Marsden, J.E., Ratiu, T. S. (eds.) The Breadth of Symplectic and Poisson Geometry. Progress in Mathematics, vol. 232, pp. 139–172. Birkhäuser Boston, Boston (2005)

[25] Ginzburg, V.L.: The Conley conjecture. Ann. Math. (2) **172**(2), 1127–1180 (2010)

[26] Gatien, D., Lalonde, F.: Holomorphic cylinders with Lagrangian boundaries and Hamiltonian dynamics. Duke Math. J. **102**(3), 485–511 (2000)

[27] Gromov, M.: Pseudo holomorphic curves in symplectic manifolds. Invent. Math. **82**(2), 307–347 (1985)

[28] Ginzburg, V.L., Shon, J.: On the filtered symplectic homology of prequantization bundles. Int. J. Math. **29**(11), 1850071, (2018)

[29] Gürel, B.Z.: Totally non-coisotropic displacement and its applications to Hamiltonian dynamics. Commun. Contemp. Math. **10**(6), 1103–1128 (2008)

[30] Gong, W., Xue, J.: Floer homology in the cotangent bundle of a closed Finsler manifold and noncontractible periodic orbits. Nonlinearity **33**(12), 6297–6348 (2020)

[31] Hermann, D.: Holomorphic curves and Hamiltonian systems in an open set with restricted contact-type boundary. Duke Math. J. **103**(2), 335–374 (2000)

[32] Hofer, H., Salamon, D.A.: Floer homology and Novikov rings. In: Hofer, H., Taubes, C.H., Weinstein, A., Zehnder, E. (eds.) The Floer Memorial Volume. Progress in Mathematics, vol. 133, pp. 483–524. Birkhäuser, Basel (1995)

[33] Hofer, H., Viterbo, C.: The Weinstein conjecture in cotangent bundles and related results. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **15**(3), 411–445 (1988)

[34] Hofer, H., Viterbo, C.: The Weinstein conjecture in the presence of holomorphic spheres. Commun. Pure Appl. Math. **45**(5), 583–622 (1992)

[35] Hofer, H., Zehnder, E.: Periodic solutions on hypersurfaces and a result by C. Viterbo. Invent. Math. **90**, 1–9 (1987)

[36] Hofer, H., Zehnder, E.: Symplectic Invariants and Hamiltonian Dynamics. Modern Birkhäuser Classics. Birkhäuser Verlag, Basel (2011). Reprint of the 1994 edition

[37] Irie, K.: Hofer–Zehnder capacity of unit disk cotangent bundles and the loop product. J. Eur. Math. Soc. **16**(11), 2477–2497 (2014)

[38] Kang, J.: Symplectic homology of displaceable Liouville domains and leafwise intersection points. Geom. Dedicata **170**, 135–142 (2014)

[39] Lê, H.V., Ono, K.: Cup-length estimates for symplectic fixed points. In: Contact and Symplectic Geometry (Cambridge, 1994). Publications of the Newton Institute, vol. 8, pp. 268–295. Cambridge University Press, Cambridge (1996)

[40] Liu, G., Tian, G.: Floer homology and Arnold conjecture. J. Differ. Geom. **49**(1), 1–74 (1998)

[41] Liu, G., Tian, G.: Weinstein conjecture and GW-invariants. Commun. Contemp. Math. **2**(4), 405–459 (2000)

[42] Lu, G.: The Weinstein conjecture on some symplectic manifolds containing the holomorphic spheres. Kyushu J. Math. **52**(2), 331–351 (1998)

[43] Lu, G.: Gromov–Witten invariants and pseudo symplectic capacities. Isr. J. Math. **156**, 1–63 (2006)

[44] Macarini, L.: Hofer–Zehnder capacity and Hamiltonian circle actions. Commun. Contemp. Math. **6**(6), 913–945 (2004)

[45] Milnor, J.: Morse Theory. Based on Lecture Notes by M. Spivak and R. Wells. Annals of Mathematics Studies, vol. 51. Princeton University Press, Princeton (1963)

[46] Macarini, L., Schlenk, F.: A refinement of the Hofer–Zehnder theorem on the existence of closed characteristics near a hypersurface. Bull. Lond. Math. Soc. **37**(2), 297–300 (2005)

[47] Oh, Y.-G.: Lectures on Floer theory and spectral invariants of Hamiltonian flows. In: Biran, P., Cornea, O., Lalonde, F. (eds.) Morse Theoretic Methods in Nonlinear Analysis and in Symplectic Topology. NATO Science Series II: Mathematics, Physics and Chemistry, vol. 217, pp. 321–416. Springer, Dordrecht (2006)

[48] Ono, K.: On the Arnol'd conjecture for weakly monotone symplectic manifolds. Invent. Math. **119**(3), 519–537 (1995)

[49] Rabinowitz, P.H.: Periodic solutions of Hamiltonian systems. Commun. Pure Appl. Math. **31**(2), 157–184 (1978)

[50] Ritter, A.F.: Novikov-symplectic cohomology and exact Lagrangian embeddings. Geom. Topol. **13**(2), 943–978 (2009)

[51] Rot, T.O.: Answer to the MathOverflow question "Relative homology of free loop space with respect to constant loops". mathoverflow.net/questions/373405 (2020)

[52] Salamon, D.: Lectures on Floer Homology. Symplectic Geometry and Topology (Park City, UT, 1997), IAS/Park City Mathematics Series, vol. 7, pp. 143–229. American Mathematical Society, Providence (1999)

[53] Schwarz, M.: On the action spectrum for closed symplectically aspherical manifolds. Pac. J. Math. **193**(2), 419–461 (2000)

[54] Schlenk, F.: Applications of Hofer's geometry to Hamiltonian dynamics. Comment. Math. Helv. **81**(1), 105–121 (2006)

[55] Struwe, M.: Existence of periodic solutions of Hamiltonian systems on almost every energy surface. Bol. Soc. Brasil. Mat. (N.S.) **20**(2), 49–58 (1990)

[56] Sugimoto, Y.: Hofer's metric on the space of Lagrangian submanifolds and wrapped Floer homology. J. Fixed Point Theory Appl. **18**(3), 547–567 (2016)

[57] Sugimoto, Y.: The sharp energy-capacity inequality on convex symplectic manifolds. J. Fixed Point Theory Appl. **21**(1), Paper No. 13, 18 (2019)

[58] Sullivan, D.: Differential forms and the topology of manifolds. In: Hattori, A. (ed.) Manifolds—Tokyo 1973 (Proceedings International Conference, Tokyo, 1973), pp. 37–49. University of Tokyo Press, Published for the Mathematical Society of Japan, (1975)

[59] Salamon, D.A., Weber, J.: Floer homology and the heat flow. Geom. Funct. Anal. **16**(5), 1050–1138 (2006)

[60] Usher, M.: Deformed Hamiltonian Floer theory, capacity estimates and Calabi quasimorphisms. Geom. Topol. **15**(3), 1313–1417 (2011)

[61] Usher, M.: Many closed symplectic manifolds have infinite Hofer–Zehnder capacity. Trans. Am. Math. Soc. **364**(11), 5913–5943 (2012)

[62] Viterbo, C.: A proof of Weinstein's conjecture in R^{2n}. Ann. Inst. H. Poincaré Anal. Non Linéaire **4**(4), 337–356 (1987)

[63] Viterbo, C.: Symplectic topology as the geometry of generating functions. Math. Ann. **292**(4), 685–710 (1992)

[64] Viterbo, C.: Functors and computations in Floer homology with applications II (1996). arXiv:1805.01316

[65] Viterbo, C.: Exact Lagrange submanifolds, periodic orbits and the cohomology of free loop spaces. J. Differ. Geom. **47**(3), 420–468 (1997)

[66] Viterbo, C.: Functors and computations in Floer homology with applications. I. Geom. Funct. Anal. **9**(5), 985–1033 (1999)

[67] Vigué-Poirrier, M., Sullivan, D.: The homology theory of the closed geodesic problem. J. Differ. Geom. **11**(4), 633–644 (1976)

[68] Weber, J.: Noncontractible periodic orbits in cotangent bundles and Floer homology. Duke Math. J. **133**(3), 527–568 (2006)

[69] Weinstein, A.: Periodic orbits for convex Hamiltonian systems. Ann. Math. (2) **108**, 507–518 (1978)

[70] Weinstein, A.: On the hypotheses of Rabinowitz' periodic orbit theorems. J. Differ. Equ. **33**(3), 353–358 (1979)

[71] Ziller, W.: The free loop space of globally symmetric spaces. Invent. Math. **41**(1), 1–22 (1977)

Gabriele Benedetti
Mathematisches Institut
Ruprecht-Karls-Universität Heidelberg
Im Neuenheimer Feld 205
69120 Heidelberg
Germany
e-mail: gbenedetti@mathi.uni-heidelberg.de

Jungsoo Kang
Department of Mathematical Sciences, Research Institute in Mathematics
Seoul National University
Gwanak-Gu
Seoul 08826
South Korea
e-mail: jungsoo.kang@snu.ac.kr

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:24
https://doi.org/10.1007/s11784-022-00934-z
Published online April 4, 2022

Journal of Fixed Point Theory and Applications

An Arnold-type principle for non-smooth objects

Lev Buhovsky, Vincent Humilière and Sobhan Seyfaddini

Dedicated to Claude Viterbo on the occasion of his 60th birthday.

Abstract. In this article, we study the Arnold conjecture in settings where objects under consideration are no longer smooth but only continuous. The example of a Hamiltonian homeomorphism, on any closed symplectic manifold of dimension greater than 2, having only one fixed point shows that the conjecture does not admit a direct generalization to continuous settings. However, it appears that the following Arnold-type principle continues to hold in C^0 settings: suppose that X is a non-smooth object for which one can define spectral invariants. If the number of spectral invariants associated to X is smaller than the number predicted by the (homological) Arnold conjecture, then the set of fixed/intersection points of X is homologically non-trivial, hence it is infinite. We recently proved that the above principle holds for Hamiltonian homeomorphisms of closed and aspherical symplectic manifolds. In this article, we verify this principle in two new settings: C^0 Lagrangians in cotangent bundles and Hausdorff limits of Legendrians in 1-jet bundles which are isotopic to 0-section. An unexpected consequence of the result on Legendrians is that the classical Arnold conjecture does hold for Hausdorff limits of Legendrians in 1-jet bundles.

Mathematics Subject Classification. 57R58, 53D12, 53D22.

Keywords. Arnold conjecture, C^0 rigidity, C^0 Lagrangian, Legendrian, spectral invariants, symplectic, Hamiltonian homeomorphisms.

Contents

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

1. Introduction and main results

The Arnold conjecture states that a Hamiltonian *diffeomorphism* of a closed and connected symplectic manifold (M, ω) must have at least as many fixed points as the minimal number of critical points of a smooth function on M. The classical Lusternik–Schnirelmann theory shows that this minimal number is always at least the *cup length* of M, a topological invariant of M defined as[1]

$$\begin{aligned}\mathrm{cl}(M) := \max\{k+1 \,:\, \exists\, a_1, \ldots, a_k \in H_*(M),\ \forall i, \deg(a_i) \neq \dim(M) \\ \text{and } a_1 \cap \cdots \cap a_k \neq 0\}.\end{aligned}$$

Therefore, a natural interpretation of the Arnold conjecture, sometimes referred to as the homological Arnold conjecture, is that a Hamiltonian diffeomorphism of (M, ω) must have at least $\mathrm{cl}(M)$ fixed points.[2] Successful efforts at resolving this conjecture were pioneered by Floer [7,8,10] and led to the development of what is now called Floer homology. The original version of the Arnold conjecture has been proven on symplectically aspherical manifolds [9,12,33] while the homological version has been proven on a larger class of manifolds, e.g., $\mathbb{C}P^n$ by Fortune–Weinstein [11], and symplectic manifolds which are negatively monotone by Lê–Ono [22].

The Arnold conjecture admits reformulations for symplectic objects other than Hamiltonian diffeomorphisms: For example, a Lagrangian version of the conjecture states that in a cotangent bundle T^*N, a Lagrangian submanifold which is Hamiltonian isotopic to the zero section must have at least $\mathrm{cl}(N)$ intersection points with the zero-section O_N (see [12,21]). Here is a Legendrian reformulation of this last statement: a Legendrian submanifold in a 1-jet bundle $J^1N = T^*N \times \mathbb{R}$, which is isotopic to the zero section

[1] Here, $\cap$ refers to the intersection product in homology. The cup length can be equivalently defined in terms of the cup product in cohomology.

[2] Note that we do not make any assumptions regarding non-degeneracy of Hamiltonian diffeomorphisms here.

through Legendrians, must have at least $\mathrm{cl}(N)$ intersections with the 0-wall $O_N \times \mathbb{R}$.[3]

The goal of this article is to understand the Arnold conjecture in settings where objects under consideration are no longer smooth but only continuous. Although the Arnold conjecture is true for Hamiltonian homeomorphisms of surfaces [26], we showed in [3] that every closed and connected symplectic manifold of dimension at least 4 admits a Hamiltonian homeomorphism with a single fixed point. Analogously, an example of a continuous Lagrangian submanifold Hamiltonian homeomorphic to the zero section and having a single intersection point with the zero section can be constructed in the cotangent bundle of any closed connected surface, see Proposition 1.2 below.

In spite of these counter-examples, it appears that certain reformulations of the Arnold conjecture do survive in C^0 settings. These reformulations, which involve counting fixed/intersection points and certain "homologically essential" critical values of the action, (*i.e., spectral invariants*), are inspired by the following statement from Lusternik–Schnirelman theory:

Let f be a smooth function on a closed manifold M. If the number of homologically essential critical values of f is smaller than $\mathrm{cl}(M)$, *then the set of critical points of f is homologically non-trivial.*

The above statement can be deduced from Proposition 3.1. Homologically essential critical values, which are usually referred to as *spectral invariants* in the symplectic literature, are defined in Sect. 3.1. A subset $A \subset M$ is homologically non-trivial if for every open neighborhood U of A the map $i_* : H_j(U) \to H_j(M)$, induced by the inclusion $i : U \hookrightarrow M$, is non-trivial for some $j > 0$. Clearly, homologically non-trivial sets are infinite.

The reformulations of the Arnold conjecture which continue to hold in C^0 settings may be summarized as follows:

Principle 1. *Suppose that X is a non-smooth object for which one can define spectral invariants. If the number of spectral invariants associated to X is smaller than the number predicted by the homological Arnold conjecture, then the set of fixed/intersection points of X is homologically non-trivial, hence it is infinite.*

In our recent article [2], we established the above principle for Hamiltonian homeomorphisms of symplectically aspherical manifolds: suppose that (M, ω) is closed, connected, and symplectically aspherical. In Theorem 1.4 of [2] we prove that if ϕ is a Hamiltonian homeomorphism of (M, ω) with fewer spectral invariants than $\mathrm{cl}(M)$, then the set of fixed points of ϕ is homologically non-trivial. A variant of this statement for negative monotone symplectic manifolds and for complex projective spaces has been proven by Y. Kawamoto in [18].

The main results of this article establish Principle 1 in two more contexts: C^0 Lagrangians in cotangent bundles and Hausdorff limits of Legendrians in 1-jet bundles.

[3] Sandon has recently presented a reformulation of the Arnold conjecture for contactomorphisms; see [34,35].

C^0 **Lagrangians:** Consider the cotangent bundle T^*N of a closed manifold N and denote by O_N its zero section. As we will see in Sect. 4, (Lagrangian) spectral invariants can be defined for a C^0 Lagrangian of the form $L = \phi(O_N)$ where ϕ is a compactly supported Hamiltonian homeomorphism of T^*N; this is proven in Theorem 4.1. We call such a C^0 Lagrangian "a C^0 Lagrangian Hamiltonian homeomorphic to the zero section". It is not difficult to see that in this setting our principle translates to the following statement.

Theorem 1.1. *Let ϕ denote a compactly supported Hamiltonian homeomorphism of T^*N and suppose that $L = \phi(O_N)$. If the number of spectral invariants of L is smaller than $\mathrm{cl}(N)$, then $L \cap O_N$ is homologically non-trivial, hence it is infinite.*

It is interesting to remark that, as for Hamiltonian homeomorphisms, the Arnold conjecture breaks down for C^0 Lagrangians; this is the content of the next result.

Proposition 1.2. *Let M be a closed connected surface. Then, there is a Hamiltonian homeomorphism ψ of T^*M such that the C^0-Lagrangian $L = \psi(O_M)$ has only one intersection with the zero-section O_M.*

Note that although we expect a similar statement to hold in higher dimensions, our proof is valid only for M of dimension two. However, the argument we present is relatively simple compared to the construction in [3].

Remark 1.3. Of course, as a consequence of Theorem 1.1, a C^0 submanifold L as in Proposition 1.2 must have at least $\mathrm{cl}(N)$ distinct spectral invariants.

Remark 1.4. It is reasonable to ask if in the above theorem the hypothesis $L = \phi(O_N)$ could be weakened to L being the Hausdorff limit of a sequence L_i, where each L_i is Hamiltonian isotopic to the zero section. This is related to a conjecture of Viterbo; see also Remark 4.4 below.

Hausdorff limits of Legendrians: Let L denote the Hausdorff limit of a sequence of Legendrians which are contact isotopic to the zero section in the 1-jet bundle J^1N. We have not been able to verify whether it is possible to define Legendrian spectral invariants for the Hausdorff limit L. However, as we will now explain, it is still possible to make sense of the action spectrum of L: Let K be a smooth Legendrian submanifold of J^1N which is contact isotopic to the zero section. Then, as we explain in Sect. 3.3, the set $\mathrm{spec}(K) = \pi_{\mathbb{R}}(K \cap (O_N \times \mathbb{R}))$ is the set of critical values of the gfqi associated to K. By analogy, we will define the *spectrum* of any subset $L \subset J^1N$ to be

$$\mathrm{spec}(L) := \pi_{\mathbb{R}}(L \cap (O_N \times \mathbb{R})).$$

Although our next theorem does not establish Principle 1 for Hausdorff limits of Legendrians, it may still be viewed as a natural incarnation of our principle.

Theorem 1.5. *Let L_i be a sequence of Legendrian submanifolds in J^1N which are contact isotopic to the zero-section $O_N \times \{0\}$. Suppose that this sequence has a limit L for the Hausdorff distance, where $L \subset J^1N$ is a compact subset.*

Assume that the cardinality $\mathrm{spec}(L)$ *is strictly less than* $\mathrm{cl}(N)$*. Then, there exists* $\lambda \in \mathrm{spec}(L)$ *such that* $L \cap (O_N \times \{\lambda\})$ *is homologically non-trivial in* $O_N \times \{\lambda\}$*. In particular,* $L \cap (O_N \times \mathbb{R})$ *is infinite.*

Note that we make no assumptions with regards to regularity of L. In fact, we do not even require L to be a C^0 submanifold of J^1N.

Remark 1.6. A careful examination of the proof of Theorem 1.5 reveals that the assumption of Hausdorff convergence of L_i to L can be relaxed to the following: any neighborhood of L contains L_i for i large.

Remark 1.7. In an ongoing project [17], the second author and N. Vichery show that Principle 1 can also be established for singular supports of sheaves (belonging to a certain subcategory of sheaves introduced by Tamarkin). These singular supports can be seen as (singular) generalizations of Legendrian submanifolds.

Organization of the paper

In Sect. 2, we recall some basic notions from symplectic geometry. In Sect. 3, we introduce preliminaries on Lusternik–Schnirelmann theory and spectral invariants.

Section 4 is dedicated to establishing Principle 1 for C^0 Lagrangians Hamiltonian homeomorphic to the zero section. The main technical step for doing so, which is of independent interest, consists of proving that Lagrangian spectral invariants can be defined for such C^0 Lagrangians. This is achieved in Sect. 4.1; see Theorem 4.1 therein. Theorem 1.1 is proven in Sect. 4.2. We prove Proposition 1.2 in Sect. 4.3. Lastly, Theorem 1.5 is proven in Sect. 5.

2. Preliminaries from symplectic geometry

For the remainder of this section (M, ω) will denote a connected symplectic manifold. Recall that a symplectic diffeomorphism is a diffeomorphism $\theta : M \to M$ such that $\theta^*\omega = \omega$. The set of all symplectic diffeomorphisms of M is denoted by $\mathrm{Symp}(M, \omega)$. Hamiltonian diffeomorphisms constitute an important class of examples of symplectic diffeomorphisms. These are defined as follows: A smooth Hamiltonian $H \in C_c^\infty([0,1] \times M)$ gives rise to a time-dependent vector field X_H which is defined via the equation: $\omega(X_H(t), \cdot) = -dH_t$. The Hamiltonian flow of H, denoted by ϕ_H^t, is by definition the flow of X_H. A compactly supported Hamiltonian diffeomorphism is a diffeomorphism which arises as the time-one map of a Hamiltonian flow generated by a compactly supported Hamiltonian. The set of all compactly supported Hamiltonian diffeomorphisms is denoted by $\mathrm{Ham}_c(M, \omega)$; this forms a normal subgroup of $\mathrm{Symp}(M, \omega)$.

2.1. Symplectic and Hamiltonian homeomorphisms

We equip M with a Riemannian distance d. Given two maps $\phi, \psi : M \to M$, we denote

$$d_{C^0}(\phi, \psi) = \max_{x \in M} d(\phi(x), \psi(x)).$$

We will say that a sequence of compactly supported maps $\phi_i : M \to M$, C^0–converges to ϕ, if there is a compact subset of M which contains the supports of all ϕ_i's and if $d_{C^0}(\phi_i, \phi) \to 0$ as $i \to \infty$. Of course, the notion of C^0–convergence does not depend on the choice of the Riemannian metric.

Definition 2.1. A homeomorphism $\theta : M \to M$ is said to be symplectic if it is the C^0–limit of a sequence of symplectic diffeomorphisms. We will denote the set of all symplectic homeomorphisms by $\mathrm{Sympeo}(M, \omega)$.

The Eliashberg–Gromov theorem states that a symplectic homeomorphism which is smooth is itself a symplectic diffeomorphism. We remark that if θ is a symplectic homeomorphism, then so is θ^{-1}. In fact, it is easy to see that $\mathrm{Sympeo}(M, \omega)$ forms a group.

Definition 2.2. A symplectic homeomorphism ϕ is said to be a Hamiltonian homeomorphism if it is the C^0–limit of a sequence of Hamiltonian diffeomorphisms. We will denote the set of all Hamiltonian homeomorphisms by $\overline{\mathrm{Ham}}(M, \omega)$.

It is not difficult to see that $\overline{\mathrm{Ham}}(M, \omega)$ forms a normal subgroup of $\mathrm{Sympeo}(M, \omega)$. It is a long standing open question whether a smooth Hamiltonian homeomorphism, which is isotopic to identity in $\mathrm{Symp}(M, \omega)$, is a Hamiltonian diffeomorphism; this is often referred to as the C^0 Flux conjecture; see [1,20,38].

We should add that alternative definitions for Hamiltonian homeomorphisms do exist within the literature of C^0 symplectic topology. Most notable of these is a definition given by Müller and Oh in [30]. A homeomorphism which is Hamiltonian in the sense of [30] is necessarily Hamiltonian in the sense of Definition 2.2 and thus, the results of this article apply to the homeomorphisms of [30] as well.

2.2. Hofer's distance

We will denote the Hofer norm on $C_c^\infty([0,1] \times M)$ by

$$\|H\| = \int_0^1 \left(\max_{x \in M} H(t, \cdot) - \min_{x \in M} H(t, \cdot) \right) dt.$$

The Hofer distance on $\mathrm{Ham}(M, \omega)$ is defined via

$$d_{\mathrm{Hofer}}(\phi, \psi) = \inf \|H - G\|,$$

where the infimum is taken over all H, G such that $\phi_H^1 = \phi$ and $\phi_G^1 = \psi$. This defines a bi-invariant distance on $\mathrm{Ham}(M, \omega)$.

Given $B \subset M$, we define its *displacement energy* to be

$$e(B) := \inf\{d_{\mathrm{Hofer}}(\phi, \mathrm{Id}) : \phi \in \mathrm{Ham}(M, \omega), \phi(B) \cap B = \emptyset\}.$$

Non-degeneracy of the Hofer distance is a consequence of the fact that $e(B) > 0$ when B is an open set. This fact was proven in [13,19,32].

3. Preliminaries on spectral invariants

We fix a ground field $\mathbb{F}$, e.g., $\mathbb{Z}_2, \mathbb{Q}$, or $\mathbb{C}$. Singular homology, Floer homology and all notions relying on these theories depend on the field $\mathbb{F}$.

3.1. Min–max critical values and Lusternik–Schnirelmann theory

Let $f \in C^\infty(M)$ a smooth function on a closed and connected manifold M. For any $a \in \mathbb{R}$, let $M^a = \{x \in M : f(x) < a\}$. Let $\alpha \in H_*(M)$ be a non-zero singular homology class and define

$$c_{\mathrm{LS}}(\alpha, f) := \inf\{a \in \mathbb{R} : \alpha \in \mathrm{Im}(i_a^*)\},$$

where $i_a^* : H_*(M^a) \to H_*(M)$ is the map induced in homology by the natural inclusion $i_a : M^a \hookrightarrow M$. The number $c_{\mathrm{LS}}(\alpha, f)$ is a critical value of f and such critical values are often referred to as *homologically essential* critical values.

The function $c_{\mathrm{LS}} : H_*(M)\backslash\{0\} \times C^\infty(M) \to \mathbb{R}$ is called a *min–max* critical value selector. In the following proposition, $[M]$ denotes the fundamental class of M and $[pt]$ denotes the class of a point.

Proposition 3.1. *The min–max critical value selector c_{LS} possesses the following properties.*

1. $c_{\mathrm{LS}}(\alpha, f)$ *is a critical value of* f,
2. $c_{\mathrm{LS}}([pt], f) = \min(f) \leq c_{\mathrm{LS}}(\alpha, f) \leq c_{\mathrm{LS}}([M], f) = \max(f)$,
3. $c_{\mathrm{LS}}(\alpha \cap \beta, f) \leq c_{\mathrm{LS}}(\alpha, f)$, *for any* $\beta \in H_*(M)$ *such that* $\alpha \cap \beta \neq 0$,
4. *Suppose that* $\deg(\beta) < \dim(M)$ *and* $c_{\mathrm{LS}}(\alpha \cap \beta, f) = c_{\mathrm{LS}}(\alpha, f)$. *Then, the set of critical points of* f *with critical value* $c_{\mathrm{LS}}(\alpha, f)$ *is homologically non-trivial.*

The above are well-known results from Lusternik–Schnirelmann theory and hence we will not present a proof here. For further details, we refer the reader to [6, 25, 42].

3.2. Spectral invariants for Lagrangians

Let N be a closed manifold. The canonical symplectic structure on the cotangent bundle T^*N is induced by the form $\omega_0 = -d\lambda$ where $\lambda = p\,dq$. We will denote by Lag the space of Lagrangian submanifolds of T^*N which are Hamiltonian isotopic to the zero section, i.e., $\mathrm{Lag} := \{\phi(O_N) : \phi \in \mathrm{Ham}_c(T^*N, \omega_0)\}$.

Consider $\phi \in \mathrm{Ham}_c(T^*N, \omega_0)$ and let $L = \phi(O_N)$. We will briefly explain how one may associate Lagrangian spectral invariants to the Hamiltonian diffeomorphism ϕ. Pick a compactly supported Hamiltonian $H \in C_c^\infty([0,1] \times T^*N)$ such that $\phi = \phi_H^1$. The action functional associated to H is defined by

$$\mathcal{A}_H : \Omega(T^*N) \to \mathbb{R}\,, \quad z \mapsto \int_0^1 H_t(z(t))\,dt - \int z^*\lambda$$

where $\Omega(T^*N) = \{z : [0,1] \to T^*N \,|\, z(0) \in O_N,\ z(1) \in O_N\}$. The critical points of $\mathcal{A}_H$ are the chords of the Hamiltonian vector field X_H which start and end on O_N. Note that such chords are in one-to-one correspondence with $L \cap O_N$. The spectrum of $\mathcal{A}_H$ consists of the critical values of $\mathcal{A}_H$. It is a nowhere dense subset of $\mathbb{R}$ which turns out to depend only on the time–1 map ϕ_H^1, hence we will denote it by $\mathrm{Spec}(L;\phi)$.

At a formal level, Lagrangian Floer homology is the Morse homology of the above action functional and, in this setting, it is canonically isomorphic to the usual singular homology of N. Now, in a manner similar to what was done in the previous section, one can define a mapping

$$\ell : H_*(N)\backslash\{0\} \times \mathrm{Ham}_c(T^*N, \omega_0) \to \mathbb{R}$$

which associates to a homology class $a \in H_*(N) \setminus \{0\}$ a value in $\mathrm{Spec}(L;\phi)$; roughly speaking, the number $\ell(a,\phi)$ is the minimal action value at which the homology class a appears in the Morse homology of $\mathcal{A}_H$.

These numbers are often referred to as the Lagrangian spectral invariants of ϕ. They were first introduced by Viterbo in [42] via generating function techniques. The Floer theoretic approach was carried out by Oh [28]. Lagrangian spectral invariants have many properties some of which are listed below. For a more comprehensive list of their properties, as well as a survey of their construction, we refer the reader to [27]; see for example Theorems 2.11 and 2.17 in [27].

Proposition 3.2. *The map* $\ell : H_*(N)\backslash\{0\} \times \mathrm{Ham}_c(T^*N, \omega_0) \to \mathbb{R}$, *satisfies the following properties:*

1. $\ell(a,\phi) \in \mathrm{Spec}(L;\phi)$,
2. $|\ell(a,\phi_H^1) - \ell(a,\phi_G^1)| \leq \|H - G\|$,
3. $\ell(a \cap b, \phi\psi) \leq \ell(a,\phi) + \ell(b,\psi)$,
4. $\ell([pt],\phi) \leq \ell(a,\phi) \leq \ell([N],\phi)$,
5. $\ell([N],\phi) = -\ell([pt],\phi^{-1})$,
6. *If* $\phi(O_N) = \psi(O_N)$, *then* $\exists\, C \in \mathbb{R}$ *such that* $\ell(a,\phi) = \ell(a,\psi) + C$ *for all* $a \in H_*(N)\backslash\{0\}$,
7. *Suppose that* $f : N \to \mathbb{R}$ *is a smooth function and define the Lagrangian* $L_f := \{(q, \partial_q f(q)) : q \in N\}$. *Denote by* F *any compactly supported Hamiltonian of* T^*N *which coincides with* $\pi^* f = f \circ \pi$ *on a ball bundle* T_R^*N *of* T^*N *containing* L_f. *Then,* $\ell(a, \phi_F^1) = c_{LS}(a,f)$ *for all* $a \in H_*(N)\backslash\{0\}$.
8. *For any other manifold* N', *the spectral invariants on* $T^*(N \times N')$ *satisfy*
$$\ell(a \otimes a', \phi \times \phi') = \ell(a,\phi) + \ell(a',\phi'),$$
for all $\phi \in \mathrm{Ham}_c(T^*N,\omega)$, $\phi' \in \mathrm{Ham}_c(T^*N',\omega)$, $a \in H_*(N)\backslash\{0\}$ *and* $a' \in H_*(N')\backslash\{0\}$.

Note that the sixth property above tells us that spectral invariants $\ell(a,\phi)$ are essentially invariants of the Lagrangian $L := \phi(O_N)$. As a consequence of this property, the set of spectral invariants of L is well defined up to a shift by a constant. In particular, we can make sense of the total number of spectral invariants of any Lagrangian L which is Hamiltonian isotopic to the zero section. Similarly, we see that $\gamma : \mathrm{Lag} \to \mathbb{R}$, defined by

$$\gamma(\phi(O_N)) := \ell([N],\phi) - \ell([pt],\phi) \tag{1}$$

is well defined, i.e., it only depends on the Lagrangian $\phi(O_N)$ and not on ϕ. Viterbo showed in [42] that γ induces a non-degenerate distance on Lag.

Finally, we should mention that Lagrangian spectral invariants have been constructed in settings more general than what is described above by Leclercq [23] and Leclercq–Zapolsky [24].

Hamiltonian Spectral Invariants: To prove that Lagrangian spectral invariants can be defined for C^0 Lagrangians Hamiltonian homeomorphic to the zero section, that is to prove Theorem 4.1 below, we will need to use certain results from the theory of Hamiltonian spectral invariants. Here, we will briefly recall the aspects of this theory which will be needed below. For further details on the construction of these invariants, see [29,36]. The specific result used here, which compares Lagrangian and Hamiltonian spectral invariants, was proven in [27].

Given $\phi \in \mathrm{Ham}_c(T^*N, \omega_0)$ and $a \in H_*(N)\backslash\{0\}$, using Hamiltonian Floer homology, one can define the Hamiltonian spectral invariant $c(a, \phi)$; this is a real number which belongs to the (Hamiltonian) action spectrum of ϕ, i.e., there exists a fixed point of ϕ whose action is the value $c(a, \phi)$. These spectral invariants satisfy a list of properties similar to those listed in Proposition 3.2. We will be needing the following property which is proven in [27]: For any $\phi \in \mathrm{Ham}_c(T^*N, \omega_0)$ and any $a \in H_*(N)\backslash\{0\}$, we have

$$c([pt], \phi) \leq \ell(a, \phi) \leq c([N], \phi). \tag{2}$$

See Proposition 2.14 and item *iv* of Theorem 2.17 in [27].

Similarly to Eq. (1), we define $\gamma : \mathrm{Ham}_c(T^*N, \omega_0) \to \mathbb{R}$ via

$$\gamma(\phi) := c([N], \phi) - c([pt], \phi). \tag{3}$$

Like its Lagrangian cousin, γ induces a non-degenerate distance on $\mathrm{Ham}_c(T^*N, \omega_0)$. We will need the following properties:

1. **Comparison inequality:** As an immediate consequence of Eq. 2, the Lagrangian version of γ is smaller than the Hamiltonian version. More precisely, for any $\phi \in \mathrm{Ham}_c(T^*N, \omega_0)$, we have
$$\gamma(\phi(O_N)) \leq \gamma(\phi). \tag{4}$$
2. **Conjugacy invariance:** For any $\phi \in \mathrm{Ham}_c(T^*N, \omega_0)$ and any symplectic diffeomorphism ψ of T^*N, we have
$$\gamma(\phi) = \gamma(\psi\phi\psi^{-1}). \tag{5}$$
3. **Triangle inequality:** For any $\phi, \psi \in \mathrm{Ham}_c(T^*N, \omega_0)$, we have
$$\gamma(\phi\psi) \leq \gamma(\phi) + \gamma(\psi). \tag{6}$$
4. **Energy-capacity inequality:** Suppose that the support of ϕ can be displaced, then
$$\gamma(\phi) \leq 2e(\mathrm{supp}(\phi)), \tag{7}$$
where $e(\mathrm{supp}(\phi))$ is the displacement energy of $\mathrm{supp}(\phi)$.

3.3. Spectral invariants for Legendrians via generating functions

Once again let N be a closed manifold. The standard contact structure on the 1-jet bundle $J^1N = T^*N \times \mathbb{R}$ is induced by the contact form $\alpha = dz - \lambda$, where z is the coordinate on $\mathbb{R}$. We will denote by Leg the space of Legendrian submanifolds of J^1N which are contact isotopic to the zero section. It was proven by Chaperon [4] and Chekanov [5] that for every $L \in \mathrm{Leg}$ there exists a generating function quadratic at infinity (gfqi) $S : N \times E \to \mathbb{R}$, where E is some auxiliary vector space, such that

$$L = \left\{ \left(q, \frac{\partial S}{\partial q}(q,e), S(q,e) \right) \, : \, \frac{\partial S}{\partial e}(q,e) = 0 \right\}.$$

Observe that critical points of S correspond to the intersection points of L with the zero wall $O_N \times \mathbb{R}$: (q,e) is a critical point of S if and only if $(q, 0, S(q,e))$ is a point on L. Note that one can obtain the critical value of a given critical point of S by simply reading the z–coordinate of the corresponding intersection point of L with the zero wall.

By applying a min–max construction similar to that of Sect. 3.1 to the gfqi S, one can define Legendrian spectral invariants of the Legendrian L:

$$\ell : H_*(N)\backslash\{0\} \times \mathrm{Leg} \to \mathbb{R}.$$

The fact that $\ell(a, L)$ does not depend on the choice of the gfqi S is a consequence of the uniqueness theorem of Théret and Viterbo [41,42]. For further details on the construction, see [43].

We will now state those properties of Legendrian spectral invariants which will be used below.

Proposition 3.3. (See [43]) *The map $\ell : H_*(N)\backslash\{0\} \times \mathrm{Leg} \to \mathbb{R}$, satisfies the following properties:*

1. *$\ell(a, L)$ is a critical value of the corresponding gfqi S,*
2. *The map $\ell(a, \cdot) : \mathrm{Leg} \to \mathbb{R}$ is continuous with respect to the C^∞ topology,*
3. *$\ell(a \cap b, L + L') \leq \ell(a, L) + \ell(b, L')$, for all $L, L' \in \mathrm{Leg}$ such that $L + L' := \{(q, p + p', z + z') : (q, p, z) \in L, (q, p', z') \in L'\}$ is a smooth Legendrian submanifold contact isotopic to the 0-section.*
4. *Suppose that $f : N \to \mathbb{R}$ is a smooth function and define the Legendrian $L_f := \{(q, \partial_q f(q), f(q)) : q \in N\}$. Then, $\ell(a, L_f) = c_{LS}(a, f)$ for all $a \in H_*(N)\backslash\{0\}$.*

Remark 3.4. A proof of item 3 in Proposition 3.3 is based on the following observation: If S, S' are gfqi's for L, L', respectively, then $S \oplus S' : N \times E \times E' \to \mathbb{R}$ defined by $S \oplus S'(q, e, e') := S(q, e) + S'(q, e')$ is a gfqi for the Legendrian $L + L'$.

4. C^0 Lagrangians, proof of Theorem 1.1 and Proposition 1.2

The first two subsections in this section are devoted to the proof of Theorem 1.1. In the third, we prove Proposition 1.2. We begin by giving a precise definition of compactly supported Hamiltonian homeomorphisms of T^*N.

Equip N with a Riemannian metric and denote by $T_r^*N := \{(q,p) \in T^*N : \|p\| < r\}$ the cotangent disc bundle of radius $r > 0$. We define $\mathrm{Ham}_c(T_r^*N, \omega_0)$ to be the set of Hamiltonian diffeomorphisms whose support is contained in T_r^*N. A compactly supported Hamiltonian homeomorphism is a homeomorphism which belongs to the uniform closure of $\mathrm{Ham}_c(T_r^*N, \omega_0)$ for some $r > 0$; we will denote their collection by $\overline{\mathrm{Ham}}_c(T^*N, \omega_0)$.

4.1. Spectral invariants for C^0 Lagrangians

We will now prove that Lagrangian spectral invariants can be defined for C^0 Lagrangians of the form $L = \phi(O_N)$ where $\phi \in \overline{\mathrm{Ham}}_c(T^*N, \omega_0)$. Below is the continuity result which allows us to define spectral invariants for such C^0 Lagrangians.

Theorem 4.1. *Lagrangian spectral invariants satisfy the following two properties:*

1. *For any homology class $a \in H_*(N)\backslash\{0\}$, the map*

$$\ell(a, \cdot) : \mathrm{Ham}_c(T^*N, \omega_0) \to \mathbb{R}$$

is continuous with respect to the C^0 topology on $\mathrm{Ham}_c(T^*N, \omega_0)$ and extends continuously to the closure $\overline{\mathrm{Ham}}_c(T^*N, \omega_0)$.

2. *If $\phi(O_N) = \psi(O_N)$, then $\exists\, C \in \mathbb{R}$ such that $\ell(a, \phi) = \ell(a, \psi) + C$ for all $a \in H_*(N)\backslash\{0\}$ and for any $\phi, \psi \in \overline{\mathrm{Ham}}_c(T^*N, \omega_0)$.*

Note that as a consequence of the second item, we can define the spectral invariants of a C^0 Lagrangian Hamiltonian homeomorphic to the zero section, up to shift. In particular, it makes sense to speak of the number of spectral invariants of such a C^0 Lagrangian.

The first part of the above theorem follows from techniques which have by now become rather standard in C^0 symplectic topology and hence, we will only sketch a proof of this part of the theorem. The second part of the statement, however, is based on a trick which was recently introduced in our article [2] in the course of proving C^0 continuity of spectral invariants for Hamiltonian diffeomorphisms; see Theorem 1.1 therein.

Proof of Theorem 4.1. We begin with the proof of the first statement. We will be needing the following claim.

Claim 4.2. For every $r > 0$, there exist constants $C, \delta > 0$, depending on r, such that for any $\psi \in \mathrm{Ham}_c(T_r^*N, \omega_0)$, if $d_{C^0}(\mathrm{Id}, \psi) \leq \delta$, then $|\ell(a, \psi)| \leq C d_{C^0}(\mathrm{Id}, \psi)$.

Proof of Claim 4.2. As a consequence of Inequality (2), it is sufficient to prove the result for the Hamiltonian spectral invariants. This follows immediately from [37, Theorem 5]. $\square$

Claim 4.2 proves continuity of our map at the identity. Next, we consider $\mathrm{Id} \neq \phi \in \mathrm{Ham}_c(T_r^*N, \omega_0)$. Properties (3), (4) and (5) in Proposition 3.2 yield

$$\ell([pt], \psi) = -\ell([N], \psi^{-1}) \leq \ell(a, \phi\psi) - \ell(a, \phi) \leq \ell([N], \psi).$$

Thus,

$$|\ell(a,\phi\psi)-\ell(a,\phi)| \le \max\{|\ell([N],\psi)|,|\ell([pt],\psi)|\}.$$

Combining this with Claim 4.2, we conclude that for any $\phi,\psi \in \mathrm{Ham}_c(T_r^*N,\omega_0)$

$$d_{C^0}(\mathrm{Id},\psi)\le\delta \implies |\ell(a,\phi\psi)-\ell(a,\phi)|\le C d_{C^0}(\mathrm{Id},\psi).$$

This proves that $\ell(a,\cdot):\mathrm{Ham}_c(T_r^*N,\omega_0)\to\mathbb{R}$ is locally Lipschitz continuous. Hence, it extends continuously to the closure $\overline{\mathrm{Ham}}_c(T_r^*N,\omega_0)$. This finishes the proof of the first statement of the theorem. □

We now turn our attention to the second statement of the theorem. We begin with the following a priori weaker statement.

Theorem 4.3. *Let $\phi\in\overline{\mathrm{Ham}}_c(T^*N,\omega_0)$ be a Hamiltonian homeomorphism. If $\phi(O_N)=O_N$, then there exists a constant C such that $\ell(a,\phi)=C$ for all $a\in H_*(N)\backslash\{0\}$.*

Note that in the case where ϕ is a smooth Hamiltonian diffeomorphism, the above theorem reduces to Property (6) in Proposition 3.2.

Remark 4.4. It can be checked that Theorem 4.3 is a consequence of the following conjecture of Viterbo: If $L_i\subset T^*N$ is a sequence of Lagrangians Hamiltonian isotopic to the zero section, which Hausdorff converges to the zero-section O_N, then $\gamma(L_i)\to 0$. This conjecture has been established in several case by Shelukhin, e.g., $N=S^n,\mathbb{C}P^n,\mathbb{T}^n$ and others; See [39,40].

We will now prove that the second item in Theorem 4.1 follows from the Theorem 4.3. Suppose that $\phi(O_N)=\psi(O_N)$, where $\phi,\psi\in\overline{\mathrm{Ham}}_c(T^*N,\omega_0)$. First, note that, as a consequence of the third item in Proposition 3.2, we have the following inequality:

$$-\ell([N],\phi^{-1}\psi)\le\ell(a,\phi)-\ell(a,\psi)\le\ell([N],\psi^{-1}\phi).$$

Hence, it is sufficient to show that $\ell([N],\psi^{-1}\phi)=-\ell([N],\phi^{-1}\psi)$. Now, by the fifth item of Proposition 3.2, $-\ell([N],\phi^{-1}\psi)=\ell([pt],\psi^{-1}\phi)$ and by Theorem 4.3, we have $\ell([pt],\psi^{-1}\phi)=\ell([N],\psi^{-1}\phi)$.

It remains to prove Theorem 4.3. The proof we present below relies on an idea similar to what was used in the proof of Theorem 1.1 of [2].

Proof of Theorem 4.3. Pick a sequence ϕ_i in $\mathrm{Ham}_c(T_\rho^*N,\omega_0)$ which converges uniformly to ϕ (for some $\rho>0$). By Theorem 4.1.1, it is enough to show that there exists a constant C such that $\ell(a,\phi_i)\to C$ for any $a\in H_*(N)\backslash\{0\}$. Denote $L_i:=\phi_i(O_N)$ and observe that, as a consequence of the fourth property in Proposition 3.2, it is sufficient to show that $\gamma(L_i)$ converges to zero.

As we will now explain, we may assume without loss of generality that ϕ admits a fixed point on the zero-section O_N. Indeed, fix $p\in O_N$. Then, $\phi^{-1}(p)\in O_N$, by assumption. Now, for any two points $x_1,x_2\in O_N$ we can find a Hamiltonian G which vanishes on O_N and such that $\phi_G^1(x_1)=x_2$. Taking $x_1=p$ and $x_2=\phi^{-1}(p)$, we obtain a Hamiltonian G which vanishes on the zero section such that $\phi\circ\phi_G^1(p)=p$. For all i, we have $\gamma(\phi_i\circ\phi_G^1)=$

$\gamma(\phi_i)$, by the sixth item of Proposition 3.2. Thus, we can replace ϕ_i by $\phi_i \circ \phi_G^1$ and ϕ by $\phi \circ \phi_G^1$.

Observe that the Lagrangians L_i converge in Hausdorff topology to the zero section, i.e., for any $\delta > 0$ we have $L_i \subset T^*_\delta N$ for i sufficiently large. We will reduce the theorem to the following lemma which was obtained jointly with R. Leclercq. A variant of this lemma was established in [16]; see Lemma 8 therein.

Given $B \subset N$, we denote $T^*B := \{(q,p) \in T^*N : q \in B\}$ and $O_B := \{(q,0) : q \in B\}$.

Lemma 4.5. *Let L_i denote a sequence of Lagrangians in T^*N which are Hamiltonian isotopic to O_N. Suppose that there exists a ball $B \subset N$ such that $L_i \cap T^*B = O_B$. If the sequence L_i Hausdorff converges to O_N, then $\gamma(L_i) \to 0$.*

Proof. Pick $\phi_i \in \mathrm{Ham}_c(T^*N, \omega_0)$ such that $\phi_i(O_N) = L_i$. We begin with the following observation: Since $L_i \cap T^*B$ is connected, any two points $(q_1, 0)$, $(q_2, 0) \in L_i \cap T^*B$ have the same action. Let C_i denote this value.

For any given $\varepsilon > 0$, pick a smooth function $f : N \to \mathbb{R}$ whose critical points are all contained in B and such that $\max(f) - \min(f) < \varepsilon$. Denote by $\pi : T^*N \to N$ the natural projection and define $F = \beta\, \pi^* f$ where $\beta : T^*N \to [0,1]$ is compactly supported and $\beta = 1$ on $T^*_R N$ where $R \gg 1$.

Note that $\phi^t_F(q,p) = (q, p + t\, df(q))$ for $t \in [0,1]$ and $(q,p) \in T^*_1 N$. Therefore, $\phi^1_F \phi_i(O_N) = L_i + L_f$ where $L_i + L_f := \{(q, p + df(q)) : (q,p) \in L_i\}$. The Hausdorff convergence of the sequence L_i to O_N and the fact that $L_i \cap T^*B = O_B$ combine together to imply that $(L_i + L_f) \cap O_N = \{(q,0) : df(q) = 0\}$ for i large enough.

It is easy to see that the action of $(q,0) \in (L_i + L_f) \cap O_N$ is given by $C_i + f(q)$ where C_i is the constant introduced above. Therefore,

$$\gamma(L_i + L_f) \leq \max(f) - \min(f) < \varepsilon.$$

On the other hand, by the second property from Proposition 3.2, we have $|\gamma(L_i + L_f) - \gamma(L_i)| \leq 2(\max(f) - \min(f)) < 2\varepsilon$. Combining this with the previous inequality we obtain $\gamma(L_i) < 3\varepsilon$ for i large enough which proves the lemma. $\square$

The end of the proof of Theorem 4.3 will consist in reducing to Lemma 4.5. We will assume from now on that N has even dimension. The case where N has odd dimension reduces to the even dimensional case by replacing N with $N \times \mathbb{S}^1$ and all ϕ_i's by $\phi_i \times \mathrm{Id}_{\mathbb{S}^1}$.

We introduce for that the auxiliary maps

$$\begin{aligned} \Phi_i = \phi_i \times \phi_i^{-1} :\ & T^*N \times T^*N \to T^*N \times T^*N, \\ & (x,y) \mapsto (\phi_i(x), \phi_i^{-1}(y)), \end{aligned}$$

where we endow $T^*N \times T^*N$ with the symplectic form $\omega_0 \oplus \omega_0$; observe that this is canonically symplectomorphic to $T^*(N \times N)$ equipped with its canonical symplectic structure.

Denote $\overline{L}_i := \phi_i^{-1}(O_N)$ and note that $\Phi_i(O_{N\times N}) = L_i \times \overline{L}_i$. The map Φ_i is a Hamiltonian diffeomorphism which is not compactly supported. To obtain a compactly supported Hamiltonian diffeomorphism, we cut off the generating Hamiltonian of Φ_i far away from $O_{N\times N}$ and obtain a new Hamiltonian diffeomorphism which we will continue to denote by Φ_i. It is not difficult to see that Φ_i remains unchanged on a large enough neighborhood of the zero section and so $\Phi_i(O_{N\times N})$ continues to be $L_i \times \overline{L}_i$.

Properties 8 and 5 of Proposition 3.2 yield

$$\gamma(L_i \times \overline{L}_i) = \gamma(L_i) + \gamma(\overline{L}_i) = 2\gamma(L_i). \tag{8}$$

Our proof crucially relies on the following lemma.

Lemma 4.6. *Fix $\varepsilon > 0$. We can find a ball $B \subset N$, and $\Psi_i \in \mathrm{Ham}_c(T^*N \times T^*N, \omega_0 \oplus \omega_0)$ such that the following properties hold :*

(i) $\gamma(\Psi_i(O_{N\times N})) < \varepsilon$ *for i sufficiently large,*
(ii) $\Psi_i\Phi_i(O_{N\times N})$ *converges in Hausdorff topology to $O_{N\times N}$,*
(iii) $\Psi_i\Phi_i(O_{N\times N}) \cap T^*(B \times B) = O_{B\times B}$ *for i sufficiently large.*

We now explain why this lemma implies that $\gamma(L_i) \to 0$. Fix $\varepsilon > 0$ and let B and Ψ_i be as provided by Lemma 4.6. Using (8), the triangle inequality and the fifth property in Proposition 3.2, we get

$$\begin{aligned}\gamma(L_i) &= \tfrac{1}{2}\gamma(L_i \times \overline{L}_i) = \tfrac{1}{2}\gamma(\Phi_i(O_{N\times N}))\\ &\le \tfrac{1}{2}\gamma(\Phi_i \circ \Psi_i(O_{N\times N})) + \tfrac{1}{2}\gamma(\Psi_i^{-1}(O_{N\times N}))\\ &< \tfrac{1}{2}\gamma(\Phi_i \circ \Psi_i(O_{N\times N})) + \tfrac{\varepsilon}{2}.\end{aligned}$$

The second and the third items of Lemma 4.6 allow us to apply Lemma 4.5 and conclude that $\gamma(\Phi_i \circ \Psi_i(O_{N\times N})) \to 0$. This implies that $\gamma(L_i) \to 0$. This concludes the proof of Theorem 4.3 assuming Lemma 4.6. □

Proof of Lemma 4.6. Fix $\varepsilon > 0$. Pick a non-empty open ball B_1 in $N \simeq O_N$ containing a fixed point p of ϕ and such that the displacement energy of $U_1 := T_1^*B_1$ in T^*N is less than $\frac{\varepsilon}{4}$. Note that the displacement energy of $U_1 \times U_1$ inside $T^*(N \times N)$ is also less than $\frac{\varepsilon}{4}$.

The following claim asserts the existence of a convenient Hamiltonian diffeomorphism which switches coordinates on a small open set.

Claim 4.7. There exist an open ball $B_2 \subset B_1$ containing the fixed point p, $0 < r_2 < 1$ and a Hamiltonian diffeomorphism f of $T^*N \times T^*N$ such that:

- $f(O_{N\times N}) = O_{N\times N}$,
- f is the time-1 map of a Hamiltonian supported in $U_1 \times U_1$,
- for all $(x, y) \in U_2 \times U_2$, we have $f(x, y) = (y, x)$, where $U_2 := T_{r_2}^*B_2$.

Proof. Since N is assumed even dimensional, there is an identity isotopy, say φ_t, of $N \times N$ which is supported in $B_1 \times B_1$ with the following property: there exists a ball $B_2 \subset B_1$ containing p such that $\varphi_1(q_1, q_2) = (q_2, q_1)$ on $B_2 \times B_2$.

Let $\tilde{\varphi}_t$ denote the canonical lift of this isotopy to $T^*N \times T^*N$. The isotopy $\tilde{\varphi}_t$ is symplectic, it preserves $O_{N\times N}$, it is supported in $T^*B_1 \times T^*B_1$, and it can be checked that $\tilde{\varphi}_1(x,y) = (y,x)$ on $T^*B_2 \times T^*B_2$. Furthermore, the isotopy is Hamiltonian. Let H denote a generating Hamiltonian of the isotopy which is supported in $T^*B_1 \times T^*B_1$.

To construct our desired Hamiltonian diffeomorphism f, we simply replace H by βH where β is a smooth cutoff function on $T^*(N \times N)$ such that $\beta = 1$ on $T^*_{1-\delta}(N \times N)$, where δ is a small positive number, and $\beta = 0$ outside $T^*_1(N \times N)$. We set f to be the time-1 map of the Hamiltonian flow of βH and leave it to the reader to check that it satisfies the requirements of the claim. $\square$

We can now complete the proof of Lemma 4.6. Since $p \in B_2$, there exists a ball $B_3 \subset B_2$ and $0 < r_3 < r_2$ such that $\phi(U_3) \Subset U_2$ (i.e., $\phi(U_3)$ is compactly contained in U_2), where $U_3 := T^*_{r_3} B_3$.

Let $\Upsilon_i = \phi_i \times \mathrm{Id}_{T^*N}$ and let

$$\Psi_i = \Upsilon_i^{-1} \circ f^{-1} \circ \Upsilon_i \circ f.$$

We will first show that $\gamma(\Psi_i(O_{N\times N})) < \varepsilon$. Note that by Eq. (4), we have $\gamma(\Psi_i(O_{N\times N})) \leq \gamma(\Psi_i)$, where $\gamma(\Psi_i)$ is the Hamiltonian spectral invariant γ which was introduced above in Eq. (3). Hence, it is sufficient to show that $\gamma(\Psi_i) < \varepsilon$. The triangle inequality for γ (Eq. (6)) and its conjugacy invariance (Eq. (5)) yield $\gamma(\Psi_i) \leq 2\gamma(f)$. Lastly, $\gamma(f) < \frac{\varepsilon}{2}$ because the displacement energy of its support is smaller than $\frac{\varepsilon}{4}$; see Eq. (7). This implies Property (i) in Lemma 4.6.

Next, we will verify the second property in Lemma 4.6. Define $\Psi := \Upsilon^{-1} \circ f^{-1} \circ \Upsilon \circ f$, where $\Upsilon := \phi \times \mathrm{Id}_{T^*N}$, and let $\Phi := \phi \times \phi^{-1}$. Since f, Υ and Φ preserve $O_{N\times N}$, we conclude that $\Phi \circ \Psi$ also preserves $O_{N\times N}$. Now, there exists a neighborhood of $O_{N\times N}$ where the sequences Ψ_i and Φ_i converge uniformly to Ψ and Φ, respectively. It follows that $\Phi_i \circ \Psi_i(O_{N\times N})$ converges in Hausdorff topology to $O_{N\times N}$.

It remains to verify the third property from the lemma. We will first show that $\Phi_i \circ \Psi_i(x,y) = (x,y)$ for all $(x,y) \in U_3 \times U_3$, when i is large enough. To do so, it is sufficient to check that $\Psi_i(x,y) = (\phi_i^{-1}(x), \phi_i(y))$, which we now do:

$$\begin{aligned}\Psi_i(x,y) &= \Upsilon_i^{-1} \circ f^{-1} \circ \Upsilon_i \circ f(x,y)\\ &= \Upsilon_i^{-1} \circ f^{-1} \circ \Upsilon_i(y,x)\\ &= \Upsilon_i^{-1} \circ f^{-1}(\phi_i(y), x)\\ &= \Upsilon_i^{-1}(x, \phi_i(y))\\ &= (\phi_i^{-1}(x), \phi_i(y)).\end{aligned}$$

The above chain of identities is an immediate consequence of the following observations: $f(x,y) = (y,x)$ on $U_2 \times U_2$, $U_3 \times U_3 \subset U_2 \times U_2$, and $\Upsilon_i(U_3 \times U_3) \subset U_2 \times U_2$ for i large enough; the last statement is a consequence of the fact that $\phi(U_3) \Subset U_2$.

Let $B = B_3 \times B_3$ and $r = r_3$, so that $T^*_r B = U_3 \times U_3$. As we have seen, for i large, $\Phi_i \circ \Psi_i$ coincides with the identity on $T^*_r B$. We claim that this implies the third property. Indeed, it clearly implies $O_B \subset \Phi_i \circ \Psi_i(O_{N\times N}) \cap$

T^*B. Furthermore, it also implies that if $\Phi_i \circ \Psi_i(O_{N\times N}) \cap T^*B$ contains a point which is not in O_B, then such a point is in $T^*B\backslash T^*_rB$. But of course this cannot happen for i large because of the Hausdorff convergence of $\Phi_i \circ \Psi_i(O_{N\times N})$ to $O_{N\times N}$. This establishes the third property in Lemma 4.6. $\square$

4.2. Proof of Theorem 1.1

By the assumptions of the theorem, one can find some $r > 0$ and a sequence $\phi_i \in \mathrm{Ham}_c(T^*_rN, \omega_0)$ such that ϕ_i converges uniformly to ϕ. Since the number of Lagrangian spectral invariants of ϕ is assumed to be less than $\mathrm{cl}(N)$, there exist some $\alpha, \beta \in H_*(N)$ with $\deg\alpha, \deg\beta < \dim N$ and $\alpha\cap\beta \neq 0$, such that $\ell(\alpha,\phi) = \ell(\alpha\cap\beta,\phi) =: \lambda$. By the continuity of spectral invariants (*i.e.* the first item of Theorem 4.1), we have $\lim \ell(\alpha,\phi_i) = \lim \ell(\alpha\cap\beta,\phi_i) = \lambda$, when $i\to\infty$.

Let $U \subset O_N$ be any neighborhood of $L\cap O_N$ in O_N. It is enough to show that the closure $\overline{U}$ is homologically non-trivial in O_N. For doing this, pick a smooth function $f : N \to \mathbb{R}$ such that $f = 0$ on $\overline{U}$ and $f < 0$ on $N\backslash\overline{U}$. Denote by $\pi : T^*N \to N$ the natural projection and define $F = \beta\pi^*f$ where $\beta : T^*N \to \mathbb{R}$ is compactly supported and $\beta = 1$ on T^*_RN where R is taken to be large in comparison to r.

Claim 4.8. There exists an integer i_0 such that for any $i \geq i_0$, and for sufficiently small values of $\varepsilon > 0$,

$$\ell(\alpha\cap\beta, \phi^\varepsilon_F\,\phi_i) = \ell(\alpha\cap\beta,\phi_i).$$

Proof. Let $L_i = \phi_i(O_N)$ and $L_{\varepsilon f} = \phi^\varepsilon_F(O_N)$. Note that $\phi^t_F(q,p) = (q, p + t\,df(q))$ for $t\in[0,1]$ and $(q,p)\in T^*_rN$. Therefore, we have $L_{\varepsilon f} = \{(q,\varepsilon df(q)) : q\in N\}$ and $\phi^\varepsilon_F\phi_i(O_N) = \phi^\varepsilon_F(L_i) = L_i + L_{\varepsilon f}$ where $L_i + L_{\varepsilon f} := \{(q, p + \varepsilon df(q)) : (q,p)\in L_i\}$.

Since $L\cap\pi^{-1}(O_N\backslash U)$ is compact and does not intersect O_N, and since the sequence ϕ_i converges uniformly to ϕ, we conclude that for small enough ε and large enough i, $(L_i + L_{\varepsilon f})\cap\pi^{-1}(O_N\backslash U)$ does not intersect O_N as well. On the other hand, since $f = 0$ on U, we get that $(L_i + L_{\varepsilon f})\cap\pi^{-1}(U) = L_i\cap\pi^{-1}(U)$. Therefore, for small enough $\varepsilon > 0$ and large enough i, the Lagrangians L_i and $L_i + L_{\varepsilon f}$ have the same intersection points with the zero-section O_N. Moreover, it is easy to see that for each such intersection point, the two action values corresponding to ϕ_i and $\phi^\varepsilon_F\phi_i$ coincide. Therefore, by fixing i and $\varepsilon > 0$, and considering the family of Lagrangians $L_i + L_{s\varepsilon f}$ when $s\in[0,1]$, we see that the action spectra $\mathrm{Spec}(L_i + L_{s\epsilon f}, \phi^{s\varepsilon}_F\phi_i)$ do not depend on s. In addition, recall that the action spectrum has an empty interior in $\mathbb{R}$. As a result, since the value $\ell(\alpha\cap\beta,\phi^\varepsilon_F\phi_i)$ depends continuously on s, we conclude that it in fact does not depend on $s\in[0,1]$. In particular, $\ell(\alpha\cap\beta,\phi_i) = \ell(\alpha\cap\beta,\phi^\varepsilon_F\phi_i)$. $\square$

The triangle inequality of Proposition 3.2 implies that, for all i, $\ell(\alpha\cap\beta,\phi^\varepsilon_F\phi_i) - \ell(\alpha,\phi_i) \leq \ell(\beta,\phi^\varepsilon_F)$. Using the above claim, for i large and ε small enough, we have $\ell(\alpha\cap\beta,\phi_i) - \ell(\alpha,\phi_i) \leq \ell(\beta,\phi^\varepsilon_F)$. Taking limit as $i\to\infty$, and recalling that $\lim\ell(\alpha,\phi_i) = \lim\ell(\alpha\cap\beta,\phi_i) = \lambda$, we obtain $0 \leq \ell(\beta,\phi^\varepsilon_F)$.

We can now conclude our proof as follows. By Proposition 3.2.7 and Proposition 3.1, we have

$$0 \leq \ell(\beta, \phi_{\varepsilon F}) = c_{LS}(\beta, \varepsilon f) \leq c_{LS}([N], \varepsilon f) = \max(\varepsilon f) = 0.$$

Hence, $c_{LS}(\beta, \varepsilon f) = c_{LS}([N], \varepsilon f) = 0$ and, by Proposition 3.1.4, it follows that the zero level set of f, that is $\overline{U}$, is homologically non-trivial.

4.3. Proof of Proposition 1.2

Let M be a closed surface. The aim of this section is to construct a Hamiltonian homeomorphism ψ of T^*M such that the C^0-Lagrangian $L = \psi(O_M)$ has a single intersection point with the zero section. This will establish Proposition 1.2.

According to [31], there exists a C^1 function $f : M \to \mathbb{R}$, whose set of critical points is an arc γ, i.e., is homeomorphic to $[0,1]$. Let us fix such a function f. Let $F = f \circ \pi$ where π denotes the canonical projection $\pi : T^*M \to M$. The intersection between the C^0-Lagrangian submanifold graph$(df) = \phi_F^1(O_M)$ and the zero section is exactly γ (where we canonically identify O_M with M). Note that such an arc γ must be very irregular. More precisely, it must have infinite length (in particular it cannot be smooth). Indeed, if γ had finite length, then f would have to be constant along γ, hence on the set of its critical points. In particular, this would imply $\max_M f = \min_M f$ and f would be constant over M, contradicting the fact that γ is an arc.

We will construct the C^0-Lagrangian L roughly by "contracting the arc to a point". More precisely, given a point $a \in \gamma$, we will construct a map $h : T^*M \to T^*M$ which is a symplectic diffeomorphism between $T^*M \setminus \gamma$ and $T^*M \setminus \{a\}$, and satisfies $h(\gamma) = a$ and $h(O_M) = O_M$. We will then prove that the map

$$\psi : \begin{cases} x \mapsto h\phi_F^1 h^{-1}(x), & \text{for } x \neq a \\ a \mapsto a. \end{cases}$$

is a Hamiltonian homeomorphism and that $L = \psi(O_M)$ has a unique intersection point with O_M.

Let us now start the construction. A version of the Jordan–Schoenflies theorem (for instance its extension due to Homma [14]) implies that the arc γ admits a basis of neighborhoods $(V_i)_{i \geq 0}$, which are all homeomorphic to open discs and satisfy $\overline{V_{i+1}} \subset V_i$ for all i. Let $(U_i)_{i \geq 1}$ be a decreasing basis of neighborhoods of a. Finally, let $(\delta_i)_{i \geq 0}$ be a decreasing sequence of real numbers converging to 0.

Let $W_0 = V_0$ and $\varepsilon_0 = \delta_0$. Since the V_i's form a basis of disc-like neighborhoods, there exists a smooth (time-dependent) vector field X_1 supported in W_0 whose time-one map ζ_1 sends V_1 into U_1. We may also assume that ζ_1 fixes p. We denote $W_1 = \zeta_1(V_1) \subset U_1$.

The Hamiltonian function $(q,p) \mapsto \langle p, X_1(q) \rangle$ vanishes on O_M and its flow is supported in T^*W_0. By multiplying it with an appropriate cutoff function which equals 1 on a neighborhood of the support of X_1 in T^*M, we obtain a Hamiltonian H_1 supported in $T^*_{\varepsilon_0} W_0$. This Hamiltonian H_1 vanishes on O_M, thus its flow preserves it. Moreover, by construction, the restriction of its flow to the zero section coincides with the flow of X_1. We denote by $h_1 = \phi^1_{H_1}$ its time-one map.

Repeating the above, we construct by induction a sequence of positive real numbers ε_k converging to 0, a decreasing sequence of open subsets (W_k) of M and a sequence of Hamiltonians (H_k) on T^*M such that for each $k \geq 1$, the three following properties hold:

(i) H_k is supported in $T^*_{\varepsilon_{k-1}} W_{k-1}$,
(ii) the time-one map $h_k = \phi^1_{H_k}$ preserves O_M,
(iii) $W_k = h_k \circ \cdots \circ h_1(V_k)$ is included in U_k,
(iv) $T^*_{\varepsilon_k} W_k$ is included in $h_k \circ \cdots \circ h_1(T^*_{\delta_k} V_k)$.

Indeed, assuming all the sequences built up to the order k, we let X_{k+1} be a vector field on M which maps the disc $h_k \circ \cdots \circ h_1(V_{k+1})$ into U_{k+1}. The Hamiltonian H_{k+1} is then obtained by cutting off $(q, p) \mapsto \langle p, X_{k+1}(q)\rangle$ appropriately, as above.

For any $x \in \gamma$, we have $h_k \circ \cdots \circ h_1(x) \subset U_k$ thus the sequence $(h_k \circ \cdots \circ h_1(x))$ converges to p. For any $x \notin \gamma$, we have $x \notin T^*_{\delta_k} V_k$ for k large enough. It follows that for k large enough, $h_k \circ \cdots \circ h_1(x)$ does not belong to $T^*_{\varepsilon_k} W_k$, hence does not belong to the support of any h_i for $i > k$. Thus, the sequence $(h_k \circ \cdots \circ h_1(x))$ stabilizes to a point different from a.

We set $h(x) = \lim_{k\to\infty} f_k(x)$, where $f_k(x) := h_k \circ \cdots \circ h_1(x)$. This limit is uniform. Indeed, given $\varepsilon > 0$, there exists an integer N such that $\mathrm{diam}(T^*_{\varepsilon_k} U_k) < \varepsilon$ for all $k \geq N$. Let $k \geq N$. Then for any $x \in f_k^{-1}(T^*_{\varepsilon_k} W_k)$, we have $f_k(x) \in T^*_{\varepsilon_k} W_k$, hence $f_{k+\ell}(x) \in T^*_{\varepsilon_k} W_k \subset T^*_{\varepsilon_k} U_k$ for any $\ell \geq 1$. Taking limit as ℓ goes to infinity, we obtain $f_k(x), h(x) \in T^*_{\varepsilon_k} U_k$ hence $d(f_k(x), h(x)) < \varepsilon$. Now, for $x \notin f_k^{-1}(T^*_{\varepsilon_k} W_k)$, we have $f_k(x) \notin T^*_{\varepsilon_k} W_k$, hence $f_{k+\ell}(x) = f_k(x)$ for all $\ell \geq 1$. We deduce that $h(x) = f_k(x)$. We have shown that for all x, $d(f_k(x), h(x)) < \varepsilon$, which proves that the limit is uniform.

As a consequence, h is continuous. Moreover the restriction of h induces a symplectic diffeomorphism $T^*M\backslash\gamma \to T^*M\backslash\{a\}$. In addition, note that h preserves the zero section O_M. As announced in the beginning of the proof, we now define

$$\psi : \begin{cases} x \mapsto h\phi^1_F h^{-1}(x), & \text{for } x \neq a, \\ a \mapsto a. \end{cases}$$

Since $\phi^1_F(O_M) \cap O_M = \gamma$, and since $h(O_M) = O_M$, we have $\psi(O_M) \cap O_M = \{a\}$. Finally, ψ is a Hamiltonian homeomorphism because it is the C^0-limit of the Hamiltonian diffeomorphisms

$$(h_k \circ \cdots \circ h_1) \circ \phi^1_F \circ (h_k \circ \cdots \circ h_1)^{-1}$$

as k goes to infinity. $\square$

5. Hausdorff limits of Legendrians and proof of Theorem 1.5

This section is dedicated to the proof of Theorem 1.5. Recall that we consider a sequence L_i of Legendrian submanifolds, contact isotopic to the zero section in $J^1N = T^*N \times \mathbb{R}$, which has a Hausdorff limit L. Denote by

$\pi_{\mathbb{R}} : J^1N = T^*N \times \mathbb{R} \to \mathbb{R}$ the natural projection. Recall that we have defined the spectrum of L by $\mathrm{spec}(L) := \pi_{\mathbb{R}}(L \cap (O_N \times \mathbb{R}))$.

Proof of Theorem 1.5. Observe that the Hausdorff convergence of L_i's to L implies that the set $L_i \cap (O_N \times \mathbb{R})$ is contained in an arbitrarily small neighborhood of $L \cap (O_N \times \mathbb{R})$ for large i. Since $\ell(a, L_i)$ corresponds to an intersection point of L_i with the zero wall, we conclude that the set of limit points of $\{\ell(a, L_i) : a \in H_*(N)\backslash\{0\}, i \in \mathbb{N}\}$ is contained in $\mathrm{spec}(L)$.

Assume that $\mathrm{spec}(L)$ has less than $\mathrm{cl}(N)$ points. It follows from the above discussion that there exist $\alpha, \beta \in H_*(N)\backslash\{0\}$ and $\lambda \in \mathrm{spec}(L)$ such that for a subsequence (i_k) of indices, we have $\ell(\alpha, L_{i_k}) \to \lambda$ and $\ell(\alpha \cap \beta, L_{i_k}) \to \lambda$ as $k \to \infty$. By passing to this subsequence, we may further assume that $\ell(\alpha, L_i) \to \lambda$ and $\ell(\alpha \cap \beta, L_i) \to \lambda$ as $i \to \infty$. Let us show that $L \cap (O_N \times \{\lambda\})$ is homologically non-trivial in $O_N \times \{\lambda\}$.

Pick any neighborhood V of $L \cap (O_N \times \{\lambda\})$ in J^1N. Denote $U := \pi_N(V)$, where $\pi_N : J^1N \to N$ is the natural projection, and pick a smooth function $f : N \to \mathbb{R}$ such that $f = 0$ on $\overline{U}$ and $f < 0$ on $N\backslash\overline{U}$.

Claim 5.1. There exists an integer i_0 such that for any $i \geq i_0$, and for sufficiently small values of $\varepsilon > 0$,

$$\ell(\alpha \cap \beta, L_i + L_{\varepsilon f}) = \ell(\alpha \cap \beta, L_i).$$

Proof. By the Hausdorff convergence of L_i to L, there exists some $\delta > 0$ such that for i large enough and $\varepsilon \geqslant 0$ small enough, we have

$$(L_i + L_{\varepsilon f}) \cap (O_N \times (\lambda - \delta, \lambda + \delta)) \subset V.$$

Furthermore, for any $(q, p, z) \in V$, we have that $q \in U$ and thus $f(q) = 0$ and $df(q) = 0$. This implies that $(L_i + L_{\varepsilon f}) \cap (O_N \times (\lambda - \delta, \lambda + \delta)) = L_i \cap (O_N \times (\lambda - \delta, \lambda + \delta))$, in particular $\mathrm{spec}(L_i + L_{\varepsilon f}) \cap (\lambda - \delta, \lambda + \delta) = \mathrm{spec}(L_i) \cap (\lambda - \delta, \lambda + \delta)$.

The continuity and spectrality properties of spectral invariants, together with the fact that the spectrum of L_i has an empty interior in $\mathbb{R}$ and that $\ell(\alpha \cap \beta, L_i) \in (\lambda - \delta, \lambda + \delta)$ for i large enough, imply that the spectral invariant $\ell(\alpha \cap \beta, L_i + L_{\varepsilon f})$ is independent of ε. □

Now, the triangle inequality of Proposition 3.3 implies that, for all i, $\ell(\alpha \cap \beta, L_i + L_{\varepsilon f}) - \ell(\alpha, L_i) \leq \ell(\beta, L_{\varepsilon f})$. Using the above claim, for i large and ε small enough, we have $\ell(\alpha \cap \beta, L_i) - \ell(\alpha, L_i) \leq \ell(\beta, L_{\varepsilon f})$. Taking limit as $i \to \infty$, and recalling that $\ell(\alpha \cap \beta, L_i), \ell(\alpha, L_i) \to \lambda$, we obtain $0 \leq \ell(\beta, L_{\varepsilon f})$.

We can now conclude our proof as follows. On the one hand, by Proposition 3.3.4, we have $\ell(\beta, L_{\varepsilon f}) = c_{LS}(\beta, \varepsilon f)$. Note that $c_{LS}(\beta, \varepsilon f) = c_{LS}([N] \cap \beta, \varepsilon f)$ and by the above paragraph this number is non-negative. On the other hand, Proposition 3.1.2 gives $c_{LS}([N], \varepsilon f) = 0$. Thus, using Proposition 3.1.3, we obtain the equality $c_{LS}(\beta, \varepsilon f) = c_{LS}([N] \cap \beta, \varepsilon f) = c_{LS}([N], \varepsilon f)$. By Proposition 3.1.4, it follows that the zero level set of f, that is the closure of $U = \pi_N(V)$, is homologically non-trivial in N. Since our choice of a neighborhood V of $L \cap (O_N \times \{\lambda\})$ was arbitrary, we conclude that $L \cap (O_N \times \{\lambda\})$ is homologically non-trivial in $O_N \times \{\lambda\}$. □

Acknowledgements

We dedicate this article to Claude Viterbo whose works in mathematics have deeply influenced ours. Not only that, Claude's constant support and interest in our research, since the very beginnings of our careers, has been a great source of encouragement to each one of us. Lemma 4.5 was proven jointly with Rémi Leclercq. We are grateful to him for generously sharing his ideas with us. We also thank Alberto Abbondandolo for pointing out to us the paper of Pierre Pageault [31] used in the proof of Proposition 1.2. Our proofs of Theorems 1.1 and 1.5 were inspired by the paper of Wyatt Howard [15]. The first author was partially supported by ERC Starting Grant 757585 and ISF Grant 2026/17. The second author was partially supported by the ANR project "Microlocal" ANR-15-CE40-0007. This material is based upon work supported by the National Science Foundation under Grant No. DMS-1440140 while the third author was in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Fall 2018 semester. The third author greatly benefited from the lively research atmosphere of the MSRI and would like to thank the members and staff of the MSRI for their warm hospitality. The third author was partially supported by ERC Starting Grant 851701.

References

[1] Buhovsky, L.: Towards the C^0 flux conjecture. Algebr. Geom. Topol. **14**(6), 3493–3508 (2014)

[2] Buhovsky, L., Humilière, V., Seyfaddini, S.: The action spectrum and C^0 symplectic topology. Math. Ann. **380**(1), 293–316 (2021)

[3] Buhovsky, L., Humilière, V., Seyfaddini, S.: A C^0 counterexample to the Arnold conjecture. Invent. Math. **213**(2), 759–809 (2018)

[4] Chaperon, M.: On generating families. In: The Floer Memorial Volume, Volume 133 of Progr. Math.. Birkhäuser, Basel, pp. 283–296 (1995)

[5] Chekanov, Y.V.: Critical points of quasifunctions, and generating families of Legendrian manifolds. Funktsional. Anal. i Prilozhen. **30**(2), 56–69 (1996)

[6] Cornea, O., Lupton, G., Oprea, J., Tanré, D.: Lusternik–Schnirelmann Category. Mathematical Surveys and Monographs, vol. 103. American Mathematical Society, Providence (2003)

[7] Floer, A.: Proof of the Arnold conjecture for surfaces and generalizations to certain Kähler manifolds. Duke Math. J. **53**(1), 1–32 (1986)

[8] Floer, A.: The unregularized gradient flow of the symplectic action. Commun. Pure Appl. Math. **41**, 775–813 (1988)

[9] Floer, A.: Cuplength estimates on Lagrangian intersections. Commun. Pure Appl. Math. **42**(4), 335–356 (1989)

[10] Floer, A.: Symplectic fixed points and holomorphic spheres. Commun. Math. Phys. **120**(4), 575–611 (1989)

[11] Fortune, B., Weinstein, A.: A symplectic fixed point theorem for complex projective spaces. Bull. Am. Math. Soc. (N.S.) **12**(1), 128–130 (1985)

[12] Hofer, H.: Lusternik-Schnirelman-theory for Lagrangian intersections. Ann. Inst. H. Poincaré Anal. Non Linéaire **5**(5), 465–499 (1988)

[13] Hofer, H.: On the topological properties of symplectic maps. Proc. R. Soc. Edinb. Sect. A **115**(1–2), 25–38 (1990)

[14] Homma, T.: An extension of the Jordan curve theorem. Yokohama Math. J. **1**, 125–129 (1953)

[15] Howard, W.: Action Selectors and the Fixed Point Set of a Hamiltonian Diffeomorphism (2012). ArXiv: 1211.0580

[16] Humilière, V., Leclercq, R., Seyfaddini, S.: Coisotropic rigidity and C^0-symplectic geometry. Duke Math. J. **164**(4), 767–799 (2015)

[17] Humilière, V., Vichery, N.: Cuplength estimates via microlocal sheaf theory. In preparation

[18] Kawamoto, Y.: On C^0-continuity of the spectral norm on non-symplectically aspherical manifolds (2019). arXiv:1905.07809

[19] Lalonde, F., McDuff, D.: The geometry of symplectic energy. Ann. Math. (2) **141**(2), 349–371 (1995)

[20] Lalonde, F., McDuff, D., Polterovich, L.: On the flux conjectures. In: Geometry, Topology, and Dynamics (Montreal, PQ, 1995), Volume 15 of CRM Proc. Lecture Notes. Amer. Math. Soc, Providence, pp. 69–85(1998)

[21] Laudenbach, F., Sikorav, J.-C.: Persistance d'intersection avec la section nulle au cours d'une isotopie hamiltonienne dans un fibré cotangent. Invent. Math. **82**(2), 349–357 (1985)

[22] Lê, H.V., Ono, K.: Cup-length estimates for symplectic fixed points. In: Contact and Symplectic Geometry (Cambridge, 1994), Volume 8 of Publ. Newton Inst.. Cambridge Univ. Press, Cambridge, pp. 268–295 (1996)

[23] Leclercq, R.: Spectral invariants in Lagrangian Floer theory. J. Mod. Dyn. **2**(2), 249–286 (2008)

[24] Leclercq, R., Zapolsky, F.: Spectral invariants for monotone Lagrangians. J. Topol. Anal. **10**(3), 627–700 (2018)

[25] Lusternik, L., Schnirelmann, L.: Méthodes topologiques dans les problèmes variationnels, p. 1930. Issledowatelskij Institut Mathematiki i Mechaniki pri J. M. G. U, Moskau (1930)

[26] Matsumoto, S.: Arnold conjecture for surface homeomorphisms. In: Proceedings of the French-Japanese Conference "Hyperspace Topologies and Applications" (La Bussière, 1997), vol. 104, pp. 191–214 (2000)

[27] Monzner, A., Vichery, N., Zapolsky, F.: Partial quasimorphisms and quasistates on cotangent bundles, and symplectic homogenization. J. Mod. Dyn. **6**(2), 205–249 (2012)

[28] Oh, Y.-G.: Symplectic topology as the geometry of action functional. II. Pants product and cohomological invariants. Commun. Anal. Geom. **7**(1), 1–54 (1999)

[29] Oh, Y.-G.: Construction of spectral invariants of Hamiltonian paths on closed symplectic manifolds. In:The Breadth of Symplectic and Poisson Geometry. Progr. Math. 232. Birkhauser, Boston, pp. 525–570 (2005)

[30] Oh, Y.-G., Müller, S.: The group of Hamiltonian homeomorphisms and C^0-symplectic topology. J. Symplect. Geom. **5**(2), 167–219 (2007)

[31] Pageault, P.: Functions whose set of critical points is an arc. Math. Z. **275**(3–4), 1121–1134 (2013)

[32] Polterovich, L.: Symplectic displacement energy for Lagrangian submanifolds. Ergod. Theory Dyn. Syst. **13**, 357–367 (1993)

[33] Rudyak, Y.B., Oprea, J.: On the Lusternik-Schnirelmann category of symplectic manifolds and the Arnold conjecture. Math. Z. **230**(4), 673–678 (1999)

[34] Sandon, S.: On iterated translated points for contactomorphisms of $\mathbb{R}^{2n+1}$ and $\mathbb{R}^{2n} \times S^1$. Int. J. Math. **23**(2), 1250042 (2012)

[35] Sandon, S.: A Morse estimate for translated points of contactomorphisms of spheres and projective spaces. Geom. Dedicata **165**, 95–110 (2013)

[36] Schwarz, M.: On the action spectrum for closed symplectically aspherical manifolds. Pac. J. Math. **193**(2), 419–461 (2000)

[37] Seyfaddini, S.: Descent and C^0-rigidity of spectral invariants on monotone symplectic manifolds. J. Topol. Anal. **4**(4), 481–498 (2012)

[38] Seyfaddini, S.: A note on C^0 rigidity of Hamiltonian isotopies. J. Symplect. Geom. **11**(3), 489–496 (2013)

[39] Shelukhin, E.: Viterbo conjecture for Zoll symmetric spaces (2018). arXiv:1811.05552

[40] Shelukhin, E.: String topology and a conjecture of Viterbo (2019). arXiv:1904.06798

[41] Théret, D.: A complete proof of Viterbo's uniqueness theorem on generating functions. Topol. Appl. **96**(3), 249–266 (1999)

[42] Viterbo, C.: Symplectic topology as the geometry of generating functions. Math. Ann. **292**, 685–710 (1992)

[43] Zapolsky, F.: Geometry of contactomorphism groups, contact rigidity, and contact dynamics in jet spaces. Int. Math. Res. Not. IMRN **20**, 4687–4711 (2013)

Lev Buhovsky
School of Mathematical Sciences
Tel Aviv University
Tel Aviv
Israel
e-mail: levbuh@tauex.tau.ac.il

Vincent Humilière and Sobhan Seyfaddini
Sorbonne Université and Université Paris Cité, CNRS
IMJ–PRG F-75005
Paris
France
e-mail: vincent.humiliere@imj-prg.fr;
sobhan.seyfaddini@imj-prg.fr

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:38
https://doi.org/10.1007/s11784-022-00947-8
Published online April 15, 2022

Journal of Fixed Point Theory and Applications

Quantitative h-principle in symplectic geometry

Lev Buhovsky and Emmanuel Opshtein

Abstract. We prove a quantitative h-principle statement for subcritical isotropic embeddings. As an application, we construct a symplectic homeomorphism that takes a symplectic disc into an isotropic one in dimension at least 6.

Mathematics Subject Classification. 53D05.

Keywords. Flexibility, Symplectic homeomorphism, H-principle, Subcritical isotropic embedding.

1. Introduction

Gromov's h-principle lies at the core of symplectic topology, by reducing many questions on the existence of embeddings or immersions to verifying their compatibility with algebraic topology. Symplectic topology focuses mainly on the other problems, that do not abide by an h-principle: Lagrangian embeddings, existence of symplectic hypersurfaces in specific homology classes, etc. In [2], we have proved a refined version of h-principle, which in turn yielded applications to $\mathcal{C}^0$-symplectic geometry. For instance, we proved in [2] that in dimension at least 6, $\mathcal{C}^0$-close symplectic 2-discs of the same area are isotopic by a small symplectic isotopy, while in dimension 4, this does no longer hold. A similar quantitative h-principle was also used in [1] to show that the symplectic rigidity manifested in the Arnold conjecture for the number of fixed points of a Hamiltonian diffeomorphism completely disappears for Hamiltonian homeomorphisms in dimension at least 4.

The goal of this note is to prove a quantitative h-principle for *isotropic* embeddings and to derive some flexibility statements on symplectic homeomorphisms.

Theorem 1. (Quantitative h-principle for subcritical isotropic embeddings) *Let V be an open subset of $\mathbb{C}^n$, $k < n$, $u_0, u_1 : D^k \hookrightarrow V$ be isotropic*

Dedicated to Claude Viterbo, on the occasion of his 60th birthday. This article is part of the topical collection "Symplectic geometry-A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

embeddings of closed discs. We assume that there exists a homotopy $F : D^k \times [0,1] \to V$ *between* u_0 *and* u_1 *(so* $F(\cdot,0) = u_0$*,* $F(\cdot,1) = u_1$*) of size less than* ε *(i.e. Diam* $F(\{z\} \times [0,1]) < \varepsilon$ *for all* $z \in D^k$*).*

Then there exists a compactly supported in V *Hamiltonian isotopy* $(\Psi^t)_{t\in[0,1]}$ *of size* 2ε *(i.e. Diam* $\{\Psi^t(z) \,|\, t \in [0,1]\} < 2\varepsilon$ *for every* $z \in V$*), such that* $\Psi^1 \circ u_0 = u_1$*.*

The proof shows that the theorem holds in the relative case, provided u_0, u_1 are symplectically isotopic, relative to the boundary. The method of the proof of theorem 1 follows a very similar track as the quantitative h-principle for symplectic discs that we established in [2]. Paralleling the construction of a symplectic homeomorphism whose restriction to a symplectic 2-disc is a contraction in dimension 6, we can deduce from theorem 1 the following statement:

Theorem 2. *There exists a symplectic homeomorphism with compact support in* $\mathbb{C}^3$ *which takes a symplectic* 2*-disc to an isotropic one.*

Of course, by considering products, we infer that there exists symplectic homeomorphisms that take some codimension 4 symplectic submanifolds to submanifolds which are nowhere symplectic.

The note is organized as follows. We prove theorem 1 in the next section. The construction of a symplectic homeomorphism that takes a symplectic disc to an isotropic one is explained in Sect. 3, where we also explain a relation to relative Eliashberg-Gromov type questions, as posed in [2].

This paper is a result of the work done during visits of the first author at Strasbourg University, and a visit of the second author at Tel Aviv University. We thank both universities and their symplectic teams for a warm hospitality. We thank Maksim Stokic for pointing our attention to a gap in our proofs in a previous version of the paper. We thank the referee for careful reading and useful comments. The first author was partially supported by ERC Starting Grant 757585 and ISF Grant 2026/17.

Conventions and Notations We convene the following in the course of this paper:

- All our homotopies and isotopies have parameter space $[0,1]$. For instance (g_t) denotes an isotopy $(g_t)_{t\in[0,1]}$.
- Similarly, by concatenation of homotopies we always mean *reparametrized* concatenation.
- If $F : [0,1] \times X \to Y$ is a homotopy with value in a metric space, $\mathrm{Size}\,(F) := \max\{\mathrm{Diam}\,\big(F([0,1] \times \{x\})\big),\ x \in X\}$.
- For $A \subset B$, $\mathrm{Op}\,(A,B)$ stands for an arbitrarily small neighbourhood of A in B. To keep light notation, we omit B whenever there is no possible ambiguity.
- A homotopy $F : [0,1] \times N \to M$ is said relative to $A \subset N$ if for every $z \in A$, $F(t,z)$ is independent on $t \in [0,1]$.
- A homotopy $G : [0,1]^2 \times N \to M$ between $F_0, F_1 : [0,1] \times N \to M$ (that is a continuous map such that $G(i,t,z) = F_i(t,z)$ for $i = 0,1$) is said

relative to A and $\{0,1\}$ if $G(s,t,z) = F_0(t,z) = F_1(t,z)$ for all $z \in A$ and if $G(s,i,z) = F_0(i,z)$ for all $s \in [0,1]$.

2. Quantitative h-principle for isotropic discs

The aim of this section is to prove theorem 1.

2.1. Standard h-principle for subcritical isotropic embeddings

We recall in this section the main properties of the action of the Hamiltonian group on isotropic embeddings, as described in [3,4]. For this purpose, we first fix some notations. In the current note, a disk D^k is always assumed to be closed, unless explicitly stated (hence an embedding of D inside an open set is always compactly embedded). Since we only deal with isotropic embeddings, it is enough to prove theorem 1 for subcritical isotropic embeddings of $[-1,1]^k$ rather than of a closed disc. By abuse of notation, in this section we denote $D^k = [-1,1]^k$. The set of isotropic framings $G^{\mathrm{iso}}(k,n)$ is the space of $(k,2n)$-matrices of rank k whose columns span an isotropic vector space in $(\mathbb{R}^{2n}, \omega_{\mathrm{st}})$.

Recall that the h-principle for subcritical isotropic embeddings provides existence of isotropic embeddings or homotopies whose derivatives realize homotopy classes of maps to $G^{\mathrm{iso}}(k,n)$. We will need a specialization of the h-principle for subcritical isotropic embeddings to $\mathbb{C}^n$, which in particular addresses a relative setting. In order to present its formulation, we will use the following terminology: if $A \subset D^k$, a homotopy of $f : D^k \to G^{\mathrm{iso}}(k,n)$ rel $\mathrm{Op}\,(A)$ is a continuous map $F : [0,1] \times D^k \to G^{\mathrm{iso}}(k,n)$ such that $F(t,z) = f(z)$ for all $z \in \mathrm{Op}\,(A)$. A homotopy $G : [0,1]^2 \times D^k \to G^{\mathrm{iso}}(k,n)$ between $F_0, F_1 : [0,1] \times D^k \to G^{\mathrm{iso}}(k,n)$ (that is a continuous map such that $G(i,t,z) = F_i(t,z)$ for $i = 0,1$) is said relative to $\mathrm{Op}\,(A)$ and $\{0,1\}$ if $G(s,t,z) = F_0(t,z) = F_1(t,z)$ for all $z \in \mathrm{Op}\,(A)$ and if $G(s,i,z) = F_0(i,z)$ for all $s \in [0,1]$ and $i \in \{0,1\}$.

Theorem 2.1. (Parametric $\mathcal{C}^0$-dense relative h-principle for isotropic embeddings [3]) *Let $k < n$:*

(a) Let $\rho : D^k \to \mathbb{C}^n$ be a continuous map whose restriction to a neighbourhood of a closed subset $A \subset D^k$ is an isotropic embedding. Assume that $d\rho$ is homotopic to a map $G : D^k \to G^{\mathrm{iso}}(k,n)$ relative to $\mathrm{Op}\,(A)$. Then, for any $\varepsilon > 0$, there exists an isotropic embedding $u : D^k \hookrightarrow \mathbb{C}^n$ which coincides with ρ on $\mathrm{Op}\,(A)$, $d_{\mathcal{C}^0}(\rho, u) < \varepsilon$ and such that $du : D^k \to G^{\mathrm{iso}}(k,n)$ is homotopic to G rel $\mathrm{Op}\,(A)$.

(b) Let $u_0, u_1 : D^k \hookrightarrow \mathbb{C}^n$ be isotropic embeddings, which coincide on a neighbourhood of a closed subset $A \subset D^k$. Let $G : [0,1] \times D^k \to G^{\mathrm{iso}}(k,n)$ be a homotopy between du_0, du_1 rel $\mathrm{Op}\,(A)$ and $\rho_t : D^k \to \mathbb{C}^n$ a homotopy between u_0, u_1 rel $\mathrm{Op}\,(A)$. For any $\varepsilon > 0$, there exists an isotropic isotopy $u_t : D^k \hookrightarrow \mathbb{C}^n$ ($t \in [0,1]$) relative to $\mathrm{Op}\,(A)$ such that $d_{\mathcal{C}^0}(\rho_t, u_t) < \varepsilon$ and $\{du_t\}$ is homotopic to G rel $\mathrm{Op}\,(A)$ and $\{0,1\}$.

The next lemma will be used in the proof of the theorem 1.

Lemma 2.2. *Let A, B be two closed subsets of D^k. Let $u_0, u_1 : D^k \hookrightarrow \mathbb{C}^n$ be subcritical isotropic embeddings that coincide on* $\mathrm{Op}\,(A)$. *Assume that we are given a homotopy $G_t : D^k \to G^{\mathrm{iso}}(k,n)$ between du_0 and* $\mathrm{d}u_1$ *rel* $\mathrm{Op}\,(A)$. *Let $v_t : D^k \hookrightarrow \mathbb{C}^n$ be an isotropic isotopy between u_0 and v_1 rel* $\mathrm{Op}\,(A)$, *such that $v_{1|\mathrm{Op}\,(B)} = u_1$, and such that $\{\mathrm{d}v_{t|\mathrm{Op}\,(B)}\}$ is homotopic to $\{G_{t|\mathrm{Op}\,(B)}\}$ relative to* $\mathrm{Op}\,(A)$ *and* $\{0,1\}$.[1] *Then* $\mathrm{d}v_1$ *and* $\mathrm{d}u_1$ *are homotopic rel* $\mathrm{Op}\,(A\cup B)$ *among maps $D^k \to G^{\mathrm{iso}}(k,n)$.*

Remark 2.3. In the setting of Lemma 2.2, since v_1 and u_1 are homotopic rel $\mathrm{Op}\,(A \cup B)$ (just consider the linear homotopy between them), the lemma and theorem 2.1 immediately imply that v_1 is in fact isotropic isotopic to u_1 rel $\mathrm{Op}\,(A \cup B)$.

Proof of lemma 2.2. Consider the homotopy $K_t := dv_t : D^k \to G^{\mathrm{iso}}(k,n)$ between du_0 and dv_1 relative to $\mathrm{Op}\,(A)$, and the homotopy $G_t : D^k \to G^{\mathrm{iso}}(k,n)$ between du_0 and du_1 rel $\mathrm{Op}\,(A)$, provided by the assumption. Letting $\overline{K}_t := K_{1-t}$, we now consider the concatenation $H_t := \overline{K}_t \star G_t$. Since $\{dv_{t|\mathrm{Op}\,(B)}\}$ is homotopic to $\{G_{t|\mathrm{Op}\,(B)}\}$ relative to $\mathrm{Op}\,(A)$ and $\{0,1\}$ (as assumed by the lemma), there exists a homotopy $H_{s,t}$ ($s \in [0,1]$) between $H_{t|\mathrm{Op}\,(B)}$ and I_t relative to $\mathrm{Op}\,(A)$ and $\{0,1\}$, where $I_t \equiv \mathrm{d}u_{1|\mathrm{Op}\,(B)} = \mathrm{d}v_{1|\mathrm{Op}\,(B)}$ is a constant homotopy. Let $\chi : D^k \to [0,1]$ be a continuous function such that $\chi(x) = 0$ on a complement of a sufficiently small neighborhood of B in D^k, and $\chi(x) = 1$ on a (smaller) neighborhood of B. Now define a homotopy $\tilde{G}_t : D^k \to G^{\mathrm{iso}}(k,n)$ ($t \in [0,1]$) by

$$\tilde{G}_t(z) := \begin{cases} H_{\chi(z),t}(z) & \text{when } z \in \mathrm{Op}\,(B), \\ G_t(z) & \text{otherwise.} \end{cases}$$

Then $\tilde{G}_t$ is a desired homotopy between $\mathrm{d}u_1$ and $\mathrm{d}v_1$ rel Op $(A \cup B)$. □

We will also need the following lemma, which allows to achieve general positions by Hamiltonian perturbations.

Lemma 2.4. *Let $V \subset \mathbb{C}^n$ be an open set. We consider the following two possible scenarios:*

1 *Let Σ_1, Σ_2 be two smooth proper submanifolds of V, which are transverse in a neighbourhood of ∂V. Then there exists an arbitrarily $\mathcal{C}^1$-small Hamiltonian flow $(\phi^t)_{t\in[0,1]}$ whose generating Hamiltonian is compactly supported in V, such that $\phi^1(\Sigma_1) \pitchfork \Sigma_2$.*

2 *Let Σ_1 be a smooth proper submanifold of V, and let Σ_2 be a smooth manifold such that* $\dim \Sigma_1 + \dim \Sigma_2 \leqslant 2n-2$. *Furthermore, let $\iota_t : \Sigma_2 \to V$ be a smooth proper family of embeddings for $t \in [0,1]$, such that Σ_1 and $\iota_t(\Sigma_2)$ do not intersect near the boundary of V (uniformly in t). Then there exists an arbitrarily $\mathcal{C}^1$-small Hamiltonian flow $(\phi^t)_{t\in[0,1]}$*

[1] Recall that this means there exists a continuous map $G : [0,1]^2 \times \mathrm{Op}\,(B) \to G^{\mathrm{iso}}(k,n)$ such that $G(0,t,z) = G_t(z)$ and $G(1,t,z) = \mathrm{d}v_t(z)$ $\forall (t,z) \in [0,1] \times \mathrm{Op}\,(B)$, $G(s,t,z) = \mathrm{d}u_0(z)$ $\forall (s,t,z) \in [0,1]^2 \times \mathrm{Op}\,(A \cap B)$, $G(s,0,z) = G_0(z) = \mathrm{d}u_0(z)$ and $G(s,1,z) = G_1(z) = \mathrm{d}v_1(z)$ $\forall (s,z) \in [0,1] \times \mathrm{Op}\,(B)$.

whose generating Hamiltonian is compactly supported in V*, such that* $\phi^1(\Sigma_1) \cap \iota_t(\Sigma_2) = \emptyset$ *for any* $t \in [0,1]$*.*

Proof. For both statements, it is enough to show the following claim: if Σ_1, Σ_2 are smooth manifolds (possibly with boundary), and if $f_1 : \Sigma_1 \to V$ and $f_2 : \Sigma_2 \to V$ are smooth proper maps such that $f_1 \pitchfork f_2$ near ∂V, then there exists an arbitrarily small Hamiltonian flow $(\phi^t)_{t\in[0,1]}$ with compact support in V, such that $\phi^1 \circ f_1 \pitchfork f_2$. Indeed, the first statement of the lemma readily follows from this, and for the second statement we can apply the claim with maps the maps $f_1 = \mathrm{Id} : \Sigma_1 \to V$ and $f_2 : \Sigma_2 \times [0,1] \to V$, $f_2(w,t) = \iota_t(w)$.

Now let us show the above claim. Assume that Σ_1, Σ_2 are smooth manifolds (possibly with boundary), and let $f_1 : \Sigma_1 \to V$ and $f_2 : \Sigma_2 \to V$ be smooth maps such that $f_1 \pitchfork f_2$ on $V \setminus K$ where $K \subset V$ is a compact subset. Pick a smooth compactly supported function $h : V \to \mathbb{R}$ such that $h = 1$ on a neighbourhood of K. Now define the smooth map $F : \Sigma_1 \times \Sigma_2 \to \mathbb{C}^n$ by $F(w_1, w_2) = f_2(w_2) - f_1(w_1)$. Then by the Sard theorem, the set of critical values of F has measure zero. Hence there exist arbitrarily small (in norm) regular critical values $v \in \mathbb{C}^n$ of F. Picking such a value v, define the autonomous Hamiltonian function $H : V \to \mathbb{R}$ by $H(z) = h(z)\omega_{std}(v, z)$, where ω_{std} is the standard symplectic form of $\mathbb{C}^n$. Then its Hamiltonian flow verifies $\phi^t_H(z) = z + v$ for $z \in \mathrm{Op}\,(K)$, and it is now easy to see that $\phi^1_H \circ f_1 \pitchfork f_2$ (provided that v is sufficiently close to the origin). □

We finally state a version of Theorem 2.1 which we will use later on:

Proposition 2.5. *Let* $V \subset \mathbb{R}^{2n}$ *be an open set,* $u_0, u_1 : \mathring{D}^l \times [-1,1]^{k-l} \hookrightarrow V$ *be proper subcritical isotropic embeddings which coincide on* $\mathrm{Op}\,(\partial D^l \times [-1,1]^{k-l})$*, such that* u_0 *and* u_1 *are homotopic in* V *relative to* $\mathrm{Op}\,(\partial D^l \times [-1,1]^{k-l})$*, and moreover their differentials* du_0*,* $\mathrm{d}u_1$ *are homotopic in* $G^{\mathrm{iso}}(k,n)$ *relative to* $\mathrm{Op}\,(\partial D^l \times [-1,1]^{k-l})$*. We fix such a relative homotopy* $G : [0,1] \times \mathring{D}^l \times [-1,1]^{k-l} \to G^{\mathrm{iso}}(k,n)$ *between* du_0 *and* du_1*. If* $l = 1$*, we further assume that the curves given by restrictions of* u_0 *and* u_1 *to* $\mathring{D}^1 \times \{0\} = (-1,1) \times \{0\} \subset \mathbb{R}^k$ *have the same actions, i.e. for a* 1*-form* λ *which is a primitive of* ω *in* V*,*

$$\int_{(-1,1)\times\{0\}} u_1^*\lambda - u_0^*\lambda = 0.$$

Then there exists a Hamiltonian isotopy (ϕ^t) *with compact support in* V *such that* $\phi^1 \circ u_0 = u_1$ *and for the induced isotropic isotopy* $u_t = \phi^t \circ u_0$*,* $\{du_t\}$ *is homotopic to* G *rel* $\mathrm{Op}\,(\partial D^l \times [-1,1]^{k-l})$ *and* $\{0,1\}$*.*

Proof. Consider the closed ball $D := D^l = \overline{B}^l(0,1)$, denote $D(r) := \overline{B}^l(0,r)$, $A_{\varepsilon',\varepsilon} := D(1-\varepsilon')\backslash\mathring{D}(1-\varepsilon)$ and $A_\varepsilon := D\backslash\mathring{D}(1-\varepsilon)$. By assumption, there exists $\varepsilon_0 > 0$ such that u_0, u_1 coincide on $A_{\varepsilon_0} \times [-1,1]^{k-l}$ and moreover the homotopy G is relative to $A_{\varepsilon_0} \times [-1,1]^{k-l}$ and $\{0,1\}$. We fix $0 < \varepsilon_1 < \varepsilon_0$.

The restrictions of the maps u_0, u_1 to $D(1-\varepsilon_1)\times[-1,1]^{k-l}$ coincide on $A := A_{\varepsilon_1,\varepsilon_0}\times[-1,1]^{k-l}$, and G provides a homotopy between their differentials relative to A. By theorem 2.1, there exists a compactly supported time-dependent Hamiltonian function $H:[0,1]\times V\to\mathbb{R}$ whose flow ϕ_H^t isotopes $u_{0|D(1-\varepsilon_1)\times[-1,1]^{k-l}}$ to u_1 relative to A, with $\{d(\phi_H^t\circ u_0)_{|D(1-\varepsilon_1)\times[-1,1]^{k-l}}\}$ homotopic to G relative to A and $\{0,1\}$. The subcritical assumption allows us to apply Lemma 2.4 and assume that

$$\varphi_H^t\circ u_0(D(1-\varepsilon_0)\times[-1,1]^{k-l})\cap u_0(A_{\varepsilon_1}\times[-1,1]^{k-l})=\emptyset \tag{2.1.1}$$

for every $t\in[0,1]$. Since we moreover have

$$\phi_H^t\circ u_{0|A_{\varepsilon_1,\varepsilon_0}\times[-1,1]^{k-l}}=u_0, \tag{2.1.2}$$

we obtain the family of embeddings

$$\begin{aligned} u_t : \overset{\circ}{D}{}^l\times[-1,1]^{k-l} &\longrightarrow V\\ (x,y) \quad &\longmapsto \begin{cases}\phi_H^t\circ u_0(x,y) & \text{if } x\in D(1-\varepsilon_1),\\ u_0(x,y) & \text{if } x\in A_{\varepsilon_1}\end{cases}\end{aligned}$$

that provides an isotropic isotopy between u_0 and u_1 relative to $A_{\varepsilon_0}\times[-1,1]^{k-l}$, whose differential realizes G. At this point a distinction is necessary.

• If $l\geq 2$, A is connected, pointwise fixed by ϕ_H^t, hence the differential of $H(t,\cdot)$ vanishes on $u_0(A)$ and in particular $H(t,\cdot)$ assumes a constant value c_t on $u_0(A)$. The Hamiltonian $H'(t,\cdot):=H(t,\cdot)-c_t$ therefore vanishes on $u_0(A)$ together with its differential, and induces the same isotopy between $u_{0|D(1-\varepsilon_1)\times[-1,1]^{k-l}}$ and u_1 relative to A. Then, (2.1.1) and (2.1.2) guarantee that if we cut H' off away from a sufficiently small neighborhood of $\cup_{t\in[0,1]}u_t(D(1-\varepsilon_0)\times[-1,1]^{k-l})$ then we obtain a compactly supported in V Hamiltonian function F such that $\phi_F^t\circ u_0=u_t$ for each $t\in[0,1]$.

• If $l=1$, A is not connected and the above argument cannot be carried out unless we ensure that

$$\alpha_t := \int_{(-1,1)\times\{0\}} u_t^*\lambda-u_0^*\lambda \tag{2.1.3}$$

vanishes. Since however this is not automatic because A is no longer connected, we first alter u_t to another isotopy u_t' that satisfies this property.

By assumption we have $\alpha_0=\alpha_1=0$. Let $K:V\to\mathbb{R}$ be a compactly supported Hamiltonian function such that

$$\begin{aligned} &K_{|\mathrm{Op}\,(u_0([-1+\varepsilon_1,-1+\varepsilon_0]\times[-1,1]^{k-1}))}\equiv 0\\ &\text{and}\ \ K_{|\mathrm{Op}\,(u_0([1-\varepsilon_0,1-\varepsilon_1]\times[-1,1]^{k-1}))}\equiv 1.\end{aligned} \tag{2.1.4}$$

Then $\tilde u_t:=\phi_K^{-\alpha_t}\circ u_t$ agrees with u_0 on $A=([-1+\varepsilon_1,-1+\varepsilon_0]\cup[1-\varepsilon_0,1-\varepsilon_1])\times[-1,1]^{k-1}$, we have $\tilde u_0=u_0$, and by $\alpha_1=0$ we moreover have $\tilde u_1=u_1$. In addition, by (2.1.4) and (2.1.3) we get

$$\int_{(-1+\varepsilon_1,1-\varepsilon_1)\times\{0\}}\tilde u_t^*\lambda-u_0^*\lambda=-\alpha_t+\int_{(-1+\varepsilon_1,1-\varepsilon_1)\times\{0\}}u_t^*\lambda-u_0^*\lambda=0. \tag{2.1.5}$$

for each $t \in [0,1]$. Now, by applying Lemma 2.4 we may assume that

$$\tilde{u}_t((-1+\varepsilon_0, 1-\varepsilon_0) \times [-1,1]^{k-1}) \cap u_0(A_{\varepsilon_1}) = \emptyset \tag{2.1.6}$$

for every $t \in [0,1]$, where $A_{\varepsilon_1} = ((-1,-1+\varepsilon_1] \cup [1-\varepsilon_1, 1)) \times [-1,1]^{k-1}$. Since we moreover have $\tilde{u}_t = u_0$ on A, we can define the family of embeddings

$$\begin{aligned} u'_t : (-1,1) \times [-1,1]^{k-1} &\longrightarrow V \\ (x,y) &\longmapsto \begin{cases} \tilde{u}_t(x,y) & \text{if } x \in (-1+\varepsilon_1, 1-\varepsilon_1), \\ u_0(x,y) & \text{if } x \in (-1,-1+\varepsilon_1] \cup [1-\varepsilon_1, 1) \end{cases} \end{aligned}$$

that provides an isotropic isotopy between u_0 and u_1 relative to $A_{\varepsilon_0} \times [-1,1]^{k-1}$. To see that the path of differentials du'_t realizes G, consider the family of isotropic *immersions* $(u'_{s,t})_{s,t\in[0,1]}$ given by

$$\begin{aligned} u'_{t,s}\ (-1,1) \times [-1,1]^{k-1} &\longrightarrow V \\ (x,y) &\longrightarrow \begin{cases} \phi_K^{-s\alpha_t} \circ u_t(x,y) & \text{if } x \in [-1+\varepsilon_1, 1-\varepsilon_1], \\ u_0(x,y) & \text{if } x \in [-1,-1+\varepsilon_1] \cup [1-\varepsilon_1, 1] \end{cases} \end{aligned}$$

and then the induced family of differentials $\mathrm{d}u'_{s,t}$ provides us a homotopy between the path $\mathrm{d}u_t = \mathrm{d}u'_{0,t}$ and $\mathrm{d}u'_t = \mathrm{d}u'_{1,t}$ relative to $A_{\varepsilon_0} \times [-1,1]^{k-1}$ and $\{0,1\}$, while the path du_t is in turn homotopic to G relative to $A_{\varepsilon_0} \times [-1,1]^{k-1}$ and $\{0,1\}$.

Now we can proceed similarly as in the previous case (of $l \geq 2$). Denoting by $\widetilde{H}$ the Hamiltonian function of the flow $\phi_K^{-\alpha_t} \circ \phi_H^t$, we have $u'_t = \phi_{\widetilde{H}}^t \circ u_0$ on $[-1+\varepsilon_1, 1-\varepsilon_1] \times [-1,1]^{k-1}$. Then by (2.1.5) we have

$$\int_{(-1+\varepsilon_1, 1-\varepsilon_1)\times\{0\}} (u'_t)^*\lambda - u_0^*\lambda = 0$$

for each $t \in [0,1]$, and moreover the flow $\phi_{\widetilde{H}}^t = \phi_K^{-\alpha_t} \circ \phi_H^t$ is the identity when restricted to $u_0(A)$ (where $A = ([-1+\varepsilon_1, -1+\varepsilon_0] \cup [1-\varepsilon_0, 1-\varepsilon_1]) \times [-1,1]^{k-1}$), therefore $\widetilde{H}(t,\cdot)$ assumes a constant value c_t on $u_0(A)$ and its differential vanishes on $u_0(A)$, for each t. Hence denoting $H'(t,\cdot) := \widetilde{H}(t,\cdot) - c_t$, the transversality property (2.1.6) implies that a Hamiltonian function F obtained as a cutoff of H' away from a sufficiently small neighborhood of $\cup_{t\in[0,1]} u'_t([-1+\varepsilon_0, 1-\varepsilon_0] \times [-1,1]^{k-1})$, satisfies $\phi_F^t \circ u_0 = u'_t$ for each $t \in [0,1]$. □

2.2. Proof of theorem 1

Let $k < n$, $D^k := [-1,1]^k$, $D^k(\mu) := [-1-\mu, 1+\mu]^k$, $u_0, u_1 : D^k \hookrightarrow V \subset \mathbb{C}^n$ be smooth isotropic embeddings, and $F : D^k \times [0,1] \to V$ a homotopy between u_0, u_1 with $\operatorname{Size} F < \varepsilon$. We need to prove that there exists a *Hamiltonian* isotopy of size 2ε, which takes u_0 to u_1 on D^k.

Before passing to the proof, we need to modify slightly the framework. First, extend the isotropic embeddings and the homotopy to slightly larger isotropic embeddings: $u_0, u_1 : D^k(\mu) \hookrightarrow V$, $F : D^k(\mu) \times [0,1] \to V$, where $D^k(\mu) = [-\mu, 1+\mu]^k$. By Lemma 2.4, we do not loose generality if we assume that the images of u_0 and u_1 are disjoint (since $k < n$), which we do henceforth. Next, the homotopy F can be turned into a more convenient object:

Lemma 2.6. (see lemma A.1[2]) *There exists a smooth embedding $\tilde{F} : D^k(\mu)\times [0,1] \hookrightarrow V$, with $\tilde{F}(x,0) = u_0(x)$, $\tilde{F}(x,1) = u_1(x)$, with Diam$(\tilde{F}(\{x\} \times [0,1])) < 2\varepsilon$ for all $x \in D^k(\mu)$. In other words, $\tilde{F}$ has size 2ε when considered as a homotopy between u_0, u_1.*

Now $\tilde{F}$ can be further extended to an embedding, still denoted $\tilde{F}$,

$$\tilde{F} : D^k(\mu) \times [-\mu, 1+\mu] \times [-\mu,\mu]^{2n-k-1} \hookrightarrow V.$$

Consider now a regular grid $\Gamma_0 := \nu\mathbb{Z}^k \cap D^k$ in $D^k \subset D^k(\mu)$, of step $\nu \ll 1$ (to be specified later), where $\nu^{-1} \in \mathbb{N}$. This grid generates a cellular decomposition of D^k, whose l-skeleton Γ_l is the union of the l-faces. The set of k-faces has a natural integer-valued distance, where the distance between k-faces x and x' is the minimal m such that there exists a sequence $x = x_0, x_1, \ldots, x_m = x'$ of k-faces and $x_j \cap x_{j+1} \neq \emptyset$ for each $j \in [0, m-1]$ (note that those intersections are not required to be along full $(k-1)$-faces). Fix some $\eta < \nu/2$, and for each $x \in \Gamma_0$, let U_x be the η-neighbourhood of $\{x\} \times [0,1] \times \{0\}^{2n-k-1}$ in $\mathbb{C}^n$, and then denote $W_x := \tilde{F}(U_x)$. Similarly, for each k-face x_k, denote by U_{x_k} the η-neighbourhood of $x_k \times [0,1] \times \{0\}^{2n-k-1}$ in $\mathbb{C}^n$, and then put $W_{x_k} := \tilde{F}(U_{x_k})$. For a k-face x and $m \geqslant 0$ we denote $W_x^m := \cup W_{x'}$, where the union is over all the k-faces x' which are at distance at most m from x. Note that $W_x^0 = W_x$, and that W_x^m is a topological ball. Finally, we put $W := \cup_x W_x \subset V$, where the union is over all the k-faces. Hence, $W = \tilde{F}(U)$ where U is the η-neighborhood of $D^k \times [0,1] \times \{0\}^{2n-k-1}$ in $\mathbb{C}^n$.

We will prove Theorem 1 by successively isotopying the l-skeleton with a control on each isotopy. Precisely, arguing by induction on l, we prove the following:

Proposition 2.7. *There exist Hamiltonian isotopies (Ψ_l^t), $l \in [0,k]$ with support in W, and modified embeddings $v_0 := \Psi_0^1 \circ u_0$, $v_l := \Psi_l^1 \circ v_{l-1}$, such that*

($\mathcal{I}1$) $v_l \equiv u_1$ on a neighbourhood of the l-skeleton Γ_l, for every $l \in [0,k]$.
($\mathcal{I}2$) $v_l(x) \subset W_x^{3^l-1}$ for each k-face x and every $l \in [0, k-1]$.
($\mathcal{I}3$) $\Psi_l^t(W_x) \subset W_x^{2\cdot 3^{l-1}}$ for each k-face x and $l \in [1, k-1]$,
and $\Psi_0^t(W_x) \subset W_x$, $\Psi_k^t(W_x) \subset W_x^{3^{k(k+1)}}$, for every k-face x.
($\mathcal{I}4$) $v_l(\mathring{x}_{l+1}) \cap u_1(\mathring{x}'_{l+1}) = \emptyset$ for every pair of distinct $(l+1)$-faces, $\forall l \in [0, k-1]$.
($\mathcal{I}5$) $\mathrm{d}v_l$ and $\mathrm{d}u_1$ are homotopic rel $\mathrm{Op}\,(\Gamma_l)$ among maps $D^k(\mu) \to G^{\mathrm{iso}}(k,n)$, for each $l \in [0, k-1]$.

Proposition 2.7 readily implies Theorem 1. Indeed, denoting by $(\Psi^t)_{t\in[0,1]}$ the (reparametrized) concatenation $\{\Psi_k^t\} \star \cdots \star \{\Psi_1^t\}$ of the flows, from ($\mathcal{I}3$) we conclude that for each k-face x and each t we have $\Psi^t(W_x) \subset W_x^{3^{k^2+k+1}}$ since $\left(\sum_{j=1}^{k-1} 2\cdot 3^j\right) + 3^{k(k+1)} < 3^{k^2+k+1}$. The flow (Ψ^t) is supported in $W = \cup_{x\in\Gamma_k} W_x \subset V$, and if the step ν of the grid is chosen to be sufficiently small, then for each k-face x, the diameter of $W_x^{3^{k^2+k+1}}$ is less than

2ε. Consequently, the size of the flow $(\Psi^t)_{t\in[0,1]}$ is less than 2ε. Moreover, by $(\mathcal{I}1)$ we have $\Psi^1 \circ u_0 = v_k = u_1$ on D^k. $\square$

Proof of proposition 2.7. As already explained, the proof goes by induction over the dimension of the skeleton Γ_l.

Since $D^k(\mu)$ is contractible, there exists a homotopy $G_t : D^k \to G^{\text{iso}}(k,n)$ between $\mathrm{d}u_0$ and $\mathrm{d}u_1$.

The 0-skeleton: Let $x \in \Gamma_0$ be a 0-face, $\rho < \eta$, and $D_\rho(x)$ the ρ-neighbourhood of x in $D^k(\mu)$. Then $u_0(D_\rho(x)), u_1(D_\rho(x))$ both lie in W_x, and $\tilde{F}$ provides an isotopy between $u_{0|D_\rho(x)}$ and $u_{1|D_\rho(x)}$ in W_x. By theorem 2.1.b), there exists a Hamiltonian isotopy (ψ_x^t) with support in W_x, such that $\psi_x^1 \circ u_0 = u_1$ on $D_\rho(x)$ and $d\psi_x^t \circ du_{0|D_\rho(x)}$ is homotopic to G_t rel $\{0,1\}$. Since $W_x \cap W_{x'} = \emptyset$ for different 0-faces x, x', the isotopies ψ_x have pairwise disjoint supports.

The flow $\psi_0^t := \circ_x \psi_x^t$ (where the composition runs over all 0-faces x of Γ) and the disc $v_0' := \psi_0^1 \circ u_0$ verify $(\mathcal{I}1)$ by construction. Moreover, the flow satisfies $(\mathcal{I}3)$ because it is supported inside the disjoint union $\cup_{x\in\Gamma_0} W_x$, and for every $x \in \Gamma_0$ and k-face x' we have either $W_x \subset W_{x'}$ or $W_x \cap W_{x'} = \emptyset$. In addition, $d\psi_0^t \circ du_{0|\mathrm{Op}\,(\Gamma_0)}$ is homotopic to G_t rel $\{0,1\}$. In the next steps of the proof we will need proposition 2.5 for performing relative isotopies *via localized Hamiltonians*. Note however that in the case of $l = 1$, in addition to the formal obstructions, the proposition requires the actions of the edges to coincide. Hence in order to proceed, *we have to adjust the actions of the edges*.

Let us show that there exists a Hamiltonian isotopy $(\psi_{\mathcal{A}}^t)$, supported in an arbitrarily small neighborhood $v_0'(\Gamma_0) = u_1(\Gamma_0)$, whose flow is the identity on a (smaller) neighbourhood of Γ_0, such that

$$\mathcal{A}\big(\psi_{\mathcal{A}}^1 \circ v_0' \circ \gamma\big) := \int_{\psi_{\mathcal{A}}^1 \circ v_0' \circ \gamma} \lambda = \int_{u_1\circ\gamma} \lambda = \mathcal{A}\big(u_1 \circ \gamma\big) \text{for every edge} \gamma of \Gamma,$$

where by an edge γ of Γ here we mean a parametrized 1-face of Γ. The argument is very similar to the one for symplectic 2-discs given in [2, Page 17], however a small modification is needed since here we are dealing with isotropic discs (instead of symplectic 2-discs). Look at the discs v_0' and u_1. For any edge (i.e. a parametrized 1-face) γ of Γ, the actions $\mathcal{A}(v_0' \circ \gamma) = \int_{v_0'\circ\gamma} \lambda$ and $\mathcal{A}(u_1 \circ \gamma) = \int_{u_1\circ\gamma} \lambda$ do not necessarily coincide. Fix a 0-face $z_0 \in \Gamma_0$, and for any other 0-face $z \in \Gamma_0$, choose a path γ_z made of successive edges of Γ which joins z_0 to z. Define

$$a_z := \int_{u_1\circ\gamma_z} \lambda - \int_{v_0'\circ\gamma_z} \lambda.$$

Notice that these numbers depend on the choice of z_0 but not of γ_z since v_0, u_1 are isotropic. Then, for each edge γ of Γ,

$$\mathcal{A}(v_0' \circ \gamma) + a_{\gamma(1)} - a_{\gamma(0)} = \mathcal{A}(u_1 \circ \gamma)$$

(because $a_{\gamma(1)}$ can be obtained by integrating λ along a path that joins z_0 to $\gamma(0)$, concatenated with γ). Now choose disjoint spherical shells $A_z = \{w \in$

$\mathbb{C}^n \mid \rho_z < |w-z| < \rho'_z\} \subset W_z$, for all $z \in \Gamma_0$. Consider a Hamiltonian function $H_{\mathcal{A}}$ with support in $\cup_z B(z, \rho'_z)$, and which is equal to $-a_z$ on $B(z, \rho_z)$. The induced Hamiltonian isotopy $(\psi^t_{\mathcal{A}})$ is supported inside $\cup_z W_z$, and its time-1 map $\psi^1_{\mathcal{A}}$ is such that for every edge γ of Γ, the area between $v'_0 \circ \gamma$ and $\psi^1_{\mathcal{A}} \circ v'_0 \circ \gamma$ equals $a_{\gamma(1)} - a_{\gamma(0)}$, hence now the actions of u_1 and $\psi^1_{\mathcal{A}} \circ v'_0$ coincide on each edge. Since $\psi^t_{\mathcal{A}} \equiv \mathrm{Id}$ near Γ_0, $\tilde{\Psi}^t_0 := (\psi^t_{\mathcal{A}}) \star (\psi^t_0)$ and $\tilde{v}_0 := \tilde{\Psi}^1_0 \circ u_0 = \psi^1_{\mathcal{A}} \circ v'_0$ still verify $(\mathcal{I}1)$, and the restriction of $d\tilde{\Psi}^t_0 \circ \mathrm{d}u_0 = \mathrm{d}\psi^t_{\mathcal{A}} \circ \mathrm{d}\psi^t_0 \circ \mathrm{d}u_0$ to $\mathrm{Op}\,(\Gamma_0)$ is still homotopic to G_t rel $\{0, 1\}$. Also, since $(\psi^t_{\mathcal{A}})$ is supported in $\cup_z W_z$, $(\mathcal{I}3)$ remains to hold for the flow $(\tilde{\Psi}^t_0)$, and in addition we have $\mathcal{A}(\tilde{v}_0 \circ \gamma) = \mathcal{A}(u_1 \circ \gamma)$ for every edge γ of Γ.

However, $\tilde{v}_0$ might not verify $(\mathcal{I}4)$. Still, since $\tilde{v}_0$ coincides with u_1 on a neighbourhood of Γ_0, there exist closed balls $\overline{B}_{x_0} = \overline{B}(u_1(x_0), r) \subset W_{x_0}$ for each 0-face x_0 of Γ, such that $(\mathcal{I}4)$ is verified inside these balls. Therefore the traces of the submanifolds $\tilde{v}_0(x_1)$ and $u_1(x'_1)$ inside $\underset{x_0 \in \Gamma_0}{\cup} \big(W_{x_0} \backslash \overline{B}_{x_0}\big)$ verify the hypothesis of Lemma 2.4 (1), for every pair of distinct 1-faces x_1, x'_1. Thus an arbitrarily $\mathcal{C}^1$-small Hamiltonian flow (ϕ^t_0) whose generating Hamiltonian is supported in $\underset{x_0 \in \Gamma_0}{\cup} \big(W_{x_0} \backslash \overline{B}_{x_0}\big) \subset \underset{x_0 \in \Gamma_0}{\cup} W_{x_0}$ achieves $\phi^1_0 \circ \tilde{v}_0(x_1) \pitchfork u_1(x'_1)$, for every pair x_1, x'_1 of different 1-faces of Γ (hence these intersections are empty). Now the (reparametrized) concatenation $\Psi^t_0 := (\phi^t_0) \star (\tilde{\Psi}^t_0)$ of the flows verifies $(\mathcal{I}4)$, still verifies $(\mathcal{I}1)$, and $(\mathcal{I}3)$ still holds for $v_0 := \Psi^1_0 \circ u_0$. Since $\phi^t_0 \equiv \mathrm{Id}$ near Γ_0, the restriction $d\Psi^t_0 \circ du_{0|\mathrm{Op}\,(\Gamma_0)}$ is still homotopic to G_t rel $\{0, 1\}$. Since the flow ϕ^t_0 is generated by a Hamiltonian function that vanishes on $\underset{x_0 \in \Gamma_0}{\cup} \overline{B}_{x_0}$, the equality of actions $\mathcal{A}(v_0 \circ \gamma) = \mathcal{A}(u_1 \circ \gamma)$ remains to hold for every edge γ of Γ. Finally, $(\mathcal{I}2)$ follows immediately from $(\mathcal{I}3)$, and v_0 satisfies $(\mathcal{I}5)$ by direct application of Lemma 2.2.

The $\boldsymbol{l}$-skeleton ($\boldsymbol{1 \leqslant l < n-1}$): Here we assume that $\Psi_1, \ldots, \Psi_{l-1}$ have been constructed, and we proceed with the induction step. Recall that $v_{l-1} = \Psi^1_{l-1} \circ \cdots \circ \Psi^1_0 \circ u_0$ coincides with u_1 on $\mathrm{Op}\,(\Gamma_{l-1})$ and that $v_{l-1}(x_k) \subset W^{3^{l-1}-1}_{x_k}$ for every k-face x_k. Recall also that we have a homotopy $G^l_t : D^k \to G^{\mathrm{iso}}(k, n)$ between dv_{l-1} and du_1 rel $\mathrm{Op}\,(\Gamma_{l-1})$. Now our aim is to find a Hamiltonian flow (Ψ^t_l) which in particular isotopes $v_{l-1|\mathrm{Op}\,(x_l)}$ to $u_{1|\mathrm{Op}\,(x_l)}$, for each l-face x_l.

Fix an l-face x_l of Γ. By $(\mathcal{I}1)$, there exists an open box $\hat{x}_l \Subset \overset{\circ}{x}_l$ such that v_{l-1} and u_1 coincide on $\mathrm{Op}\,(x_l \backslash \hat{x}_l)$. Choose a k-face x_k which contains x_l. Since $u_1(\hat{x}_l)$ and $v_{l-1}(\hat{x}_l)$ both lie in the topological ball $W^{3^{l-1}-1}_{x_k}$ and coincide near their boundary, there exists a homotopy

$$\sigma_{x_l} : \hat{x}_l \times [0, 1] \to W^{3^{l-1}-1}_{x_k}$$

such that $\sigma_{x_l}(\cdot, 0) = v_{l-1}$, $\sigma_{x_l}(\cdot, 1) = u_1$, and $\sigma_{x_l}(z, t) = u_1(z)\ \forall z \in \mathrm{Op}\,(\partial \hat{x}_l)$, $t \in [0, 1]$. Since $\hat{x}_l \Subset \overset{\circ}{x}_l$ and $l < n$, $(\mathcal{I}4)$ allows to use a general position argument to ensure that moreover $\mathrm{Im}\,\sigma_{x_l}$ admits a regular neighbourhood $\mathcal{V}_{x_l} \subset W^{3^{l-1}-1}_{x_k}$ (a topological ball), such that all these neighbourhoods $\mathcal{V}_{x_l}$ are pairwise disjoint when x_l runs over the l-faces (this is the only point in the proof where we need that $l < n-1$), and such that the restrictions of v_{l-1} and u_1 to $\hat{x}_l$ are proper embeddings into $\mathcal{V}_{x_l}$ for every l-face x_l of Γ.

By assumption, there exists a homotopy $G_l^t : [0,1] \times D^k \to G^{\mathrm{iso}}(k,n)$ between dv_{l-1} and du_1, with $G^t_{l|\mathrm{Op}\,(\Gamma_{l-1})} = \mathrm{d}u_1 = \mathrm{d}v_{l-1}$. Also, $v_{l-1|\mathrm{Op}\,(\hat{x}_l)}$ is clearly homotopic to $u_{1|\mathrm{Op}\,(\hat{x}_l)}$ rel $\mathrm{Op}\,(\partial\hat{x}_l)$ in $\mathcal{V}_{x_l}$, and when $l=1$, $\mathcal{A}(v_{l-1}(\hat{x}_l)) = \mathcal{A}(u_1(\hat{x}_l))$ (in this equality of actions, $\hat{x}_l \subset x_l$ is equipped with a chosen orientation, and the equality holds since the actions of $v_{l-1}(x_l)$ and $u_1(x_l)$ coincide and since v_{l-1} and u_1 agree on $x_l \backslash \hat{x}_l$). Hence by Proposition 2.5, there exist Hamiltonian diffeomorphisms $\psi^t_{x_l}$, where x_l runs over the l-faces, which have support in $\mathcal{V}_{x_l}$, and are such that $\psi^1_{x_l} \circ v_{l-1|\mathrm{Op}\,(\hat{x}_l)} = u_1$, and the restriction of $d(\psi^t_{x_l} \circ v_{l-1})$ to $\mathrm{Op}\,(\partial\hat{x}_l)$ is homotopic to G_l^t relative to $\mathrm{Op}\,(\partial\hat{x}_l)$ and $\{0,1\}$. Let now $\psi_l^t := \circ\psi^t_{x_l}$ and $\hat{v}_l := \psi^1_l \circ v_{l-1}$. Since the $(\psi^t_{x_l})$ have pairwise disjoint supports, we have $\hat{v}_{l|\mathrm{Op}\,(x_l)} = u_{1|\mathrm{Op}\,(x_l)}$ for each l-face x_l of Γ. Hence $\hat{v}_l$ and u_1 coincide on a neighbourhood of the l-skeleton of Γ, so $\hat{v}_l$ verifies $(\mathcal{I}1)$. By Lemma 2.2, $\hat{v}_l$ verifies $(\mathcal{I}5)$ as well.

The flow (ψ^t_l) is supported in the disjoint union $\cup_{x_l\in\Gamma_l}\mathcal{V}_{x_l}$. Let x be any k-face, and assume that we have an l-face x_l such that $\mathcal{V}_{x_l} \cap W_x \neq \emptyset$. Let $x_k \supset x_l$ be a k-face as above, so that $\mathcal{V}_{x_l} \subset W^{3^{l-1}-1}_{x_k}$. Then the distance between x and x_k is not larger than 3^{l-1}, and we conclude $\mathcal{V}_{x_l} \subset W^{3^{l-1}-1}_{x_k} \subset W^{2\cdot 3^{l-1}-1}_{x}$. To summarise, for any k-face x, if x_l is an l-face with $\mathcal{V}_{x_l} \cap W_x \neq \emptyset$, then $\mathcal{V}_{x_l} \subset W^{2\cdot 3^{l-1}-1}_x$. As a result, we get

$$\psi^t_l(W_x) \subset W_x^{2\cdot 3^{l-1}-1}\ . \tag{2.2.7}$$

The embedding $\hat{v}_l$ may fail to satisfy $(\mathcal{I}4)$: there might be two different $(l+1)$-faces x_{l+1}, x'_{l+1} such that

$$\hat{v}_l(\overset{\circ}{x}_{l+1}) \cap u_1(\overset{\circ}{x}{}'_{l+1}) \neq \emptyset.$$

Notice however that since $\hat{v}_l$ and u_1 coincide on a neighbourhood of Γ_l, the set $\hat{v}_l(x_{l+1}) \cap u_1(x'_{l+1})$ is compactly contained in $W\backslash u_1(\Gamma_l)$. By Lemma 2.4 (1), there exists an arbitrarily small Hamiltonian flow $(\phi^t_l)_{t\in[0,1]}$, with compact support in $W\backslash\Gamma_l$ such that $v_l := \phi^1_l \circ \hat{v}_l$ verifies $(\mathcal{I}4)$. By the smallness of the flow (ϕ^t_l) and by (2.2.7), the flow $(\Psi^t_l) := (\phi^t_l) \star (\psi^t_l)$ satisfies $\Psi^t_l(W_x) \subset W_x^{2\cdot 3^{l-1}}$ for any k-face x. Hence $(\mathcal{I}3)$ holds for (Ψ^t_l). Since the support of (ϕ^t_l) is compactly contained in $W\backslash\Gamma_l$, $(\mathcal{I}1)$ and $(\mathcal{I}5)$ still holds for v_l. Finally, $(\mathcal{I}2)$ follows as well: if x is any k-face, then by assumption, $v_{l-1}(x) \subset W_x^{3^{l-1}-1}$, hence by (2.2.7) and $(\mathcal{I}3)$ we get

$$\begin{aligned} v_l(x) = \Psi^1_l \circ v_{l-1}(x) \subset \Psi^1_l(W_x^{3^{l-1}-1}) = \bigcup_{d(x,y)\leqslant 3^{l-1}-1} \Psi^1_l(W_y) \subset \\ \subset \bigcup_{d(x,y)\leqslant 3^{l-1}-1} W_y^{2\cdot 3^{l-1}} = W_x^{3^{l-1}-1+2\cdot 3^{l-1}} = W_x^{3^l-1}. \end{aligned} \tag{2.2.8}$$

The k-skeleton: When $k < n-1$, the procedure described above works perfectly. However, when $k = n-1$, the last step of the induction requires some adjustment. As before, for every k-face x_k, $v_{k-1}(x_k)$ and $u_1(x_k)$ both lie in the topological ball $W^{3^{k-1}-1}_{x_k}$ and coincide near the boundary, hence there exist homotopies

$$\sigma_{x_k} : \hat{x}_k \times [0,1] \to W_{x_k}^{3^{k-1}-1}$$

such that $\sigma_{x_k}(\cdot,0) = v_{k-1|x_k}$, $\sigma_{x_k}(\cdot,1) = u_{1|x_k}$ and $\sigma_{x_k}(z,t) = u_1(z)$ for all $t \in [0,1]$, $z \in \mathrm{Op}\,(\partial x_k)$ (as before, $\hat{x}_k \subset \overset{\circ}{x}_k$ is a closed box such that u_1 and v_{k-1} coincide on $\mathrm{Op}\,(x_k \backslash \overset{\circ}{\hat{x}}_k)$). The difference with the previous steps of the induction is that general position does not make the sets $\mathrm{Im}\,\sigma_{x_k}$ pairwise disjoint. Instead we proceed as follows.

By $(\mathcal{I}4)$, $v_{k-1}(\hat{x}_k) \cap u_1(x'_k) = u_1(\hat{x}_k) \cap u_1(x'_k) = \emptyset$ for every pair of different k-faces x_k, x'_k. By a standard general position argument, since $k < n$, we can therefore assume that $\mathrm{Im}\,\sigma_{x_k} \cap u_1(x'_k) = \emptyset$, and that we have a regular neighbourhood $\mathcal{V}_{x_k} \subset W_{x_k}^{3^{k-1}-1}$ of $\mathrm{Im}\,\sigma_{x_k}$, such that

$$\mathcal{V}_{x_k} \cap u_1(x'_k) = \emptyset \qquad \forall x_k \neq x'_k. \tag{2.2.9}$$

By $(\mathcal{I}5)$, and since $v_{k-1}(\hat{x}_k), u_1(\hat{x}_k)$ are homotopic relative to $\partial\hat{x}_k$ in $\mathcal{V}_{x_k}$, there exists a Hamiltonian isotopy $(\psi^t_{x_k})$ with support in $\mathcal{V}_{x_k}$ such that $\psi^1_{x_k} \circ v_{k-1|x_k} = u_1$.

Consider now a partition of the set of k-faces into $(2\cdot 3^{k-1})^k = 2^k \cdot 3^{k(k-1)}$ subsets F_i $(i = 1,\dots,2^k\cdot 3^{k(k-1)})$, such that any two faces $x_k, x'_k \in F_i$ are at distance at least $2\cdot 3^{k-1}$ from each other. Then for any i and any pair $x_k, x'_k \in F_i$ of distinct k-faces, we have $W_{x_k}^{3^{k-1}-1} \cap W_{x'_k}^{3^{k-1}-1} = \emptyset$. Define $(\psi^t_{k,i}) := \underset{x_k\in F_i}{\circ} \psi^t_{x_k}$, which is a composition of Hamiltonian isotopies, compactly supported in the disjoint union $\cup_{x_k\in F_i} W_{x_k}^{3^{k-1}-1}$. For any k-face x, if we have some $x_k \in F_i$ such that $W_x \cap W_{x_k}^{3^{k-1}-1} \neq \emptyset$, then the distance between x and x_k is at most 3^{k-1}, and hence $W_{x_k}^{3^{k-1}-1} \subset W_x^{2\cdot 3^{k-1}-1}$. We conclude that for any k-face x we have $\psi^t_{k,i}(W_x) \subset W_x^{2\cdot 3^{k-1}-1}$.

Now, letting $(\Psi^t_k) := (\psi^t_{k,2^k\cdot 3^{k(k-1)}})\star\cdots\star(\psi^t_{k,1})$ and arguing as in (2.2.8), we get for any k-face x

$$\Psi^t_k(W_x) \subset W_x^{N_k} \subset W_x^{3^{k(k+1)}},$$

where $N_k = 2^k\cdot 3^{k(k-1)}\cdot(2\cdot 3^{k-1}-1) < 3^{k(k+1)}$. Therefore, $(\mathcal{I}3)$ holds for (Ψ^t_k).

Finally, $\psi^1_{k,i} \circ v_{k-1|\mathrm{Op}\,(x_k)} = u_{1|\mathrm{Op}\,(x_k)}$ for all $x_k \in F_i$, and by (2.2.9), $\psi^1_{x'_k} \circ u_{1|\mathrm{Op}\,(x_k)} = u_{1|\mathrm{Op}\,(x_k)}$ for any pair of k-faces $x'_k \neq x_k$. Thus,

$$\Psi^1_k \circ v_{k-1|\mathrm{Op}\,(x_k)} = u_1 \text{ for every } k-\text{face } x_k \text{ of } \Gamma,$$

which just means that $\Psi^1_k \circ v_{k-1|\mathrm{Op}\,(D^k)} \equiv u_{1|\mathrm{Op}\,(D^k)}$. We have verified $(\mathcal{I}1)$ for $v_k := \Psi^1_k \circ v_{k-1}$. □

3. Action of symplectic homeomorphisms on symplectic submanifolds

3.1. Taking a symplectic disc to an isotropic one

We aim now at proving Theorem 2. The proof relies on Theorem 1 and is similar to the proof of the flexibility of disc area in the context of symplectic 2-discs considered in [2].

Proof of theorem 2. Let

$$\begin{array}{rcl} i_0: \quad D & \longrightarrow & \mathbb{C}\times\mathbb{C}\times\mathbb{C}=\mathbb{C}^3, \\ x+iy & \longmapsto & (x,y,0) \end{array} \quad \begin{array}{rcl} u_0: D & \longrightarrow & \mathbb{C}\times\mathbb{C}\times\mathbb{C} \\ z & \longmapsto & (z,0,0) \end{array}$$

be the standard isotropic and symplectic embeddings of D into $\mathbb{C}^3$. Let also $f_k : D(2) \to D(1/2^k)$ be an area-preserving immersion and

$$\begin{array}{rcl} u_k: \quad D & \longrightarrow & \mathbb{C}\times\mathbb{C}\times\mathbb{C} \\ x+iy & \longrightarrow & (x,y,f_k(x+iy)). \end{array}$$

Then, u_k is a symplectic embedding of D into $\mathbb{C}^3$ with $d_{\mathcal{C}^0}(u_k,i_0)<\frac{1}{2^k}$. Let finally consider an isotropic embedding i_k^l of D into $\mathbb{C}^3$ with $d_{\mathcal{C}^0}(i_k^l,u_k)<\frac{1}{2^l}$. Although less explicit than the previous embedding in dimension 6, it certainly exists because one can approximate the standard symplectic embedding u_0 by isotropic ones of the form $z\mapsto(z,\overline{f_l(z)},0)$. We also define

$$\begin{array}{rl} & W_k(\delta):=\{z\in\mathbb{C}^3 \mid \mathrm{d}(z,\operatorname{Im}u_k)<\delta\} \\ \text{and} & W^0(\varepsilon):=\{z\in\mathbb{C}^3 \mid \mathrm{d}(z,\operatorname{Im}i_0)<\varepsilon\}. \end{array}$$

It is enough to construct a sequence $\phi_0,\phi_1,\dots$ of compactly supported in $\mathbb{C}^3$ symplectic diffeomorphisms, such that for an increasing sequence of indices $k_0=0<k_1<k_2<\dots$ we have $\phi_i\circ u_{k_i}=u_{k_{i+1}}$, and such that moreover, the sequence $\Phi_i=\phi_i\circ\phi_{i-1}\circ\cdots\circ\phi_0$ uniformly converges to a homeomorphism Φ of $\mathbb{C}^3$. We construct such a sequence ϕ_i by induction. Let $\mathbb{C}^3=U_0\supset U_1\supset U_2\supset\cdots\supset u_0(D)$ be a decreasing sequence of open sets such that $\cap U_i=u_0(D)$. In the step 0 of the induction, we let $k_1=1$, and choose ϕ_0 to be any symplectic diffeomorphism with compact support in $\mathbb{C}^3$ such that $\phi_0\circ u_0=u_{k_1}$.

Now we describe a step $i\geqslant 1$. From the previous steps we get $k_1<\cdots<k_i$, and symplectic diffeomorphisms $\phi_0,\dots,\phi_{i-1}$. Denote $\Phi_{i-1}=\phi_{i-1}\circ\cdots\circ\phi_0$. By the step $i-1$, we have $u_{k_i}=\Phi_{i-1}\circ u_0$ and $\Phi_{i-1}(U_{i-1})\supset W^0(\varepsilon_i)$, where $\varepsilon_i=\frac{1}{2^{k_i}}$ (the inclusion $\Phi_{i-1}(U_{i-1})\supset W^0(\varepsilon_i)$ clearly holds when $i=1$ because $U_0=\mathbb{C}^3$, and for $i>1$ it follows from (3.1.1) below which was obtained in the previous step $i-1$). The choice for ε_i implies that $W^0(\varepsilon_i)\supset u_{k_i}(D)$, and moreover by $u_{k_i}=\Phi_{i-1}\circ u_0$ we get $\Phi_{i-1}(U_i)\supset u_{k_i}(D)$, so we conclude $\Phi_{i-1}(U_i)\cap W^0(\varepsilon_i)\supset u_{k_i}(D)$. Hence we can choose a sufficiently large $l_i\geqslant k_i$ such that $\Phi_{i-1}(U_i)\cap W^0(\varepsilon_i)\supset W_{k_i}(\delta_i)\supset i_{k_i}^{l_i}(D)$, where $\delta_i=\frac{1}{2^{l_i}}\leqslant\varepsilon_i$. Note that

$$d_{\mathcal{C}^0}(i_{k_i}^{l_i},i_0)\leqslant d_{\mathcal{C}^0}(i_{k_i}^{l_i},u_{k_i})+d_{\mathcal{C}^0}(u_{k_i},i_0)<\frac{1}{2^{l_i}}+\frac{1}{2^{k_i}}\leqslant 2\varepsilon_i,$$

and moreover $i_0(D), i_{k_i}^{l_i}(D)\subset W^0(\varepsilon_i)$. Hence by the convexity of $W^0(\varepsilon_i)$ and by theorem 1, there exists a Hamiltonian diffeomorphism ϕ_i' supported in

$W^0(\varepsilon_i)$ such that $i_0 = \phi'_i \circ i^{l_i}_{k_i}$ and $d_{\mathcal{C}^0}(\phi'_i, \mathrm{Id}\,) < 4\varepsilon_i$. Note that in particular, $\phi'_i(W_{k_i}(\delta_i)) \supset i_0(D)$.

We claim that there exists a homotopy of *a small size* between the (symplectic) disc $\phi'_i \circ u_{k_i}$ and the (isotropic) disc i_0, inside $\phi'_i(W_{k_i}(\delta_i))$. Indeed, the open set $W_{k_i}(\delta_i)$ contains the discs $u_{k_i}(D), i^{l_i}_{k_i}(D)$. Also we have $d_{\mathcal{C}^0}(u_{k_i}, i^{l_i}_{k_i}) < \delta_i$. Hence the linear homotopy $\rho_i(z,t) := (1-t)u_{k_i}(z)+t i^{l_i}_{k_i}(z)$, ($z \in D$, $t \in [0,1]$), satisfies $d_{\mathcal{C}^0}(u_{k_i}(z), \rho_i(z,t)) < \delta_i$ for all $z \in D$, $t \in [0,1]$, and so by definition of the neighbouhood $W_{k_i}(\delta_i)$, this homotopy ρ_i lies inside $W_{k_i}(\delta_i)$. We moreover conclude that the size of ρ_i is less than δ_i, and therefore the homotopy $\phi'_i \circ \rho_i$ between $\phi'_i \circ u_{k_i}$ and $\phi'_i \circ i^{l_i}_{k_i} = i_0$, lies inside $\phi'_i(W_{k_i}(\delta_i))$, and has size less than $\delta_i + 8\varepsilon_i \leqslant 9\varepsilon_i$ (recall that $d_{\mathcal{C}^0}(\phi'_i, \mathrm{Id}\,) < 4\varepsilon_i$).

We therefore have $\phi'_i(W_{k_i}(\delta_i)) \supset i_0(D)$, and moreover the homotopy $\phi'_i \circ \rho_i$ between $\phi'_i \circ u_{k_i}$ and i_0, lies inside $\phi'_i(W_{k_i}(\delta_i))$, and is of size less than $9\varepsilon_i$. Hence by choosing a sufficiently large $k_{i+1} > k_i$ and denoting $\varepsilon_{i+1} = \frac{1}{2^{k_{i+1}}}$, we get

$$\phi'_i(W_{k_i}(\delta_i)) \supset W^0(\varepsilon_{i+1}) \supset u_{k_{i+1}}(D),$$

and moreover the homotopy between $\phi'_i \circ u_{k_i}$ and $u_{k_{i+1}}$, given by the concatenation of $\phi'_i \circ \rho_i$ and of the linear homotopy between i_0 and $u_{k_{i+1}}$, lies in $\phi'_i(W_{k_i}(\delta_i))$ and still has size less than $9\varepsilon_i$. Applying the quantitative h-principle for symplectic discs [2, Theorem 2], we get a Hamiltonian diffeomorphism ϕ''_i supported in $\phi'_i(W_{k_i}(\delta_i))$, such that $\phi''_i \circ \phi'_i \circ u_{k_i} = u_{k_{i+1}}$ and $d_{\mathcal{C}^0}(\phi''_i, \mathrm{Id}\,) < 18\varepsilon_i$.

As a result, the composition $\phi_i := \phi''_i \circ \phi'_i$ is supported in $W^0(\varepsilon_i) \subset \Phi_{i-1}(U_{i-1})$, we have $\phi_i \circ u_{k_i} = u_{k_{i+1}}$,

$$\phi_i \circ \Phi_{i-1}(U_i) = \phi''_i \circ \phi'_i \circ \Phi_{i-1}(U_i) \supset \phi''_i \circ \phi'_i(W_{k_i}(\delta_i)) = \phi'_i(W_{k_i}(\delta_i)) \supset W^0(\varepsilon_{i+1}) \tag{3.1.1}$$

and

$$d_{\mathcal{C}^0}(\mathrm{Id}\,, \phi_i) \leqslant d_{\mathcal{C}^0}(\mathrm{Id}\,, \phi'_i) + d_{\mathcal{C}^0}(\mathrm{Id}\,, \phi''_i) < 22\varepsilon_i.$$

This finishes the step i of the inductive construction.

To summarize, we have inductively constructed a sequence of Hamiltonian diffeomorphisms $\phi_0, \phi_1, \ldots$ with uniformly bounded compact supports in $\mathbb{C}^3$, such that:

(i) ϕ_i has support in $W^0(\varepsilon_i) \subset \Phi_{i-1}(U_{i-1})$ where $\Phi_{i-1} = \phi_{i-1} \circ \cdots \circ \phi_0$,
(ii) $d_{\mathcal{C}^0}(\mathrm{Id}\,, \phi_i) < 22\varepsilon_i = \frac{22}{2^{k_i}}$,
(iii) $u_{k_{i+1}} = \phi_i \circ u_{k_i}$.

It follows by (ii) that Φ_i is a Cauchy sequence in the $\mathcal{C}^0$ topology, hence uniformly converges to some continuous map $\Phi : \mathbb{C}^3 \to \mathbb{C}^3$. Next, since $u_{k_{i+1}} = \phi_i \circ u_{k_i}$ for every $i \geqslant 0$, we have $i_0 = \Phi \circ u_0$. Finally, we claim that Φ is an injective map, hence a homeomorphism. To see this, consider two points $x \neq y \in U_0 = \mathbb{C}^3$. If $x, y \in u_0(D)$, then by (iii), $\Phi(x) = i_0 \circ u_0^{-1}(x) \neq i_0 \circ u_0^{-1}(y) = \Phi(y)$. If $x, y \notin u_0(D)$, then $x, y \in {}^cU_i$ for i large enough, so by (i), $\Phi_i(x) = \Phi_{i+1}(x) = \Phi_{i+2}(x) = \cdots = \Phi(x)$, and similarly $\Phi_i(y) = \Phi(y)$ (because for each $j > i$, the support of ϕ_j lies in $\Phi_{j-1}(U_{j-1}) \subset \Phi_{j-1}(U_i)$), so $\Phi(x) = \Phi_i(x) \neq \Phi_i(y) = \Phi(y)$. Finally, if $x \in u_0(D)$ and $y \notin u_0(D)$, then

$y \in {}^cU_i$ for i large enough, and so $\Phi(y) = \Phi_i(y) \in \Phi_i({}^cU_i) \subset {}^cW^0(\varepsilon_{i+1})$ by (i). Since $\Phi(x) \in \operatorname{Im} i_0 \subset W^0(\varepsilon_{i+1})$, we conclude that also in this case we have $\Phi(x) \neq \Phi(y)$. $\square$

3.2. Relative Eliashberg–Gromov $\mathcal{C}^0$-rigidity

Here we address the following question which appeared in our earlier work [2]:

Question 1. *Assume that a symplectic homeomorphism h sends a smooth submanifold N to a submanifold N', and that $h_{|N}$ is smooth. Under which conditions $h^*\omega_{|N'} = \omega_{|N}$*

Of course, that question is non-trivial only when $\dim N$ is at least 2, which we assume henceforth. The question is particularly interesting in the setting of pre-symplectic submanifolds. Recall that a submanifold $N \subset (M, \omega)$ is called pre-symplectic if ω has constant rank on M. The *symplectic dimension* $\dim^\omega N$ of a pre-symplectic submanifold N is the minimal dimension of a symplectic submanifold that contains N. One checks immediately that $\dim^\omega N = \dim N + \operatorname{Corank} \omega_{|N}$.

In [2], we answered question 1 in various cases of the pre-symplectic setting. Theorem 2 allows to address almost all the remaining cases. Our next result incorporates these remaining cases, together with those verified in [2]:

Theorem 3. *Let $N \subset (M^{2n}, \omega)$ be a pre-symplectic disc. Then the answer to question 1 is*

- *Negative if* $\dim^\omega N \leqslant 2n-4$*, or if* $\dim^\omega N = 2n-2$ *and* $\operatorname{Corank} \omega_{|N} \geqslant 2$.
- *Positive if* $\dim^\omega N = 2n$*, or if* $\dim^\omega N = 2n-2$ *and* $\operatorname{Corank} \omega_{|N} = 0$.

The only case that remains open is when $\dim^\omega N = 2n-2$ and $\operatorname{Corank} \omega_{|N} = 1$ (i.e. $\dim N = 2n-3$, $\operatorname{Corank} \omega_{|N} = 1$).

Proof of theorem 3. When $\dim^\omega N \leqslant 2n-4$ and N is not isotropic, the answer is negative because we can find a symplectic homeomorphism that fixes N and contracts the symplectic form (by [2]). When $\dim^\omega N \leqslant 2n-2$ and $r := \operatorname{corank} \omega_{|N} \geqslant 2$, there is a local symplectomorphism that takes N to $[0,1]^r \times D^k \times \{0\} \subset \mathbb{C}^r_{(z)} \times \mathbb{C}^k_{(z')} \times \mathbb{C}^m_{(w)}$, where $m \geq 1$ and $r \geq 2$. By Theorem 2, we can find a symplectic homeomorphism $f(z_1, z_2, w_1)$ of $\mathbb{C}^2 \times \mathbb{C}$ which takes $[0,1]^2 \times \{0\}$ to a symplectic disc. The induced map

$$\tilde{f}\mathbb{C}^2_{(z_1,z_2)} \times \mathbb{C}_{(w_1)} \times \mathbb{C}^{r-2} \times \mathbb{C}^k \times \mathbb{C}^{m-1} \to \mathbb{C}^n$$
$$(z_1, z_2, w_1, z_3, \ldots, z_r, z'_1, \ldots, z'_k, w_2, \ldots, w_m) \to f(z_1, z_2, w_1) \times \mathrm{Id}$$

is obviously a symplectic homeomorphism which takes N to a submanifold on which the co-rank of the symplectic form is reduced by 2. Note that this argument also works when $\dim^\omega N \leqslant 2n-4$ and N is isotropic. The second item of the theorem was proved in [2]. $\square$

Acknowledgements

We are very much indebted to Claude Viterbo for his support and interest in our research. Claude's fundamental contributions to symplectic geometry and

topology, and in particular to the field $\mathcal{C}^0$ symplectic geometry, are widely recognized. We wish Claude all the best, and to continue enjoying math and delighting us with his creative and inspiring mathematical works.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Buhovsky, L., Humilière, V., Seyfaddini, S.: A C^0 counterexample to the Arnold conjecture. Invent. Math. **213**(2), 759–809 (2018)

[2] Buhovsky, L., Opshtein, E.: Some quantitative results in $\mathcal{C}^0$ symplectic geometry. Invent. Math. **205**(1), 1–56 (2016)

[3] Eliashberg, Y., Mishachev, N.: Introduction to the h-principle. Graduate Studies in Mathematics, vol. 48. American Mathematical Society, Providence, RI (2002)

[4] Gromov, M.: Partial differential relations. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 9. Springer, Berlin (1986)

Lev Buhovsky
School of Mathematical Sciences
Tel Aviv University
Tel Aviv
Israel
e-mail: `levbuh@tauex.tau.ac.il`

Emmanuel Opshtein
Institut de Recherche Mathématique Avancée
Université de Strasbourg et CNRS, UMR
Strasbourg 7501
France
e-mail: `opshtein@math.unistra.fr`

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:33
https://doi.org/10.1007/s11784-022-00938-9
Published online April 8, 2022

Journal of Fixed Point Theory and Applications

On symplectomorphisms and Hamiltonian flows

Franco Cardin

This paper is dedicated to Claude Viterbo, friend and master, on the occasion of his 60th birthday.

Abstract. We propose the construction of a sequence of time one flows of autonomous Hamiltonian vector fields, converging to a fixed near the identity C^1 symplectic diffeomorphism ψ. This convergence is proved to be uniformly exponentially fast, in a non analytic symplectic topology framework.

Mathematics Subject Classification. 53D05, 53D22, 37J11.

Keywords. Symplectomorphisms, hamiltonian systems, symplectic topology.

1. Introduction

The aim of this note is to revisit the classical issue to find in correspondence of any fixed small (i.e. close to identity) symplectic diffeomorphism ψ a sequence of compositions of time one flows $\Phi^1_{X_{F_j}}$ of autonomous Hamiltonian vector fields X_{F_j},

$$\cdots \Phi^1_{X_{F_j}} \circ \Phi^1_{X_{F_{j-1}}} \circ \cdots \circ \Phi^1_{X_{F_1}} \circ \Phi^1_{X_{F_0}}$$

and to discuss its possible convergence to ψ. This subject is crucial in the C^ω Hamiltonian perturbation theory. There is a long history around this matter. After early pioneering papers [5,6], a rigorous setting has been provided mainly in [8,9], and other interesting references therein quoted. More recently, Giorgilli [10] introduces in a genuinely innovative way a 'Lie transform', generalizing the Lie series and having a number of nice algebraic properties; this object appears as a useful tool which could give new help to a greater comprehension of the Baker-Campbell-Hausdorff (BCH) matter. Unlike the analytic framework of the above quoted papers, the novelty in the present work is to

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

act in a C^1 environment for diffeomorphisms ψ, possibly for small compact perturbations of the identity, which are typical in many questions of symplectic topology. Our construction will lead to an infinite product converging uniformly exponentially fast to the fixed ψ:

$$|\Phi^1_{X_{F_j}} \circ \cdots \circ \Phi^1_{X_{F_1}} \circ \Phi^1_{X_{F_0}} - \psi|_{C^0} \leq |\psi - \mathrm{id}|^{2^{j+1}}_{C^0} \tag{1}$$

It will be provided by a global realization of Weinstein's neighbourhood theorem in $T^*\mathbb{R}^n$ proposed by Viterbo (see [4, 20]), and the exponential estimate (1) will be produced by means of standard facts from mid-point approximation theory.

This results suggests that, by means of an adequate version of the BCH theorem, one could solve the problem of the construction of an autonomous Hamiltonian vector field X_F whose time-one flow is a very fine approximation of an assigned symplectic diffeomorphism ψ, all this in a non analytic category. The effort to obtain anyway some good approximations in such a direction has been made inside the community of the Hamiltonian perturbation theory, see the discussion in [2] and also [13] for allied topics in the analytic category. As a matter of fact, the paper [2] by Benettin and Giorgilli is a point of arrival in the analytic realm: the authors show that ε-small symplectic diffeomorphisms can be asymptotically, up to terms of order $O(\exp(-\varepsilon^*/\varepsilon))$, approximated by autonomous Hamiltonian time one flows.

2. Preliminary background. Time one flows of time dependent Hamiltonians.

We start by illustrating the lines of thought about the (radically simpler) comparison of the symplectomorphisms with the Hamiltonian time one flows related to *time dependent* Hamiltonian systems. Let (M, ω) be a symplectic manifold. Denote by $\mathrm{Diff}_{\omega,0}(M)$ the component of the symplectomorphisms $\mathrm{Diff}_\omega(M) = \{\varphi \in \mathrm{Diff}(M) : \varphi^*\omega = \omega\}$ in which the identity lies. The set of all diffeomorphisms that can be obtained as the time one flow $\Phi^1_{X_H}$ of some possibly time dependent Hamiltonian vector field X_H,

$$i_{X_H}\omega = -dH, \qquad H\colon M \times [0,1] \to \mathbb{R}, \tag{2}$$

is a subgroup of $\mathrm{Diff}_{\omega,0}(M)$ called $Ham(M, \omega)$. As Polterovich and Rosen point out [17], although in general the inclusion $\mathrm{Ham}(M, \omega) \subset \mathrm{Diff}_{\omega,0}(M)$ is strict, the difference between the two groups is 'not too big'. Actually, when $\omega = d\vartheta$, the gap between $\mathrm{Ham}(M, \omega)$ and $Diff_{\omega,0}(M)$, is essentially cohomological one, and, e.g. as in [21], this fact can be seen directly in a clear way: on one hand, given $\psi \in \mathrm{Diff}_{\omega,0}(M)$, we have that $\psi^*\vartheta - \vartheta$ is *closed*,

$$d(\psi^*\vartheta - \vartheta) = \psi^* d\vartheta - d\vartheta = 0, \tag{3}$$

on the other hand, if $\psi = \Phi^1_{X_H}$,

$$\begin{aligned} \tfrac{\mathrm{d}}{\mathrm{d}t}\left(\Phi^t_{X_H}\right)^*\vartheta &= (\Phi^t_{X_H})^*(L_{X_H}\vartheta) \\ &= (\Phi^t_{X_H})^*(d\, i_{X_H}\vartheta + i_{X_H}\, d\vartheta) \\ &= (\Phi^t_{X_H})^*(d\, i_{X_H}\vartheta - dH) \\ &= (\Phi^t_{X_H})^* d(i_{X_H}\vartheta - H) \\ &= d(\Phi^t_{X_H})^*(i_{X_H}\vartheta - H), \end{aligned} \tag{4}$$

so that

$$(\Phi^1_{X_H})^*\vartheta - \vartheta = d\int_0^1 (\Phi^t_{X_H})^*(i_{X_H}\vartheta - H)\mathrm{d}t \tag{5}$$

is *exact*. In the cotangent bundle case, $M = T^*Q$ with $\vartheta = pdq$, the generating function in the r.h.s. of (5) becomes the well known $\int_0^1 (p\dot{q} - H)dt$. Hence we can define the Flux map

$$\begin{aligned} \text{Flux}: \ \text{Diff}_{\omega,0}(M) &\longrightarrow H^1(M,\mathbb{R}) \\ \psi &\longmapsto [\psi^*\vartheta - \vartheta], \end{aligned} \tag{6}$$

and we have seen right now that

$$\text{Ham}(M,\omega) \subseteq \ker(\text{Flux}) \tag{7}$$

Banyaga [1] showed that actually (see also [16] Sec. 14.1, and [21] Ex. 6.3 p. 23), since ω is exact,

$$\text{Ham}(M,\omega) = \ker(\text{Flux}) \tag{8}$$

In a topologically trivial environment—like $T^*\mathbb{R}^n$, where $H^1 = \{0\}$—Banyaga's result tells us that $\text{Diff}_{\omega,0}(M) = Ham(M,\omega)$, in other words, any symplectomorphism is the time one flow for some possibly time dependent Hamiltonian; this fact has been known already for a long time inside the specific world of the Hamiltonian perturbations theory—[5], [15]—, as we can read e.g. in [2]. The time dependence is crucial, because in such a case $\text{Ham}(M,\omega)$ turns out to be even a *subgroup* of $Diff_{\omega,0}(M)$: the composition of several time one flows related to (even time independent) Hamiltonian vector fields is a time one flow, for some, possibly time dependent, Hamiltonian vector field.

In the next Sections we revisit the problem of the approximation of the symplectomorphisms by time one flows related to *time independent* Hamiltonian systems. As already said, this subject has long been studied in the context of the theory of perturbation of Hamiltonian systems. We tackle this matter by investigation of simple techniques borrowed from symplectic topology. More precisely, this note would represent an attempt to realise some first steps towards a perturbative symplectic topology scheme; in other words, we direct our attention towards reaching perturbation results starting from a C^1 context, even though iterated Lie brackets or vector fields do force us immediately in a C^∞ and then Gevrey environment, well adequate to consider compactly supported objects, even though the first result, see (43) and (45), is a purely topological one. By the way, we find along our road map a quadratic estimate—see (35)—evoking an analogue one arising in Hamiltonian perturbation theory, producing in that context well known Newton-like efficient approximation algorithms.

3. Symplectomorphisms as Lagrangian submanifolds

The following Lemma is a summary of Weinstein's Lagrangian neighborhood theorem (see [22], and Prop. 3.4.13 - 14 of [14]) in a form that fits our needs.

Lemma 3.1. *Let (M, ω) be a closed symplectic manifold. There exists a symplectomorphism Ψ between a neighborhood $\mathcal{N}_1(\Delta)$ of the diagonal Δ in $(M \times M, \omega \oplus -\omega)$ and a neighborhood $\mathcal{N}_2(\mathcal{O}_{T^*M})$ of the zero section $\mathcal{O}_{T^*M}$ of T^*M:*

$$\Psi : \mathcal{N}_1(\Delta) \subset (M \times M, \omega \oplus -\omega) \longrightarrow \mathcal{N}_2(\mathcal{O}_{T^*M}) \subset (T^*M, \omega_M) \tag{9}$$

and

$$\Psi(\Delta) = \mathcal{O}_{T^*M} \tag{10}$$

Here and inafter, by ω_M we mean the standard symplectic 2-form on T^*M. As a consequence of this Lemma, we have the next fact. Let $\psi \in Diff_{\omega,0}(M)$ be a symplectomorphism sufficiently C^1-close to the identity,

$$\mathrm{graph}(\psi) \subset \mathcal{N}_1(\Delta), \tag{11}$$

then the image by Ψ of the a Lagrangian submanifold graph(ψ) is Lagrangian in T^*M which is candidate to be transverse to the fibers of π_M.

When M is topologically trivial, like $M = T^*\mathbb{R}^n = \mathbb{R}^{2n}$, versions of this Lemma have been utilised in some directions. The following ingenious linear symplectomorphism by Viterbo [20] realises explicitly the above task:

$$\begin{array}{ccc} (T^*\mathbb{R}^n \times T^*\mathbb{R}^n, \underbrace{pr_1^*\omega_{\mathbb{R}^n} - pr_2^*\omega_{\mathbb{R}^n}}_{\Omega}) & \xrightarrow{f} & (T^*(T^*\mathbb{R}^n), \omega_{T^*\mathbb{R}^n}) \\ (q, p, Q, P) & \longmapsto & \left(\frac{q+Q}{2}, \frac{p+P}{2}, p - P, Q - q\right) \end{array} \tag{12}$$

We have denoted the standard projections by $T^*\mathbb{R}^n \xleftarrow{pr_1} T^*\mathbb{R}^n \times T^*\mathbb{R}^n \xrightarrow{pr_2} T^*\mathbb{R}^n$ and we may easily verify the main property $f^*(\omega_{T^*\mathbb{R}^n}) = \Omega$.
This transformation has been e.g. utilized in [3], other authors, as Traynor [19] and Sandon [18], introduced the alternative symplectic map

$$\tau : (q, p, Q, P) \rightarrow (q, P, P - p, q - Q), \tag{13}$$

at first glance simpler, but not possessing the precious property leading us to the midpoint finite reduction of Hamiltonian systems, see Theorem 3.1 right below.

Coming back to our canonical transformation or symplectomorphism $\psi \in C^1(T^*\mathbb{R}^n; T^*\mathbb{R}^n)$,

$$\psi : T^*\mathbb{R}^n \ni (q, p) \longmapsto (Q(q, p), P(q, p)) \in T^*\mathbb{R}^n, \tag{14}$$

let us denote by ℓ its deviation from the identity,

$$\ell = \psi - \mathrm{id} \tag{15}$$

which we suppose compactly supported in $T^*\mathbb{R}^n$, and we ask that

$$\mathrm{Lip}(\ell) = \sup_{x \in T^*\mathbb{R}^n} |\nabla(\psi(x) - x)| < 2. \tag{16}$$

Set

$$\mathrm{graph}(\psi) = \{(x, \psi(x)) \in T^*\mathbb{R}^n \times T^*\mathbb{R}^n : \ x = (q,p) \in T^*\mathbb{R}^n\}, \tag{17}$$

we notice that $f(\mathrm{graph}(\psi))$,

$$\left\{ \left(\frac{q + Q(q,p)}{2}, \frac{p + P(q,p)}{2}, p - P(q,p), Q(q,p) - q \right) : \ (q,p) \in T^*\mathbb{R}^n \right\}, \tag{18}$$

is a deformation of the null section, and also a Lagrangian submanifold Λ of $T^*(T^*\mathbb{R}^n)$. Indeed $\mathrm{graph}(\psi)$ is a Lagrangian submanifold ($\Omega|_{\mathrm{graph}(\psi)} = 0$) and f is a symplectomorphism, hence $\omega_{T^*\mathbb{R}^n}\big|_{f(\mathrm{graph}(\psi))} = 0$.

By denoting as usual $J = \begin{pmatrix} \mathbb{O} & \mathbb{I} \\ -\mathbb{I} & \mathbb{O} \end{pmatrix}$, we write the inclusion of this Lagrangian submanifold into $T^*(T^*\mathbb{R}^n)$,

$$\begin{array}{ccccc} T^*\mathbb{R}^n \overset{j}{\hookrightarrow} & \mathrm{graph}(\psi) \subset T^*\mathbb{R}^n \times T^*\mathbb{R}^n & \overset{f}{\hookrightarrow} & T^*(T^*\mathbb{R}^n) & \xrightarrow{\pi_{T^*\mathbb{R}^n}} T^*\mathbb{R}^n \\ x \longmapsto & (x, \psi(x)) & \longmapsto & \left(\frac{x+\psi(x)}{2}, J(x - \psi(x)) \right) & \longmapsto \ \frac{x+\psi(x)}{2} \end{array} \tag{19}$$

The composed map

$$\pi_{T^*\mathbb{R}^n} \circ f \circ j : \ x \longmapsto \frac{x + \psi(x)}{2} \tag{20}$$

is a diffeomorphism of $T^*\mathbb{R}^n$ into itself, which is a small deformation of the identity. Indeed, it is a perturbation of the identity by the map $\frac{\ell}{2}$ which is contractive by (16) : $\mathrm{Lip}(\frac{\ell}{2}) < 1$. This fact is telling us that $f(\mathrm{graph}(\psi))$ is globally transverse to the fibers of $\pi_{T^*\mathbb{R}^n}$. Hence, there exists a *generating function* $F \in C^2(T^*\mathbb{R}^n; \mathbb{R})$ for it (without Maslov-Hörmander auxiliary parameters, see e.g. [4, 12, 23]),

$$F : T^*\mathbb{R}^n \to \mathbb{R}, \qquad (\xi, \eta) \mapsto F(\xi, \eta), \tag{21}$$

such that

$$f(\mathrm{graph}(\psi)) = \mathrm{im}(dF) \subset T^*(T^*\mathbb{R}^n) \tag{22}$$

Recalling $\mathrm{im}(dF)$ is

$$\left\{ \left(\xi, \eta, \frac{\partial F}{\partial \xi}(\xi,\eta), \frac{\partial F}{\partial \eta}(\xi,\eta) \right) : \ (\xi,\eta) \in T^*\mathbb{R}^n \right\}, \tag{23}$$

linking (18) with (23) by the diffeomorphism

$$(q,p) = x \mapsto \frac{x + \psi(x)}{2} = (\xi, \eta), \tag{24}$$

we obtain

$$\begin{cases} p - P(q,p) = \frac{\partial F}{\partial \xi}\left(\frac{q+Q(q,p)}{2}, \frac{p+P(q,p)}{2} \right) \\ \\ Q(q,p) - q = \frac{\partial F}{\partial \eta}\left(\frac{q+Q(q,p)}{2}, \frac{p+P(q,p)}{2} \right) \end{cases} \tag{25}$$

or

$$\psi - \mathrm{id} = J dF \left(\frac{\mathrm{id} + \psi}{2} \right) \tag{26}$$

Actually, from (26) we see that the generating function F can be interpreted as a genuine *time-independent* Hamiltonian function: more precisely, the above discussion can be summarized in the following

Theorem 3.1. *Let ψ be a $C^1(T^*\mathbb{R}^n;T^*\mathbb{R}^n)$ symplectic diffeomorphism satisfying* $\mathrm{Lip}(\psi-\mathrm{id})<2$ *with $\psi-\mathrm{id}$ compactly supported—see (16). Then there exists a time independent Hamiltonian function $F:\mathbb{R}^{2n}\to\mathbb{R}$ such that ψ is exactly the solution $\varphi^1_{X_F}$ of the* **step one Euler midpoint**[1]*, related to the Hamiltonian vector field $X_F=JdF$:*

$$\varphi^1_{X_F}-\mathrm{id}=JdF\left(\frac{\mathrm{id}+\varphi^1_{X_F}}{2}\right), \tag{27}$$

that is, see (26): $\psi=\varphi^1_{X_F}$.

Note that so far the compact support hypothesis has not yet been used: this will instead be essential in the next Section to guarantee the completeness of the flows.

4. Iterations and Euler midpoint estimates in action

4.1. Setting and definitions

We define

$$h:=\sup_{\mathbb{R}^{2n}}|\psi(x)-x|=|\psi-\mathrm{id}|_{C^0}=|\ell|_{C^0} \tag{28}$$

In view of the iteration implemented in the next Sect. 4.2, we will insert the 'zero' subscript into the mathematical objects introduced above. From (25) we see that

$$(h=)\,h_0=\mathrm{Lip}\,F_0=\sup_{(\xi,\eta)\in\mathbb{R}^{2n}}|\nabla_{\mathbb{R}^{2n}}F_0(\xi,\eta)|=|X_{F_0}|_{C^0} \tag{29}$$

Moreover we define $\mathcal{F}_0$,

$$F_0=\mathcal{F}_0\,h_0 \qquad (\mathrm{Lip}\,\mathcal{F}_0=1) \tag{30}$$

so that (25) reads

$$\begin{cases} p-P(q,p)=\frac{\partial\mathcal{F}_0}{\partial\xi}\big(\frac{q+Q(q,p)}{2},\frac{p+P(q,p)}{2}\big)h_0 \\ \\ Q(q,p)-q=\frac{\partial\mathcal{F}_0}{\partial\eta}\big(\frac{q+Q(q,p)}{2},\frac{p+P(q,p)}{2}\big)h_0 \end{cases} \tag{31}$$

The hypothesis made above that $\psi-\mathrm{id}$ be compactly supported in $T^*\mathbb{R}^n$ assures us that the C^1 Hamiltonian vector field X_{F_0} admits a complete flow $\Phi^t_{X_{F_0}}$.

To look for time one flow of X_{F_0} is equivalent to look for time h_0 flow of $X_{\mathcal{F}_0}$:

$$\Phi^1_{X_{F_0}}=\Phi^{h_0}_{X_{\mathcal{F}_0}} \qquad \text{(flow)} \tag{32}$$

and analogously—by (31)—for the Euler midpoint:

$$\varphi^1_{X_{F_0}}=\varphi^{h_0}_{X_{\mathcal{F}_0}}=\psi \qquad \text{(Eulermidpoint)} \tag{33}$$

[1] See e.g. [7]. difference scheme reduction

Note that if

$$d := \operatorname{diam}\operatorname{supp}(\psi - \mathrm{id}),$$

by asking $F_0|_{\mathbb{R}^{2n}\setminus\operatorname{supp}(\psi - id)} \equiv 0$ (normalization), then

$$|F_0|_{C^0} \le d\, h_0, \qquad \text{and} \quad |F_0|_{C^1} \le (1+d)h_0 \tag{34}$$

4.2. Main estimate and its iterations

Given ψ, by (25) we get a function F_0 such that ψ is *exactly* the step one Euler midpoint representation of the flow for X_{F_0}, $\psi = \varphi^1_{X_{F_0}} = \varphi^{h_0}_{X_{\mathcal{F}_0}}$. By using the Euler midpoint approximation settled in Section A, the deviation of ψ from $\Phi^1_{X_{F_0}}$ is estimated as follows (recall that $h_0 = |\psi - \mathrm{id}|_{C^0} = |X_{F_0}|_{C^0}$)

$$|\Phi^1_{X_{F_0}} - \psi|_{C^0} = |\Phi^{h_0}_{X_{\mathcal{F}_0}} - \psi|_{C^0} =_{(33)} |\Phi^{h_0}_{X_{\mathcal{F}_0}} - \varphi^{h_0}_{X_{\mathcal{F}_0}}|_{C^0} \le_{(60)} h_0^2 = |\psi - id|^2_{C^0} \tag{35}$$

We will iterate this procedure: instead of ψ, in the next step we consider the symplectomorphism $\psi \circ \Phi^{-1}_{X_{F_0}}$, uniformly closer to the identity than ψ.

Preliminarily, recall that, for any function $g : \mathbb{R}^m \to \mathbb{R}^m$ with $|g - \mathrm{id}|_{C^0} < +\infty$ and any diffeomorphism $f : \mathbb{R}^m \to \mathbb{R}^m$, we have that

$$|(g - \mathrm{id}) \circ f|_{C^0} = |g - \mathrm{id}|_{C^0}, \quad \text{and, for } g = f^{-1}, \quad |f - \mathrm{id}|_{C^0} = |f^{-1} - \mathrm{id}|_{C^0}. \tag{36}$$

Coming back to our task:

$$|\psi \circ \Phi^{-1}_{X_{F_0}} - \mathrm{id}|_{C^0} = |\psi - \Phi^1_{X_{F_0}}|_{C^0} \le h_0^2. \tag{37}$$

In the Sect. 3 above we asked on ψ, a given 'first order' diffeomorphism near the identity, the requirement (17) precisely to guarantee that $\frac{1}{2}(\mathrm{id} + \psi)$ is a diffeomorphism too. Differently, now we observe that the composition $\psi \circ \Phi^{-1}_{X_{F_0}}$ is 'second order' diffeomorphism near to the identity, so we claim that the analogous $\frac{1}{2}(\mathrm{id} + \psi \circ \Phi^{-1}_{X_{F_0}})$ is safely a diffeomorphism of $T^*\mathbb{R}^n$ into itself.

$$T^*\mathbb{R}^n \overset{j}{\hookrightarrow} \operatorname{graph}(\psi \circ \Phi^{-1}_{X_{F_0}}) \subset T^*\mathbb{R}^n \times T^*\mathbb{R}^n \overset{f}{\hookrightarrow} T^*(T^*\mathbb{R}^n) \xrightarrow{\pi_{T^*\mathbb{R}^n}} T^*\mathbb{R}^n$$

$$x \mapsto (x, \psi \circ \Phi^{-1}_{X_{F_0}}(x)) \mapsto \left(\frac{x + \psi\circ\Phi^{-1}_{X_{F_0}}(x)}{2}, J\left(x - \psi \circ \Phi^{-1}_{X_{F_0}}(x)\right)\right) \mapsto \frac{x + \psi\circ\Phi^{-1}_{X_{F_0}}(x)}{2} \tag{38}$$

The global transversality with respect to the fibers of $\pi_{T^*\mathbb{R}^n}$ is offering the existence of a new generating function $F_1 : \mathbb{R}^{2n} \to \mathbb{R}$, such that

$$f(\operatorname{graph}(\psi \circ \Phi^{-1}_{X_{F_0}})) = \operatorname{im}(dF_1), \qquad \varphi^1_{X_{F_1}} = \psi \circ \Phi^{-1}_{X_{F_0}}. \tag{39}$$

We rewrite the relations (35) for this second step (note that $h_1 = |\psi \circ \Phi^{-1}_{X_{F_0}} - \mathrm{id}|_{C^0} = |X_{F_1}|_{C^0}$)

$$|\Phi^1_{X_{F_1}} - \psi \circ \Phi^{-1}_{X_{F_0}}|_{C^0} = |\Phi^1_{X_{F_1}} - \varphi^1_{X_{F_1}}|_{C^0} = |\Phi^{h_1}_{X_{\mathcal{F}_1}} - \varphi^{h_1}_{X_{\mathcal{F}_1}}|_{C^0} \le_{(60)} h_1^2 \tag{40}$$

We link h_1 to h_0,

$$h_1 = |\psi \circ \Phi^{-1}_{X_{F_0}} - \mathrm{id}|_{C^0} = |(\psi - \Phi^1_{X_{F_0}}) \circ \Phi^{-1}_{X_{F_0}}|_{C^0} =_{(36_1)} |\psi - \Phi^1_{X_{F_0}}|_{C^0} \le h_0^2 \tag{41}$$

and from

$$|\Phi^1_{X_{F_1}} \circ \Phi^1_{X_{F_0}} \circ \psi^{-1} - \mathrm{id}|_{C^0} = |(\Phi^1_{X_{F_1}} - \psi \circ \Phi^{-1}_{X_{F_0}}) \circ (\Phi^1_{X_{F_0}} \circ \psi^{-1})|_{C^0},$$

we obtain

$$|\Phi^1_{X_{F_1}} \circ \Phi^1_{X_{F_0}} \circ \psi^{-1} - \mathrm{id}|_{C^0} = |\Phi^1_{X_{F_1}} - \psi \circ \Phi^{-1}_{X_{F_0}}|_{C^0} = |\Phi^1_{X_{F_1}} - \varphi^1_{X_{F_1}}|_{C^0} \leq h_1^2 \leq h_0^4 \tag{42}$$

By iterating k-times this procedure, we get

$$|\Phi^1_{X_{F_k}} \circ \Phi^1_{X_{F_{k-1}}} \circ \ldots \Phi^1_{X_{F_1}} \circ \Phi^1_{X_{F_0}} \circ \psi^{-1} - id|_{C^0} \leq |\psi - \mathrm{id}|^{2^{k+1}}_{C^0} = h_0^{2^{k+1}} \tag{43}$$

Furthermore, recalling (25),

$$|X_{F_\alpha}|_{C^0} = |\Phi^1_{X_{F_{\alpha-1}}} \circ \ldots \Phi^1_{X_{F_1}} \circ \Phi^1_{X_{F_0}} \circ \psi^{-1} - \mathrm{id}|_{C^0} \leq |\psi - \mathrm{id}|^{2^\alpha}_{C^0} = h_0^{2^\alpha} \tag{44}$$

In other words, we achieve a sequence of time one Hamiltonian flows $\{\Phi^1_{X_{F_k}}\}_{k\in\mathbb{N}}$ uniformly convergent (and highly fast) to the assigned ψ:

$$C^0\text{-}\lim_{k\to+\infty} \Phi^1_{X_{F_k}} \circ \Phi^1_{X_{F_{k-1}}} \circ \ldots \Phi^1_{X_{F_1}} \circ \Phi^1_{X_{F_0}} = \psi \tag{45}$$

Here below, we list the hierarchy of the steps giving the sequence, where we denote by $\varphi^1_{X_{F_k}}$ the solution of the step one Euler midpoint difference reduction related to X_{F_k}:

$$\begin{aligned}
&\psi - \mathrm{id} = JdF_0\left(\tfrac{\mathrm{id}+\psi}{2}\right)\\
&\psi = \varphi^1_{X_{F_0}}, \quad \Phi^1_{X_{F_0}}\\
&\\
&\psi \circ \Phi^{-1}_{X_{F_0}} - \mathrm{id} = JdF_1\left(\frac{\mathrm{id}+\psi\circ\Phi^{-1}_{X_{F_0}}}{2}\right)\\
&\psi \circ \Phi^{-1}_{X_{F_0}} = \varphi^1_{X_{F_1}}, \quad \Phi^1_{X_{F_1}}\\
&\\
&\psi \circ \Phi^{-1}_{X_{F_0}} \circ \Phi^{-1}_{X_{F_1}} - \mathrm{id} = JdF_2\left(\frac{\mathrm{id}+\psi\circ\Phi^{-1}_{X_{F_0}}\circ\Phi^{-1}_{X_{F_1}}}{2}\right)\\
&\psi \circ \Phi^{-1}_{X_{F_0}} \circ \Phi^{-1}_{X_{F_1}} = \varphi^1_{X_{F_2}}, \quad \Phi^1_{X_{F_2}}\\
&\ldots
\end{aligned} \tag{46}$$

Obviously, (45) does work if $|\psi - \mathrm{id}|_{C^0} = h_0 < 1$, so that, together with condition (16), we could summarize our requirements by

$$|\psi - \mathrm{id}|_{C^1} = |\ell|_{C^1} < 1 \tag{47}$$

The above discussion can be summarized in the following

Theorem 4.1. *Let ψ be a $C^1(T^*\mathbb{R}^n; T^*\mathbb{R}^n)$ symplectic diffeomorphism satisfying $|\psi - \mathrm{id}|_{C^1} < 1$ with $\psi - \mathrm{id}$ compactly supported. Then there exists a sequence of time independent Hamiltonian function $F_k : \mathbb{R}^{2n} \to \mathbb{R}$, $k \in \mathbb{N}$, such that ψ is represented by the uniform composition limit (45).*

This result restores a further known fact: since $Ham(T^*\mathbb{R}^n, d\theta)$ is a group, any symplectomorphism ψ as above in (45) is generated by the time one flow of some, possibly *time dependent*, Hamiltonian vector field.

Acknowledgements

This research was partially supported by the Project MIUR "Regular and stochastic behaviour in dynamical systems" and by Project DOR "Meccanica analitica, teoria del controllo, mezzi continui" of the University of Padova.

Funding Open access funding provided by Università degli Studi di Padova within the CRUI-CARE Agreement

A Synopsis on the implicit Euler midpoint scheme

A.1 Basic theory

What follows is standard in numerical analysis and we reproduce here for the sake of clarity. Consider a curve $x(t)$ and the discrete substitution $t = n\tau$, with step $\tau > 0$,

$$x(t),\ t \in \mathbb{R} \Rightarrow x_n = x(n\tau),\ n \in \mathbb{Z} \tag{48}$$

The following approximations will be justified below:

$$\begin{cases} x(t+\frac{\tau}{2}) \approx \frac{x(t)+x(t+\tau)}{2} & \left(\text{or: } \frac{x_n+x_{n+1}}{2}\right) \\ \dot{x}(t+\frac{\tau}{2}) \approx \frac{x(t+\tau)-x(t)}{\tau} & \left(\text{or: } \frac{x_{n+1}-x_n}{h}\right) \end{cases} \tag{49}$$

First at all, we have to give an estimate of the approximations '$\approx$' in (49).

Beginning from $(49)_1$,

$$x(t+\tfrac{\tau}{2}) = \underbrace{x(t)+\dot{x}(t)\frac{\tau}{2}}_{*} + \tfrac{1}{2}\ddot{x}(t)\tfrac{\tau^2}{4} + O(\tau^3),$$

$$\tfrac{x(t)+x(t+\tau)}{2} = \underbrace{x(t)+\dot{x}(t)\frac{\tau}{2}}_{*} + \tfrac{1}{4}\ddot{x}(t)\tau^2 + O(\tau^3),$$

thus

$$\frac{x(t)+x(t+\tau)}{2} = x\left(t+\frac{\tau}{2}\right) + O(\tau^2). \tag{50}$$

Considering (49)$_2$,

$$\begin{aligned}
\dot{x}(t+\tfrac{\tau}{2}) &= \underbrace{\dot{x}(t)+\ddot{x}(t)\frac{\tau}{2}}_{**}+\tfrac{1}{2}\dddot{x}(t)\tfrac{\tau^2}{4}+O(\tau^3),\\
x(t+\tau) &= x(t)+\dot{x}(t)\tau+\tfrac{1}{2}\ddot{x}(t)\tau^2+O(\tau^3),\\
\tfrac{x(t+\tau)-x(t)}{\tau} &= \underbrace{\dot{x}(t)+\ddot{x}(t)\frac{\tau}{2}}_{**}+O(\tau^2),
\end{aligned}$$

hence

$$\frac{x(t+\tau)-x(t)}{\tau}=\dot{x}\left(t+\frac{\tau}{2}\right)+O(\tau^2). \tag{51}$$

Now, supposing the curve $x(t)$ is solving the ode

$$\dot{x}(t)=X(x(t)), x(0)=x_0, \tag{52}$$

then

$$\begin{aligned}
&\frac{x(t+\tau)-x(t)}{\tau}=\dot{x}(t+\frac{\tau}{2})+O(\tau^2)=\\
&\quad=X\left(x\left(t+\frac{\tau}{2}\right)\right)+O(\tau^2)=X\left(\frac{x(t)+x(t+\tau)}{2}+O(\tau^2)\right)+O(\tau^2),
\end{aligned}$$

and eventually:

$$x(t+\tau)-x(t)=X\left(\frac{x(t)+x(t+\tau)}{2}\right)\tau+O(\tau^3). \tag{53}$$

A.2 Estimates

Here we denote by $y(t+\tau)$ the τ Euler midpoint approximation:

$$y(t+\tau)-x(t)=X\left(\frac{x(t)+y(t+\tau)}{2}\right)\tau \tag{54}$$

and compare it with the general Euler midpoint representation (53) of the exact solution $x(t+\tau)$:

$$x(t+\tau)-x(t)=X\left(\frac{x(t)+x(t+\tau)}{2}\right)\tau+O(\tau^3). \tag{55}$$

By the trivial identity

$$\frac{x(t)+z(t+\tau)}{2}=x(t)+\frac{z(t+\tau)-x(t)}{2}, \tag{56}$$

we can write (54) using (56) with $z=y$,

$$y(t+\tau)-x(t)=X(x(t))\tau+X'(x(t))\frac{y(t+\tau)-x(t)}{2}\tau+O(\tau^3), \tag{57}$$

analogously, we can write (55) using (56) with $z=x$,

$$x(t+\tau)-x(t)=X(x(t))\tau+X'(x(t))\frac{x(t+\tau)-x(t)}{2}\tau+O(\tau^3) \tag{58}$$

Furthermore, since reasonably $\mathbb{I} - \frac{\tau}{2}X'(x(t))(\approx \mathbb{I})$ is invertible,

$$\begin{cases} y(t+\tau) - x(t) = \left[\mathbb{I} - \frac{\tau}{2}X'(x(t))\right]^{-1} X(x(t))\tau + O(\tau^3) \\ \\ x(t+\tau) - x(t) = \left[\mathbb{I} - \frac{\tau}{2}X'(x(t))\right]^{-1} X(x(t))\tau + O(\tau^3) \end{cases}$$

we get

$$y(t+\tau) - x(t+\tau) = O(\tau^3), \tag{59}$$

and, for small τ,

$$|y(t+\tau) - x(t+\tau)| \leq \tau^2 \qquad (y(t) = x(t)) \tag{60}$$

References

[1] Banyaga, A.: Sur la structure du groupe des difféomorphismes qui préservent une forme symplectique. Comment. Math. Helv. **53**(2), 174–227 (1978)

[2] Benettin, G., Giorgilli, A.: On the Hamiltonian interpolation of near-to-the-identity symplectic mappings with application to symplectic integration algorithms. J. Stat. Phys. **74**(5–6), 1117–1143 (1994)

[3] Bernardi, O., Cardin, F.: Poincaré-Birkhoff periodic orbits for mechanical Hamiltonian systems on $T^*\mathbb{T}^n$. J. Math. Phys. **47**(7), 15 (2006)

[4] Cardin, F.: Elementary Symplectic Topology and Mechanics, p. 222. Springer, Berlin (2015)

[5] Deprit, A.: Canonical transformations depending on a small parameter. Celestial Mech. **1**, 12–30 (1970)

[6] Dragt, A.J., Finn, J.M.: Lie series and invariant functions for analytic symplectic maps. J. Math. Phys. **17**(12), 2215–2227 (1976)

[7] Feng, K., Qin, M.: Symplectic Geometric Algorithms for Hamiltonian Systems, p. xxiv+676. Springer, Heidelberg (2010)

[8] Giorgilli, A., Galgani, L.: Formal integrals for an autonomous Hamiltonian system near an equilibrium point. Celestial Mech. **17**(3), 267–280 (1978)

[9] Giorgilli, A., Galgani, L.: Luigi Rigorous estimates for the series expansions of Hamiltonian perturbation theory. Celestial Mech. **37**(2), 95–112 (1985)

[10] Giorgilli, A.: On the representation of maps by Lie transforms. Rendiconti dell'Istituto Lombardo Accademia di Scienze e Lettere. Class. Sci. Mate. Natur. **146**, 251–277 (2012)

[11] Giorgilli, A.: Perturbation methods in celestial mechanics, pp.59. In: Lectures given at the International School SDSM-2017—satellite dynamics and space missions: theory and applications of celestial mechanics, San Martino al Cimino (VT), Italy (2017)

[12] Guillemin, V., Sternberg, S.: Geometric asymptotics. Mathematical Surveys, No. 14. In: American Mathematical Society, Providence, R.I., pp. xviii+474 (1977)

[13] Kuksin, S., Pöschel, J.: On the inclusion of analytic symplectic maps in analytic Hamiltonian flows and its applications. In: Seminar on Dynamical Systems (St. Petersburg, 1991), pp. 96-116, Progr. Nonlinear Differential Equations Appl., vol. 12, Birkhäuser, Basel (1994)

[14] McDuff, D., Salamon, D.: Introduction to Symplectic Topology, Third edition. Oxford University Press, New York, pp. xi+623 (2017)

[15] Moser, J.: Lectures on Hamiltonian systems. Mem. Amer. Math. Soc. No. 81, 60 pp., Amer. Math. Soc., Providence, R.I (1968)

[16] Polterovich, L.: The geometry of the group of symplectic diffeomorphisms. In: Lectures in Mathematics ETH Zürich. Birkhäuser, Basel, pp. xii+132 (2001)

[17] Polterovich, L., Rosen, D.: Function theory on symplectic manifolds. CRM Monograph Series, vol. 34. American Mathematical Society, Providence, RI, pp. xii+203 (2014)

[18] Sandon, S.: Contact homology, capacity and non-squeezing in $\mathbb{R}^{2n} \times S^1$via generating functions. Ann. Inst. Fourier (Grenoble) **61**(1), 145–185 (2011)

[19] Traynor, L.: Symplectic homology via generating functions. Geom. Funct. Anal. **4**(6), 718–748 (1994)

[20] Viterbo, C.: Symplectic topology as the geometry of generating functions. Math. Ann. **292**(1), 685–710 (1992)

[21] Viterbo, C.: An Introduction to Symplectic Topology through Sheaf theory. Princeton, Fall 2010.- Spring, New York (2011)

[22] Weinstein, A.: Symplectic manifolds and their Lagrangian submanifolds. Adv. Math. **6**(1971), 329–346 (1971)

[23] Weinstein, A.: Lectures on symplectic manifolds. In: Corrected reprint. CBMS Regional Conference Series in Mathematics, 29. American Mathematical Society, Providence, R.I., pp. ii+48 (1979)

Franco Cardin
Dipartimento di Matematica Tullio Levi-Civita
Università degli Studi di Padova
via Trieste
63-35121 Padua
Italy
e-mail: cardin@math.unipd.it

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:34
https://doi.org/10.1007/s11784-022-00939-8
Published online April 8, 2022

Journal of Fixed Point Theory and Applications

Lagrangian skeleta and plane curve singularities

Roger Casals

Dedicated to Claude Viterbo on the occasion of his 60th birthday.

Abstract. We construct closed arboreal Lagrangian skeleta associated to links of isolated plane curve singularities. This yields closed Lagrangian skeleta for Weinstein pairs $(\mathbb{C}^2, \Lambda)$ and Weinstein 4-manifolds $W(\Lambda)$ associated to max-tb Legendrian representatives of algebraic links $\Lambda \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$. We provide computations of Legendrian and Weinstein invariants, and discuss the contact topological nature of the Fomin–Pylyavskyy–Shustin–Thurston cluster algebra associated to a singularity. Finally, we present a conjectural ADE-classification for Lagrangian fillings of certain Legendrian links and list some related problems.

Mathematics Subject Classification. 53D12, 57K33, 14H20.

1. Introduction

The object of this note is to study a relation between the theory of isolated plane curve singularities,[1] as developed by Arnol'd and Gusein-Zade [8–10, 61] A'Campo [1–4] Milnor [76] and others, and arboreal Lagrangian skeleta of Weinstein 4-manifolds. In particular, we construct closed Lagrangian skeleta for the infinite class of Weinstein 4-manifolds obtained by attaching Weinstein 2-handles [28, 108] to the link of $f : \mathbb{C}^2 \longrightarrow \mathbb{C}$, where f defines an isolated plane curve singularity at the origin. These closed Lagrangian skeleta allow for an explicit computation of the moduli of microlocal sheaves [60, 80, 98] and also explain the symplectic topology origin of the Fomin–Pylyavskyy–Shustin–Thurston cluster algebra [47] of an isolated singularity.

1.1. Main results

The advent of Lagrangian skeleta and sheaf invariants have underscored the relevance of Legendrian knots in the study of symplectic 4-manifolds [21,

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

[1] The reader is referred to [54] for a beautiful and welcoming introduction to the subject.

28,49,97,98]. The theory of arboreal singularities, as developed by Nadler [78,79], provides a local-to-global method for the computation of categories of microlocal sheaves [80]. These invariants, in turn, yield results in terms of Fukaya categories [49,50]. The existence of arboreal Lagrangian skeleta has been crystallized by L. Starkston [100] in the context of Weinstein 4-manifolds, where this article takes place.

Given a Weinstein 4-manifold $(W, \lambda_{\rm st})$, it is presently a challenge to describe an associated arboreal Lagrangian skeleta $\mathbb{L} \subseteq W$. In particular, there is no general method for finding *closed* arboreal Lagrangian skeleta,[2] or deciding whether these exist. This manuscript explores this question by introducing a new type of closed arboreal Lagrangian skeleta for Legendrian links $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{\rm st})$ which are maximal-tb Legendrian representatives of the smooth link of an holomorphic germ f in $(\mathbb{C}^2, 0)$. In practice, we restrict to studying polynomials $f : \mathbb{C}^2 \longrightarrow \mathbb{C}$, $f \in \mathbb{C}[x,y]$, which define an isolated singularity at the origin, and also suppose that a real morsification $\tilde{f}_t \in \mathbb{R}[x,y]$ of f exists, $t \in (0,1]$. This is an assumption, and we will always take $f \in \mathbb{R}[x,y]$ as our germs. For simplicity of notation, we denote by $\tilde{f}$ a real morsification $\tilde{f}_t \in \mathbb{R}[x,y]$ for some generic but fixed choice of the deformation parameter $t \in (0,1]$. The discussion in this note unravels thanks to the following geometric fact.

Theorem 1.1. *Let $f \in \mathbb{C}[x,y]$ define an isolated singularity at the origin, $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{st})$ be its associated Legendrian link and $\tilde{f} \in \mathbb{R}[x,y]$ a real morsification. Then, the Weinstein pair $(\mathbb{C}^2, \Lambda_f)$ admits the closed arboreal Lagrangian skeleton*

$$\mathbb{L}(\tilde{f}) = L_{\tilde{f}} \cup \mathscr{T}(\vartheta_{\tilde{f}}),$$

obtained by attaching the Lagrangian $\mathbb{D}^2$-thimbles $\mathscr{T}(\vartheta_{\tilde{f}})$ of $\tilde{f}$ to an embedded exact Lagrangian surface $L_{\tilde{f}} \subseteq \mathbb{C}^2$, where $L_{\tilde{f}} \subseteq \mathbb{C}^2$ is (compactly supported) smoothly isotopic to the Milnor fiber $M_f \subseteq \mathbb{C}^2$ of f. □

The two objects Λ_f and $\mathbb{L}(\tilde{f})$ in the statement of Theorem 1.1 require an explanation, which will be given. We rigorously define the notion of a *Legendrian* link $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{\rm st})$ associated to the germ $f \in \mathbb{C}[x,y]$ of an isolated curve singularity in Sect. 2. Note that the smooth link of the singularity $f \in \mathbb{C}[x,y]$, as defined by Milnor [76], and canonically associated to f, is naturally a *transverse* link $T_f \subseteq (\mathbb{S}^3, \xi_{\rm st})$ [38,53,56]. The Legendrian link $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{\rm st})$ will be a maximal-tb Legendrian approximation of T_f. The notation $(\mathbb{C}^2, \Lambda_f)$ refers to the Weinstein pair $(\mathbb{C}^2, \mathscr{R}(\Lambda_f))$, where $\mathscr{R}(\Lambda_f) \subseteq (\mathbb{S}^3, \xi_{\rm st})$ is a small (Weinstein) annular ribbon for the Legendrian link Λ_f.

The Lagrangian skeleton $\mathbb{L}(\tilde{f})$ is also defined in Sect. 2. Note that the Milnor fibration of $f \in \mathbb{C}[x,y]$ is a *symplectic* fibration on $(\mathbb{C}^2, \omega_{\rm st})$, whose symplectic fibers bound the transverse link $T_f \subseteq (\mathbb{S}^3, \xi_{\rm st})$. Nevertheless, the Lagrangian skeleton $\mathbb{L}(\tilde{f})$ is built from an exact Lagrangian surface $L_{\tilde{f}}$ and the vanishing cycles $\vartheta_{\tilde{f}}$ associated to a real morsification $\tilde{f}$. The Lagrangian

[2]That is, a compact arboreal Lagrangian skeleta $\mathbb{L} \subseteq (W\ \lambda)$ such that $\partial\mathbb{L} = 0$.

surface $L_{\tilde{f}}$ is also introduced in Sect. 2. Intuitively, in the same manner that $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{\text{st}})$ is a Legendrian approximation of $T_f \subseteq (\mathbb{S}^3, \xi_{\text{st}})$, the exact Lagrangian surfaces $L_{\tilde{f}} \subseteq (\mathbb{C}^2, d\lambda_{\text{st}})$ are Lagrangian analogues of the symplectic Milnor fiber $M_f \subseteq (\mathbb{C}^2, d\lambda_{\text{st}})$. Indeed, $L_{\tilde{f}}$ are smoothly indistinguishable from M_f, and they only become different geometric objects once we incorporate the symplectic structure $(\mathbb{C}^2, d\lambda_{\text{st}})$. Theorem 1.1 is a *relative* statement, being about a Weinstein *pair* $(\mathbb{C}^2, \Lambda_f)$ and not just about a Weinstein 4-manifold. Hence, it is useful in the *absolute* context, as follows. Consider a Legendrian knot $\Lambda \subseteq (\mathbb{S}^3, \xi_{\text{st}})$ in the standard contact 3-sphere and the Weinstein 4-manifold $W(\Lambda) = \mathbb{D}^4 \cup_\Lambda T^*\mathbb{D}^2$ obtained by performing a 2-handle attachment along Λ, i.e. its Weinstein trace. A front projection for Λ (almost) provides an arboreal skeleton for the Weinstein 4-manifold $W(\Lambda)$, as explained in [100]. Nevertheless, the computation of microlocal sheaf invariants from this model is far from immediate, nor exhibits the cluster nature of the moduli space of Lagrangian fillings. The symplectic topology of a Weinstein manifold is much more visible, and invariants more readily computed, from a *closed* arboreal Lagrangian skeleton, i.e. an arboreal Lagrangian skeleton which is compact and without boundary. In particular, Theorem 1.1 provides such a closed Lagrangian skeleton associated to a real morsification:

Corollary 1.2. *Let $f \in \mathbb{C}[x, y]$ define an isolated curve singularity at the origin, $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{st})$ be its associated Legendrian link and $\tilde{f} \in \mathbb{R}[x, y]$ a real morsification. The four-dimensional Weinstein manifold*

$$W(\Lambda_f) = \mathbb{D}^4 \cup_{\Lambda_f} (T^*\mathbb{D}^2 \cup \overset{\pi_0(\Lambda_f)}{\cdots} \cup T^*\mathbb{D}^2))$$

admits the closed arboreal Lagrangian skeleton

$$\mathbb{L}(\tilde{f}) \cup_\partial (\mathbb{D}^2 \cup \overset{\pi_0(\Lambda_f)}{\cdots} \cup \mathbb{D}^2),$$

obtained by attaching the Lagrangian $\mathbb{D}^2$-thimbles of $\tilde{f}$ to the compactified surface $\overline{L}_{\tilde{f}} := L_{\tilde{f}} \cup_\partial (\mathbb{D}^2 \cup \overset{\pi_0(\partial L_{\tilde{f}})}{\cdots} \cup \mathbb{D}^2)$. □

Let us see how Theorem 1.1 and Corollary 1.2 can be applied for two simple singularities, corresponding to the D_5 and the E_6 Dynkin diagrams. As we will see, part of the strength of these results is the explicit nature of the resulting Lagrangian skeleta and the direct bridge they establish between the theory of singularities and symplectic topology.

Example 1.3. (i) First, consider the germ of the D_5-singularity $f(x, y) = xy^2 + x^4$, the Legendrian link associated to this singularity is depicted in Fig. 1 (Left). The Weinstein 4-manifold $W(\Lambda_f) = \mathbb{D}^4 \cup_{\Lambda_f} (T^*\mathbb{D}^2 \cup T^*\mathbb{D}^2)$ admits the closed arboreal Lagrangian skeleton depicted in Fig. 1 (Right). This Lagrangian skeleton is associated to a real morsification $\tilde{f}(x, y) = (x + 1)(4x^3 - 3x + 2y^2 - 1)$ of $f(x, y)$, whose divide $\{(x, y) \in \mathbb{R}^2 : (x + 1)(4x^3 - 3x + 2y^2 - 1) = 0\}$ is depicted in Fig. 4. The D_5-Dynkin diagram is readily seen in the unoriented intersection quiver of the boundaries of the Lagrangian 2-disks added to the (smooth compactification) of the genus 2 Milnor fiber;

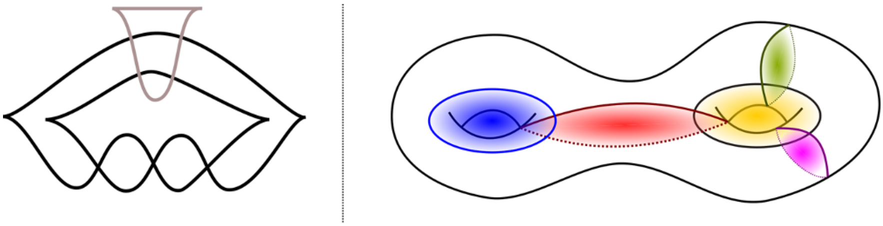


FIGURE 1. The D_5-Legendrian link $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$ (Left) and a closed Lagrangian arboreal skeleton for the Weinstein 4-manifold $W(\Lambda_f)$ (Right), obtained by attaching 5 Lagrangian 2-disks to the cotangent bundle $(T^*\Sigma_2, \lambda_{\mathrm{st}})$

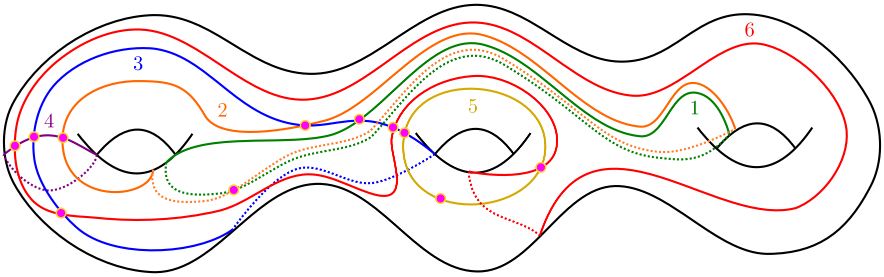


FIGURE 2. Closed Lagrangian arboreal skeleton associated to the simple E_6-singularity $f(x,y) = x^3 + y^4$, according to Corollary 1.2

this unoriented intersection quiver for the vanishing cycles is also drawn in Fig. 4 (Left).

(ii) Second, consider the germ of the singularity $f(x,y) = x^3 + y^4$, the link of the singularity is the maximal-tb positive torus knot $\Lambda_f \cong \Lambda(3,4) \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$. The Weinstein 4-manifold $W(\Lambda_f) = \mathbb{D}^4 \cup_{\Lambda_f} T^*\mathbb{D}^2$ admits the closed arboreal Lagrangian skeleton depicted in Fig. 2. This Lagrangian skeleton is associated to a real morsification $\tilde{f}(x,y) = 4x^3 - 3x + 8y^4 - 8y^2 + 1$ of $f(x,y)$; the Lagrangian skeleton is built by attaching six Lagrangian 2-disks to the Lagrangian zero section Σ_3 of the cotangent bundle $(T^*\Sigma_3, \lambda_{\mathrm{st}})$ of a genus 3 surface. These 2-disks are attached along the six curves in Fig. 2, whose intersection quiver is (mutation equivalent to) the E_6 Dynkin diagram; this unoriented intersection quiver is also drawn in Fig. 4 (Right). See also Fig. 3 for an alternative closed Lagrangian arboreal skeleton, also associated to the simple E_6-singularity $f(x,y) = x^3 + y^4$. □

In the two cases of Example 1.3, the real morsifications can be explicitly obtained using Chebyshev polynomials $T_n(w)$, which are (uniquely) defined by the functional equations $T_n(\cos(t)) = \cos(nt)$, $n \in \mathbb{N} \cup \{0\}$. It can be shown that $T_n(x) + T_m(y)$ is a real morsification of the singularity $f(x,y) = x^n + y^m$ and thus, for example, the expression $T_3(x) + T_4(y) = 4x^3 - 3x + 8y^4 - 8y^2 + 1$ is a real morsification of E_6, as used above and depicted in Fig. 4. In general, we

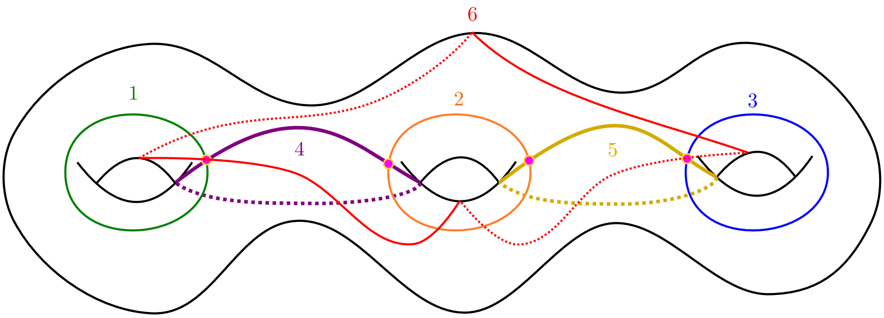


FIGURE 3. Another closed Lagrangian arboreal skeleton for the simple E_6-singularity $f(x,y) = x^3 + y^4$. This is a more symmetric alternative to the closed Lagrangian skeleton in Fig. 2

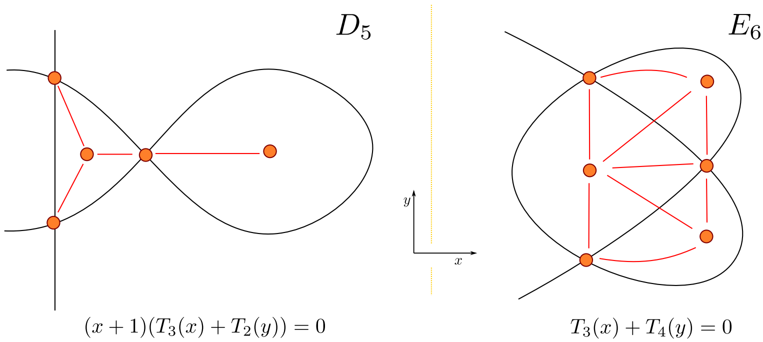


FIGURE 4. The two divides associated to the real morsifications that yield the Lagrangian skeleta in Figs. 1 and 2. The implicit equations for the divides are written in terms of the Chebyshev polynomials $T_n(w)$, determined by the relations $T_n(\cos(t)) = \cos(nt)$. The (unoriented) quivers associated to these two divides are depicted with orange vertices and red edges. Note that the diagram obtained for E_6 is not the E_6 Dynkin diagram; once the quiver is properly oriented, it is mutation equivalent to an orientation of the E_6 Dynkin diagram

will see that the vanishing cycles of a real morsification can be oriented, and then an oriented quiver can be associated to the skew-symmetric intersection form.

From now onward, we abbreviate "closed arboreal Lagrangian skeleton" to *Cal*-skeleton.[3] Let (W, λ) be a Weinstein 4-manifold, e.g. described by a

[3]This seems appropriate, as D. Nadler (UC Berkeley) and L. Starkston (UC Davis), the initial developers of arboreal Lagrangian skeleta, hold their positions in the University of California.

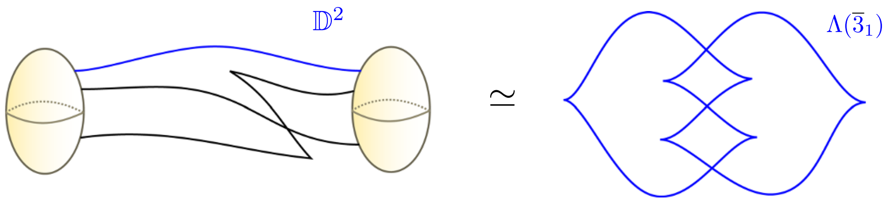


FIGURE 5. Cal-skeleton $\mathbb{RP}^2 \cup_{\mathbb{S}^1} \mathbb{D}^2$ associated to $\Lambda(\overline{3}_1) \subseteq (\partial\mathbb{D}^4, \lambda_{st})$

Legendrian handlebody, a Lefschetz fibration or analytic equations in $\mathbb{C}^n$. There are two basic nested questions: Does it admit a Cal-skeleton? If so, how do you find one ? For instance, consider a max-tb Legendrian representative $\Lambda \subseteq (\partial\mathbb{D}^4, \lambda_{st})$ of any smooth knot, does $W(\Lambda)$ admit a Cal-skeleton ? It might be that not all these Weinstein 4-manifolds $W(\Lambda)$ admit such a skeleton: it is certainly not the case if the Legendrian knot Λ were stabilized, hence the max-tb hypothesis is necessary. In general, the lack of exact Lagrangians in $W(\Lambda)$ would provide an obstruction.

Remark 1.4. For simplicity, we focus on *oriented* exact Lagrangians. Non-orientable Cal-skeleta should also be of interest. For instance, consider the max-tb Legendrian *left*-handed trefoil knot $\Lambda(\overline{3}_1) \subseteq (\partial\mathbb{D}^4, \lambda_{st})$. Figure 5 (Right) depicts a planar front for it. Then the Weinstein 4-manifold $W(\Lambda(\overline{3}_1))$ admits a Cal-skeleton $\mathbb{RP}^2 \cup_{\mathbb{S}^1} \mathbb{D}^2$ given by attaching a Lagrangian 2-disk to a Lagrangian $\mathbb{RP}^2$, as shown in Fig. 5. Indeed, the Weinstein 4-manifold given by Fig. 5 (Left), described by one Weinstein 1-handle and the (black) Weinstein 2-handle passing through it twice, is Weinstein equivalent to the standard cotangent bundle $(T^*\mathbb{RP}^2, \lambda_{st}, \varphi_{st})$, see e.g. [58]. The zero section $\mathbb{RP}^2$ is chosen as its Lagrangian skeleton, and then a Lagrangian 2-disk—core of a Weinstein 2-handle—is attached along the blue circle depicted in the Weinstein handlebody diagram in Fig. 5 (Left). At this stage, we simplify the diagram by handle-sliding the black Legendrian knot along the blue Legendrian boundary of the Lagrangian 2-disk, and then cancel the Weinstein 1-handle with this latter (blue) Weinstein 2-handle; see [21]. This yields a front for the max-tb Legendrian *left*-handed trefoil knot $\Lambda(\overline{3}_1) \subseteq (\partial\mathbb{D}^4, \lambda_{st})$, as required. □

Symplectic invariants of Weinstein 4-manifolds W include (partially) wrapped Fukaya categories [12,101] and categories of microlocal sheaves [80]. Microlocal sheaf invariants should be particularly computable if a Cal-skeleton $\mathbb{L} \subseteq W$ is given, yet worked out examples are scarce in the literature. In Sect. 4, we use[4] Theorem 1.1 to compute the moduli space of simple microlocal sheaves on some of the Cal-skeleta $\mathbb{L}$ from Corollary 1.2.

Finally, Theorem 1.1 provides a context for the study of exact Lagrangian fillings of Legendrian links $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{st})$ associated to isolated

[4]The correspondence [84, Theorem 1.3] and T. Kálmán's description [66] of augmentation varieties $\mathrm{Aug}(\Lambda)$ are also useful tools in this context.

plane curve singularities. Indeed, let

$$\mathbb{L}(\tilde{f}) = L_{\tilde{f}} \cup \vartheta(\tilde{f})$$

be a Cal-skeleton for the Weinstein pair $(\mathbb{C}^2, \Lambda_f)$ for a real morsification $\tilde{f}$, as produced in Theorem 1.1. The exact Lagrangian filling $L_{\tilde{f}}$ may serve as a starting exact Lagrangian filling for the Legendrian link Λ_f, and then performing Lagrangian disk surgeries [96,109] along the Lagrangian thimbles in ϑ is a method to construct additional[5] exact Lagrangian fillings. In general, this strategy might be potentially obstructed, as the Lagrangian disks might acquire immersed boundaries when the Lagrangian surgeries are performed. That said, since Lagrangian disks surgeries yield combinatorial mutations of a quiver, Theorem 1.1 might hint towards a structural conjecture: we expect as many exact Lagrangian fillings Λ_f as elements in the cluster mutation class of the intersection quiver for the vanishing thimbles ϑ. It should be noted that C. Viterbo's work is abundant in useful and remarkable results, but also bountiful in insightful questions and conjectures[6]: trying to follow his steps, Sect. 5 concludes with a discussion on such conjectural matters.

2. Lagrangian skeleta for isolated singularities

In this section we introduce the necessary ingredients for Theorem 1.1 and prove it. We refer the reader to [9,54,75] for the basics of plane curve singularities and [37,38,53,85] for background on 3-dimensional contact topology.

2.1. The legendrian link of an isolated singularity

Let $f \in \mathbb{C}[x, y]$ be a bivariate complex polynomial which defines an isolated complex singularity at the origin $(x, y) = (0, 0) \in \mathbb{C}^2$. The *link of the singularity* $T_f \subseteq (\mathbb{S}^3, \xi_{\text{st}})$ is the intersection

$$T_f = V(f) \cap \mathbb{S}^3_\varepsilon = \{(x, y) \in \mathbb{C}^2 : f(x, y) = 0\} \cap \{(x, y) \in \mathbb{C}^2 : |x|^2 + |y|^2 = \varepsilon\},$$

where $\varepsilon \in \mathbb{R}^+$ is small enough. The intersection is transverse for $\varepsilon \in \mathbb{R}^+$ small enough [31,76], and thus T_f is a smooth link. The link T_f is in fact a transverse link for the contact structure $\xi_{\text{st}} = T\mathbb{S}^3 \cap i(T\mathbb{S}^3)$, as is the boundary of the (Milnor) fiber M_f for the Milnor fibration [53,56]. Equivalently, it is the transverse binding of the contact open book generated by

$$\frac{f}{\|f\|} : \mathbb{S}^3 \backslash T_f \longrightarrow \mathbb{S}^1.$$

The link of a singularity was first introduced by Wirtinger and Brauner [19] and masterfully studied by Milnor [76]. The book [31] comprehensively develops[7] the smooth topology of link of singularities and their connection to 3-manifold topology. The contact topological nature of the associated open book was developed by Giroux [56].

[5]Potentially not Hamiltonian isotopic.

[6]E.g. I recently attended a conference at IMPA where several talks discussed "the Viterbo conjecture". As it turned out, the conjectures the speakers discussed were all different, yet all clearly impactful in their respective areas.

[7]See also W. Neumann's article in Kähler's volume [65].

Let us suppose that the germ of our singularity is irreducible.[8] From a smooth perspective, the smooth isotopy class of T_f is that of an iterated cable of the unknot [31]. Let $K_{l,m}$ be the oriented (l,m)-cable of a smooth link $K \subseteq \mathbb{S}^3$, i.e. an embedded curve in the boundary $\partial\mathcal{O}p(K)$ of the solid torus $\mathcal{O}p(K)$ in the homology class $l \cdot [\lambda] + m \cdot [\mu]$, with λ the longitude and μ the meridian of $\mathcal{O}p(K)$. It is shown in [31, Chapter IV.7] that an iterated cable $K_{(l_1,\mu_1),(l_2,\mu_2),\dots,(l_r,\mu_r)} \subseteq \mathbb{S}^3$ is the link of an isolated singularity if and only if $\mu_{i+1} > (l_i\mu_i)l_{i+1}$, for $1 \le i \le r-1$.

Remark 2.1. Given an isolated singularity $f(x,y)$, there are algorithms for determining the smooth type of T_f, i.e., the sequence of pairs $\{(l_1,\mu_1),(l_2,\mu_2),\dots,(l_r,\mu_r)\}$. For instance, by applying the Newton–Puiseux algorithm to $f(x,y)$ we may write

$$y = a_1x^{\frac{m_1}{n_1}} + a_2x^{\frac{m_2}{n_1n_2}} + a_3x^{\frac{m_3}{n_1n_2n_3}} + \dots, \quad a_i \in \mathbb{C}^*$$

at each branch, where the exponents $m_1/n_1 < m_2/(n_1n_2) < m_3/(n_1n_2n_3) < \cdots$ are increasing and $\gcd(m_i,n_i) = 1$, for all $i \in \mathbb{N}$. The pairs $(n_i,m_i) \in \mathbb{N}^2$ are called the Puiseux pairs. For reference, the Newton pairs are then (p_i,q_i) with $p_i = n_i$, $q_1 = m_1$ and $q_i = m_i - m_{i-1}n_i$ for $i \ge 2$, and the cabling algebraic condition reads $p_i,q_i > 0$. The topological pairs (l_i,μ_i) are given by $l_i = p_i = n_i$, $\mu_1 = q_1$ and $\mu_{i+1} = q_{i+1} + p_ip_{i+1}\mu_i$ for $i \ge 1$, and the cabling algebraic condition translates into $l_i = p_i > 0$ and $q_{i+1} = \mu_{i+1} - l_il_{i+1}\mu_i > 0$, as above. The algorithm and these relations are explained in [31, Appendix to Chapter I]. □

In the finer context of contact topology, the transverse link $T_f \subseteq (\mathbb{S}^3,\xi_{\mathrm{st}})$ is an iterated cable with maximal self-linking number $sl(T_f) = \overline{sl}$, as it bounds the symplectic Milnor fiber $M_f \subseteq \mathbb{C}^2$ of $f \in \mathbb{C}[x,y]$, equiv. the symplectic page of the contact open book [39,56]. By the transverse Bennequin bound [14], this self-linking must be equal to the Euler characteristc $-\chi(M_f)$. A fact about the smooth isotopy class of links of singularities is their Legendrian simplicity:

Proposition 2.2. *Let $f \in \mathbb{C}[x,y]$ define an isolated singularity at the origin and $T_f \subseteq (\mathbb{S}^3,\xi_{st})$ be its associated transverse link. There exists a unique maximal Thurston–Bennequin Legendrian approximation $\Lambda_f \subseteq (\mathbb{S}^3,\xi_{st})$ of the transverse link T_f.*

Proof. The classification of Legendrian representatives of iterated cables of positive torus knots is established in [71, Corollary 1.6], building on [40,41]. The sufficent numerical condition for Legendrian simplicity is $\mu_{i+1}/l_{i+1} > \overline{tb}(K_i)$, where K_i is the ith iterated cable in $K_{(l_1,\mu_1),(l_2,\mu_2),\dots,(l_r,\mu_r)} \subseteq \mathbb{S}^3$. The maximal Thurston-Bennequin equals $\overline{tb}(K_i) = A_i - B_i$, where $A_i, B_i \in \mathbb{N}$ are given by

$$A_i := \sum_{\alpha=1}^{i} p_\alpha \prod_{\beta=\alpha+1}^{i} q_\beta \prod_{\beta=\alpha}^{i} q_\beta, \qquad B_i := \sum_{\alpha=1}^{i}\left(p_\alpha \prod_{\beta=\alpha+1}^{i} q_\beta\right) + \prod_{\alpha=1}^{i} q_\alpha, \quad i \in \mathbb{N},$$

[8] For the general case, we refer the reader to [31] and their splice diagrams.

as defined in [71, Equation (2)], and satisfy $\mu_i l_i > A_i - B_i$. In particular, an algebraic link satisfies $\mu_{i+1}/l_{i+1} > \mu_i l_i > A_i - B_i = \overline{tb}(K_i)$, for all $1 \leq i \leq r-1$, and its max-tb representative is unique. □

Proposition 2.2 implies that there exists a *unique* Legendrian link $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$, up to contact isotopy, whose positive transverse push-off $\tau(\Lambda_f)$, as defined in [53, Section 3.5.3], is transverse isotopic to the transverse link T_f. Note that two distinct Legendrian approximations of a transverse link [35, Theorem 2.1] differ by Legendrian stabilizations, which necessarily decrease the Thurston-Bennequin invariant.

Remark 2.3. Proposition 2.2 does not hold for $K \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$ an arbitrary smooth link. For instance, the smooth isotopy classes of the mirrors $\overline{5}_2, \overline{6}_1$ of the three-twist knot and the Stevedore knot admit *two* distinct maximal-tb Legendrian representatives each [27, Section 4]. That said, the knots $\overline{5}_2, \overline{6}_1$ are not links of singularities, as their Alexander polynomials are not monic, and thus they are not fibered knots [83]. □

Proposition 2.2 allows us to canonically define a *Legendrian* link associated to an isolated singularity:

Definition 2.4. Let f be the germ of an isolated singularity at the origin. A Legendrian link $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$ is associated to f if it is a maximal-tb Legendrian link $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$ whose positive transverse push-off $\tau(\Lambda_f)$ is transversely isotopic to the link of the singularity $T_f \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$. □

Proposition 2.2 shows that the Legendrian isotopy class of a Legendrian link $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$ associated to f is unique. Thus, we refer to $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$ in Definition 2.4 as *the* Legendrian link associated to the germ f.

Example 2.5. (ADE Singularities) Let us consider the three ADE families of simple isolated singularities [11, Chapter 2.5]. Their germs are given by

$$(A_n)\quad f(x,y) = x^{n+1} + y^2, \qquad (D_n)\quad f(x,y) = xy^2 + x^{n-1}, \qquad n \in \mathbb{N},$$
$$(E_6)\quad f(x,y) = x^3 + y^4, \qquad (E_7)\quad f(x,y) = x^3 + xy^3, \qquad (E_8)\quad f(x,y) = x^3 + y^5.$$

The Legendrian link associated to the A_n-singularity is the positive $(2, n+1)$-torus link, with $\overline{tb} = n-1$. These links are associated to the braid σ_1^{n+1}, as depicted in Fig. 6 (Left). The Legendrian link associated to the D_n-singularity is the link consisting of the link associated to the A_{n-3}-singularity and the standard Legendrian unknot, linked as in Fig. 6 (Right). This is the topological consequence of the factorization $f(x,y) = x(y^2 + x^{n-2})$. These D_n-links are associated to the (rainbow closure of the) positive braid $\sigma_1^{n-2}\sigma_2\sigma_1^2\sigma_2$, $n \geq 3$. Each of the three components K_1, K_2, K_3 of the D_2-link is a max-tb Legendrian unknot, with $K_1 \cup K_2$ and $K_2 \cup K_3$ forming each a (max-tb) Hopf link and $K_1 \cup K_3$ forming the 2-unlink. The D_3-link is Legendrian isotopic to the A_3-link, i.e. a max-tb positive $T(2,4)$-torus link.

The Legendrian links associated to the E_6 and E_8 singularities are the maximal-tb positive $(3,4)$-torus Legendrian link and the Legendrian $(3,5)$-torus link, as depicted in Fig. 7. The E_7 is a maximal-tb Legendrian link

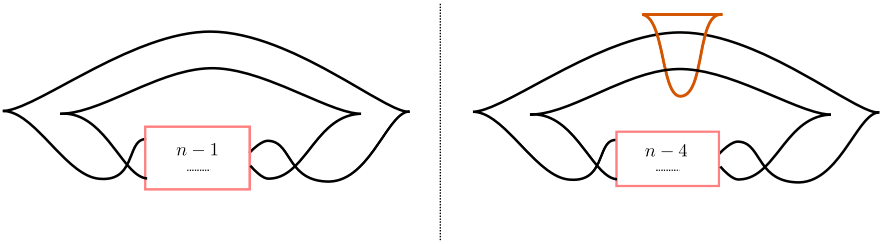


FIGURE 6. The Legendrian link for the A_n-singularity is the max-tb $(2, n+1)$-torus link (Left). The Legendrian link for the D_n-singularity is the link given by the union of a max-tb $(2, n-2)$-torus link and a standard Legendrian unknot, in orange, linked as in the Legendrian front on the right (Right) (colour figure online)

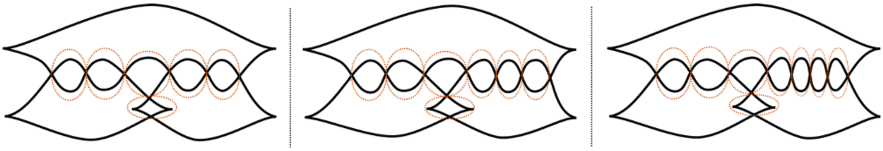


FIGURE 7. The Legendrian links for the E_6, E_7 and E_8 simple singularities

consisting of a trefoil knot and a standard Legendrian unknot, linked as in the center Legendrian front in Fig. 7. This is implied by the $f(x,y) = x(x^2+y^3)$ factorization of the E_7 singularity. The Legendrian links for E_6, E_7 and E_8 can also be obtained as the closure of the three braids $\sigma_1^{n-3}\sigma_2\sigma_1^3\sigma_2$, $n = 6, 7, 8$. Figure 7 also depicts generators of the first homology group of the minimal genus Seifert surface; these generate the first homology of each Milnor fiber, and the E_6, E_7 and E_8 Dynkin diagrams are readily exhibited from their intersection pattern. □

The singularities $f(x,y) = x^a + y^b$, $a \geq 3, b \geq 6$, or $(a,b) = (4,4), (4,5)$, yield an infinite family of non-simple isolated singularities for which the associated Legendrian is readily computed to be the maximal-tb positive (a,b)-torus link, confer Remark 2.1. Two more instances are illustrated in the following:

Example 2.6. (Two Iterated Cables) Consider the isolated curve singularity

$$g(x,y) = x^7 - x^6 + 4x^5y + 2x^3y^2 - y^4.$$

The Puiseux expansion yields the Newton solution $y = x^{3/2}(1 + x^{1/4})$ and thus $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{st})$ is the maximal-tb Legendrian representative of the $(2,13)$-cable of the trefoil knot. This Legendrian knot is depicted in Fig. 8 (Left). The reader is invited to show that the Legendrian knot $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{st})$ of the singularity

$$h(x,y) = x^9 - x^{10} + 6x^8y - 3x^6y^2 + 2x^5y^3 + 3x^3y^4 - y^6,$$

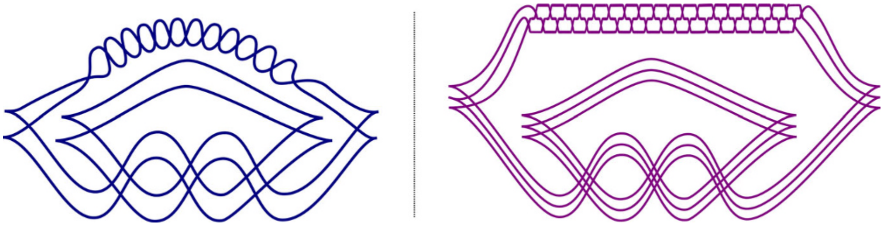


Figure 8. The Legendrian links Λ_g and Λ_h associated to the singularity $g(x,y) = x^7 - x^6 + 4x^5y + 2x^3y^2 - y^4$, on the left, and the singularity $h(x,y) = x^9 - x^{10} + 6x^8y - 3x^6y^2 + 2x^5y^3 + 3x^3y^4 - y^6$, on the right

is the maximal-tb Legendrian representative of the $(3,19)$-cable of the trefoil knot [54], as depicted in Fig. 8 (Right). (For that, start by writing the relation as $y(x) = x^{3/2} + x^{5/3}$.) $\square$

2.2. A'Campo's divides and their conormal lifts

Let $f \in \mathbb{C}[x,y]$ define an isolated singularity at the origin, $\mathbb{D}^4_\varepsilon \subseteq \mathbb{C}^2$ be a Milnor ball for this singularity [75, Corollary 4.5], $\varepsilon \in \mathbb{R}^+$, $\mathbb{R}^2 = \{(x,y) \in \mathbb{C}^2 : \Im(x) = 0, \Im(y) = 0\} \subseteq \mathbb{C}^2$ the real 2-plane, and $\mathbb{D}^2_\varepsilon = \mathbb{D}^4_\varepsilon \cap \mathbb{R}^2$ a real Milnor 2-disk. First, we need the notion of a *divide*, called *partage* in [2], as follows:

Definition 2.7. [2] Let $\mathbb{D}^2_\varepsilon \subseteq \mathbb{R}^2$ be the 2-disk of radius $\varepsilon \in \mathbb{R}^+$. A divide is a proper generic immersion $\gamma : I \longrightarrow \mathbb{D}^2$ of a 1-manifold I into $\mathbb{D}^2$. $\square$

The image $\gamma(I) \subseteq \mathbb{D}^2_\varepsilon$ is also referred to as a divide, in a slight abuse of notation. Definition 2.7 belongs to the realm of real differential topology. A remarkable fact is that A'Campo explained how to associate a divide to certain real morsifications of a singularity. For that, consider a real morsification $\tilde{f}_t(x,y)$, $t \in [0,1]$, such that, for $t \in (0,1]$, $f_t(x,y)$ has only A_1-singularities, its critical values are real and the level set $\tilde{f}_t^{-1}(0) \cap \mathbb{D}^4_\varepsilon$, contains all the saddle points of the restriction $(\tilde{f}_t)|_{\mathbb{D}^2_\varepsilon}$. Then, the intersection $D_{\tilde{f}} := \tilde{f}^{-1}(0) \cap \mathbb{D}^2_\varepsilon \subseteq \mathbb{R}^2$, where $\tilde{f} = f_1$, is a divide, and it is known as the *divide* of the real morsification $\tilde{f}_t$ [3,9,63].

Let us denote by D_f a divide $D_{\tilde{f}}$ obtained from a real morsification $\tilde{f}_t$ of f. A divide D_f is also referred to as an A'Campo divide for the singularity f. As in Definition 2.7, it is the image of a union of a smooth 1-manifold I under an immersion $i : I \longrightarrow \mathbb{R}^2$ [55,62,64], and it is a generic such immersion. In this manuscript, we assume that the germs of singularities that we consider admit such real morsifications. See [2,61] for the existence and details of real morsifications, and see Fig. 4 for divides associated to real morsifications of the simple singularities D_5 and E_6.

Let us now move towards contact topology. By considering a divide $D_f \subseteq \mathbb{R}^2$ as a wavefront co-oriented in both conormal directions, its (biconormal) Legendrian lift is a Legendrian link $\Lambda_0(D_f)$ in the (ideal) contact boundary $(\partial(T^*\mathbb{R}^2), \lambda_{\mathrm{st}}|_{\partial(T^*\mathbb{R}^2)})$. In this case, $(\partial(T^*\mathbb{R}^2), \lambda_{\mathrm{st}}|_{\partial(T^*\mathbb{R}^2)})$ is considered with its Legendrian projection onto the zero section $\partial(T^*\mathbb{R}^2) \longrightarrow \mathbb{R}^2$, who fibers are Legendrian 1-spheres $\mathbb{S}^1 \subseteq \partial(T^*\mathbb{R}^2)$. See [8, Section 3.1] for fronts and Legendrian fibrations and, e.g. [97, Section 2] and [53, Section 3.2].

The biconormal lift $\Lambda_0(D_f) \subseteq \partial(T^*\mathbb{R}^2)$ of the immersed curve D_f to the (unit) boundary of the cotangent bundle $T^*\mathbb{R}^2$ can be constructed using the three local models:

(i) The biconormal lift near a smooth interior point $P \in D_f$ is defined as

$$\{u = (q, u_q) \in T^*\mathcal{O}p(P) : \|u_q\| = 1, T_qD_f \subseteq \ker(u_q) \text{ for } q \in D_f \cap \mathcal{O}p(P)\},$$

for an arbitrary fixed choice of metric in $\mathbb{R}^2$, and neighborhood $\mathcal{O}p(P) \subseteq \mathbb{R}^2$. See the first row of Fig. 9.

(ii) The biconormal lift near an immersed point $P \in D_f$ is defined as the (disjoint) union of the conormal lifts of each of its embedded branches through P. See the second row of Fig. 9.

(iii) Finally, at the endpoint $P \in D_f$, the biconormal lift is defined as the closure in $T^*_P\mathbb{R}^2$ of one of the components of

$$T^*_P\mathbb{R}^2 \setminus \{u \in T^*_P\mathbb{R}^2 : \|u_q\| = 1, T_PD_f \subseteq \ker(u_q) \text{ for } q \in D_f \cap \mathcal{O}p(P)\},$$

where the tangent line T_PD_f is defined as the (ambient) smooth limit of the tangent lines $T_{q_i}D_f$ for a sequence $\{q_i\}_{i\in\mathbb{N}}$ of interior points $q_i \in D_f$ convering to $P \in D_f$. There are two such components, but our arguments are independent of such a choice. See the third row of Fig. 9.

Remark 2.8. The restriction of the canonical projection $\pi : \partial(T^*\mathbb{R}^2) \longrightarrow \mathbb{R}^2$ is finite two-to-one onto the image of the interior points of I. The pre-image of π at (the image of) endpoints contains an open interval of the Legendrian circle fiber. For instance, the full conormal lift of a point $p \in \mathbb{R}^2$ is Legendrian isotopic to the zero section $\mathbb{S}^1 \subseteq (J^1\mathbb{S}^1, \xi_{\mathrm{st}})$, as is the conormal lift of an embedded closed segment. □

These local models define the Legendrian biconormal lift $\Lambda_0(D_f) \subseteq (\partial(T^*\mathbb{R}^2), \xi_{\mathrm{st}})$ of the divide of the Morsification $\tilde{f}$. Let $\iota_0 : \mathbb{S}^1 \longrightarrow (\mathbb{S}^3, \xi_{\mathrm{st}})$ be a Legendrian embedding in the isotopy class of the standard Legendrian unknot. A small neighborhood $\mathcal{O}p(\iota(\mathbb{S}^1))$ is contactomorphic to the 1-jet space $(J^1\mathbb{S}^1, \xi_{\mathrm{st}}) \cong (T^*\mathbb{S}^1 \times \mathbb{R}_t, \ker\{\lambda_{\mathrm{st}} - dt\})$, yielding a contact inclusion $\iota : (J^1\mathbb{S}^1, \xi_{\mathrm{st}}) \longrightarrow (\mathbb{S}^3, \xi_{\mathrm{st}})$. Note that there exists a contactomorphism $\Psi : (\partial(T^*\mathbb{R}^2), \xi_{\mathrm{st}}) \longrightarrow (J^1\mathbb{S}^1, \xi_{\mathrm{st}})$, where the zero section in the 1-jet space bijects to the Legendrian boundary of a Lagrangian cotangent fiber in $T^*\mathbb{R}^2$. This leads to the following:

Definition 2.9. Let $D_f \subseteq \mathbb{R}^2$ be the divide associated to a real morsification of a germ f defining an isolated singularity. The biconormal lift $\Lambda(D_f) \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$ is the image $\iota(\Psi(\Lambda_0(D_f)))$. That is, the biconormal lift $\Lambda(D_f) \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$ is the satellite of the biconormal lift $\Lambda_0(D_f) \subseteq (\partial(T^*\mathbb{R}^2), \xi_{\mathrm{st}})$ with companion knot the standard Legendrian unknot in $(\mathbb{S}^3, \xi_{\mathrm{st}})$. □

The central result in N. A'Campo's articles [3,4] is that the Legendrian link $\Lambda(D_f) \subseteq S^3$ is *smoothly* isotopic to the transverse link T_f, see also [64]. The formulation above, in terms of the satellite to the Legendrian unknot,

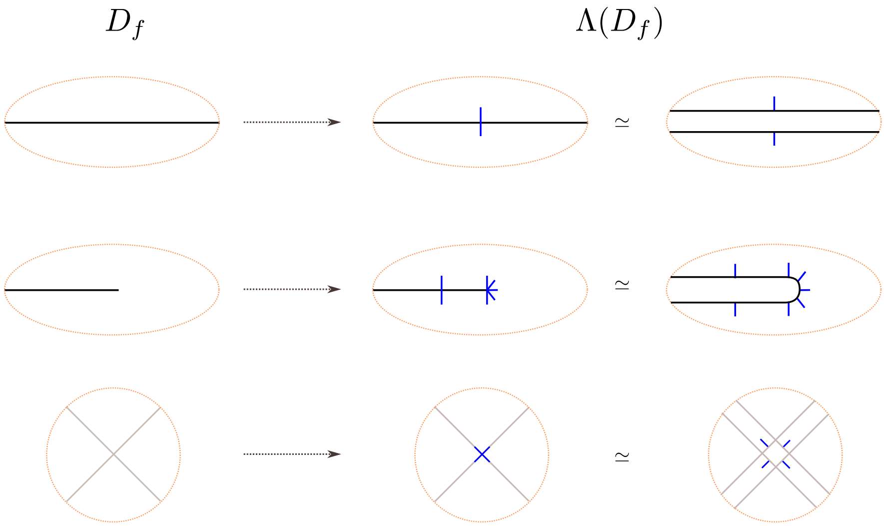


FIGURE 9. Local models for the divides D_f, on the left column, and their corresponding biconormal lifts, on the right column. Note that we have depicted the biconormal lift in its non-generic form (matching D_f at the boundary), at the left of the right column, and also after a Legendrian front perturbation, at the right of the right column. The local model of the crossing is depicted in gray so that the conormal direction (in blue) is visible (colour figure online)

is not necessarily explicit in the literature on divides and their Legendrian lifts, but probably known to the experts, as it is effectively being used in Hirasawa's visualization [62, Figure 2]. See also the work of Kawamura [70, Figure 2], Ishikawa and Gibson [55,63] and others [26,64]. The phrasing in Definition 2.9 might help crystallize the contact topological characteristics of each object.

Example 2.10. (i) The A_1-singularity admits two real morsifications $\tilde{f}_1(x,y) = x^2+y^2-1$ and $\tilde{f}_2(x,y) = x^2-y^2$, with corresponding divides

$$D_1 = \{(x,y)\in\mathbb{R}^2 : x^2+y^2-1=0\}, \quad D_2 = \{(x,y)\in\mathbb{R}^2 : x^2-y^2=0\}.$$

The biconormal lift $\Lambda_0(D_1) \subseteq (\partial(T^*\mathbb{R}^2),\xi_{\mathrm{st}})$ consists of two copies of the Legendrian fibers of the fibration $\pi : \partial(T^*\mathbb{R}^2) \longrightarrow \mathbb{R}^2$. Each of these two copies is satellited to the standard Legendrian unknot, forming a maximal-tb Hopf link $\Lambda(D_1) \subseteq (\mathbb{S}^3,\xi_{\mathrm{st}})$. Indeed, the second Legendrian fiber can be assumed to be the image of the first Legendrian fiber under the Reeb flow. Hence, the Legendrian link $\Lambda(D_1) \subseteq (\mathbb{S}^3,\xi_{\mathrm{st}})$ must consist of the standard Legendrian unknot union a small Reeb push-off. Similarly, the biconormal lift $\Lambda_0(D_2) \subseteq (\partial(T^*\mathbb{R}^2),\xi_{\mathrm{st}})$ equally consists of two copies of the Legendrian fibers of the fibration $\pi : \partial(T^*\mathbb{R}^2) \longrightarrow \mathbb{R}^2$,

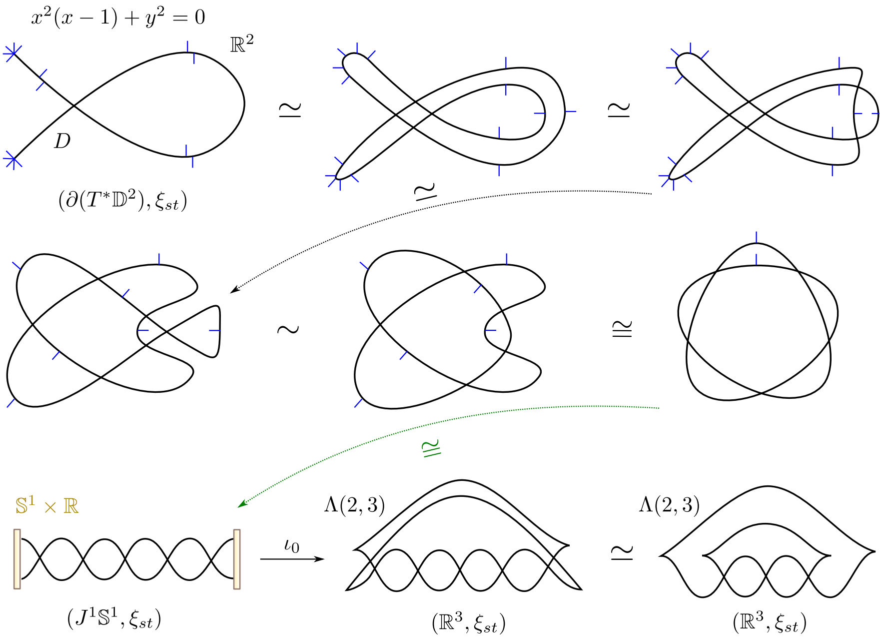


FIGURE 10. A co-oriented divide D for the A_2-singularity $f(x,y) = x^3 + y^2$, as a front for its Legendrian link $\Lambda(D) \subseteq (\partial(T^*\mathbb{D}^2), \xi_{\mathrm{st}})$. That is, the biconormal lift of D is $\Lambda(D)$. Its satellite along the standard unknot is the (unique) max-tb Legendrian trefoil $\Lambda(2,3) \subseteq (\mathbb{R}^3, \xi_{\mathrm{st}})$

and thus both Legendrian links $\Lambda(D_1), \Lambda(D_2)$ are Legendrian isotopic in $(\mathbb{S}^3, \xi_{\mathrm{st}})$.

(ii) The A_2-singularity $f(x,y) = x^3 + y^2$ admits the real morsification $\tilde{f}(x,y) = x^2(x-1) + y^2$, whose divide is $D = \{(x,y) \in \mathbb{R}^2 : x^2(x-1) + y^2 = 0\}$. The divide $D \subseteq \mathbb{R}^2$ with its co-orientations is depicted in Fig. 10 (upper left). It depicts a wavefront homotopy, which yields a Legendrian isotopy in $(\partial(T^*\mathbb{R}^2), \xi_{\mathrm{st}})$, and an additional move equivalence (as in [47, Definition 8.2]). In the first row, the first move separates the two conormals pictorially and the second move is a Reidemeister II, i.e. a safe (non-dangerous) self-tangency. The transition to the second row starts with a Reidemeister III move, which is a front homotopy. The first move in the second row is undoing the kink, also known as a U-turn—see [47, Figure 30]—and the second is a planar isotopy. Finally, the third row starts by depicting the change of front projections induced by the contactomorphism Ψ, and performs the satellite to the standard Legendrian unknot. The resulting Legendrian $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$ is the max-tb Legendrian trefoil knot $\Lambda(2,3)$ presented in one of its common fronts for $(\mathbb{R}^3, \xi_{\mathrm{st}})$. $\square$

Remark 2.11. In general, divides for A_n-singularities are depicted in [47, Figure 4]. We invite the reader to study the A_5-singularity $f(x,y) = x^5 + y^2$ with its divide

$$D = \{(x,y) \in \mathbb{R}^2 : x^2(x^3 + x^2 - x - 1) + y^2 = 0\}$$

and discover the corresponding Legendrian isotopy, as in Fig. 10. The isotopy should end with the max-tb Legendrian link $\Lambda(2,5) \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$, e.g. expressed as the (rainbow) closure of the positive braid σ_1^5, equiv. the (-1)-framed closure of σ_1^7. The general case $n \in \mathbb{N}$ is similar. □

Before we proceed with the proof of Theorem 1.1, we note the following contact topological property for the Legendrian links $\Lambda(D_{\tilde{f}})$ associated to divides of real morsifications $\tilde{f}$:

Proposition 2.12. *Let $f \in \mathbb{C}[x,y]$ define an isolated singularity, $D_f \subseteq \mathbb{R}^2$ be the divide associated to a real morsification and $\Lambda(D_f) \subseteq (\mathbb{S}^3, \xi_{st})$ its biconormal lift. Then $\Lambda(D_f)$ admits an embedded exact Lagrangian filling in $(\mathbb{D}^4, \lambda_{st})$. In particular, the Thurston-Bennequin invariant of $\Lambda(D_f)$ is maximal.*

Proof. Consider the plabic graph associated to the divide D_f as in [47, Definition 6.11] and note that the alternating strand diagram associated to a plabic graph is Legendrian isotopic to $\Lambda(D_f)$. Indeed, they only differ by U-turns, at the boundary endpoints, and safe tangencies [47, Section 8] at the interior crossings. Now, from a smooth perspective, we can consider the Goncharov-Kenyon conjugate surface [59, Section 2.1] associated to this plabic graph, which bounds its alternating strand diagram. Thus, this is a smooth embedded surface in $\mathbb{S}^3$ bounding $\Lambda(D_f) \subseteq \mathbb{S}^3$ which can be pushed into an embedded surface $\mathbb{D}^4$, relative to the boundary. In short, the conjugate surface is a smooth surface filling for $\Lambda(D_f)$. This surface can be turned in an embedded exact Lagrangian, as done in [98, Proposition 4.9], which proves the first statement. The statement on the Thurston-Bennequin invariant follows from [24, Theorem 1.4]. □

Figure 11 depicts a piece of such a Lagrangian filling near a crossing of the divide. See [98, Section 4] and [47, Section 6] for further details on the construction. Observe that the plabic graph associated to D_f is not unique, e.g. it is possible to perform a square move at each crossing. The Hamiltonian isotopy of the Lagrangian filling, relative to the boundary, does typically depend on this choice and one should expect to build more than one Hamiltonian isotopy class of Lagrangian fillings with the method of Proposition 2.12.[9]

2.3. Proof of Theorem 1.1

There is an interesting dissonance at this stage. The Legendrian link $\Lambda(D_f) \subseteq \mathbb{S}^3$ in Definition 2.9 and the transverse link $T_f \subseteq \mathbb{S}^3$ of the singularity are smoothly isotopic, yet certainly *not* contact isotopic. Their relationship is described by the following:

[9] Naively applied, this method seems to yield finitely many possible Hamiltonian isotopy classes of Lagrangian fillings. Note that we have proven in [20] that most max-tb Legendrian algebraic links admit infinitely many such classes.

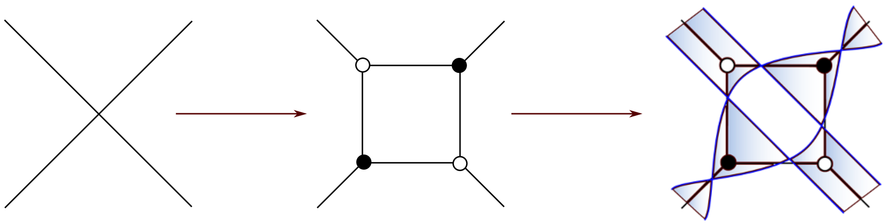


FIGURE 11. A local depiction of the (Lagrangian) conjugate surface near a crossing of the divide (Right). The surface is depicted in darker blue, and it bounds a front, in blue, for the Legendrian link. The plabic graph associated to a crossing (Left) is shown the center. Note that there are two choices of (bi)coloring for the vertices, and the two surfaces differ by a square move, i.e., a Lagrangian mutation; both such choices yield embedded exact Lagrangian fillings (though not necessarily in the same Hamiltonian isotopy class)

Proposition 2.13. *Let $f \in \mathbb{C}[x,y]$ define an isolated singularity and $D_f \subseteq \mathbb{R}^2$ be the divide associated to a real morsification. The positive transverse push-off $\tau(\Lambda(D_f)) \subseteq (\mathbb{S}^3, \xi_{st})$ of the Legendrian link $\Lambda(D_f)$ is contact isotopic to the transverse link $T_f \subseteq (\mathbb{S}^3, \xi_{st})$. In particular, $\Lambda(D_f) \subseteq (\mathbb{S}^3, \xi_{st})$ is Legendrian isotopic to the Legendrian link $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{st})$ associated to the isolated singularity of $f \in \mathbb{C}[x,y]$.* □

Proof. First, we note that $\Lambda(D_f)$ is a maximal-tb Legendrian representative by Proposition 2.12. Thus the latter part of statement follows from the former and Proposition 2.2. Hence we now focus on the first part of the statement. In A'Campo's isotopy [3, Section 3] from the link associated to the divide to the link of the singularity, the key step is the *almost complexification* of the Morsification $\tilde{f} : \mathbb{R}^2 \longrightarrow \mathbb{R}$. This replaces the $\mathbb{R}$-valued function $\tilde{f}$ by an expression of the form

$$\tilde{f}_{\mathbb{C}} : T^*\mathbb{R}^2 \longrightarrow \mathbb{C}, \quad \tilde{f}_{\mathbb{C}}(x,u) := \tilde{f}(x) + id\tilde{f}(x)(u) - \frac{1}{2}\chi(x)H(f(x))(u,u),$$

which is a $\mathbb{C}$-valued function, where $u = (u_1, u_2) \in \mathbb{R}^2$ are Cartesian coordinates in the fiber. Here $H(f(x))$ is the Hessian of f, which is a quadratic form, and $\chi(x)$ is a bump function with $\chi(x) \equiv 1$ near double-points of the divide $D_f \subseteq \mathbb{R}^2$ and $\chi(x) \equiv 0$ away from them. The results in [3], see also [63,64], imply that the transverse link of the singularity is isotopic to the intersection $\partial_\varepsilon(T^*\mathbb{R}^2) \cap \widetilde{f}_{\mathbb{C}}^{-1}(0) \subseteq (\partial_\varepsilon(T^*\mathbb{R}^2), \xi_{\mathrm{st}})$ of the ε-unit cotangent bundle with the 0-fiber of $\widetilde{f}_{\mathbb{C}}$, $\varepsilon \in \mathbb{R}^+$ small enough.[10] It thus suffices to compare this transverse link to the Legendrian lift $\Lambda(D_f) \subseteq (\partial_\varepsilon(T^*\mathbb{R}^2), \xi_{\mathrm{st}})$, which we can check in each of the two local models: near a smooth interior point of the divide D_f and near each of its double points. Note that the case of boundary

[10]This mimicks S. Donaldson's construction of Lefschetz pencils, where the boundary of a fiber is a transverse link at the boundary, see also E. Giroux's construction of the contact binding of an open book [56,57].

points can be perturbed to that of smooth interior points, as in the second row of the local models depicted in Fig. 9 or the first perturbation in Fig. 10. We detail the computation in the first local model, the case of double points follows similarly.

The contact structure $(\partial_\varepsilon(T^*\mathbb{R}^2), \xi_{\rm st})$ admits the contact form $\xi_{\rm st} = \ker\{\cos(\theta)dx_1 - \sin(\theta)dx_2\}$, $(x_1, x_2) \in \mathbb{R}^2$ and $\theta \in \mathbb{S}^1$ is a coordinate in the fiber – this is the angular coordinate in the (u_1, u_2)-coordinates above. The divide can be assumed to be cut locally by $D = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 = 0\} \subseteq \mathbb{R}^2$, as we can write $\tilde{f}(x_1, x_2) = x_2$, and thus its bi-conormal Legendrian lift is

$$\Lambda(D) = \{(x_1, x_2, \theta) \in \mathbb{R}^2 \times \mathbb{S}^1 : x_2 = 0, \theta = \pm\pi/2\}.$$

Note that the tangent space $T_{(x_1,x_2)}\Lambda(D)$ of $\Lambda(D)$ is spanned by ∂_{x_1}, which satisfies

$$\langle \partial_{x_1} \rangle = \ker\{\cos(\theta)dx_1 - \sin(\theta)dx_2\}, \text{ as } \cos(\theta) = 0 \text{ at } \theta = \pm\pi/2.$$

Since the model is away from a double point, $\tilde{f}_{\mathbb{C}}(x, u) := x_2 + i(0, 1) \cdot (u_1, u_2)^t = x_2 + iu_2$ becomes the standard symplectic projection $\mathbb{R}^2 \times \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ onto the second (symplectic) factor. The zero set is thus $x_2 = 0$ and $u_2 = 0$ and so the intersection with $T^\varepsilon\mathbb{R}^2$ is

$$\kappa = \{(x_1, x_2, \theta) \in \mathbb{R}^2 \times \mathbb{S}^1 : x_2 = 0, \theta = 0, \pi\},$$

as the points with $|u_1|^2 = \varepsilon$ are at θ-coordinates $\theta = 0, \pi$. The tangent space $T\kappa = \langle \partial_{x_1} \rangle$ is spanned by ∂_{x_1}, which is transverse to the contact structure along κ:

$$(\cos(\theta)dx_1 - \sin(\theta)dx_2)(\partial_{x_1}) = \pm 1, \quad \text{at } \theta = 0, \pi.$$

It evaluates positive for $\theta = 0$ and negative for $\theta = \pi$, which corresponds to each of the two branches in the biconormal lift. It is readily verified [53, Section 3.1] that κ is the transverse push-off, positive *and* negative,[11] of $\Lambda(D)$, e.g. observe that the annulus $\{(x_1, x_2, \theta) \in \mathbb{R}^2 \times \mathbb{S}^1 : x_2 = 0, 0 \le \theta \le \pi\}$ is a (Weinstein) ribbon for the Legendrian segment $\{(x_1, x_2, \theta) \in \mathbb{R}^2 \times \mathbb{S}^1 : x_2 = 0, \theta = \pi/2\}$. □

Proposition 2.13 implies that real morsifications $\tilde{f}$ yield models for the Legendrian link $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{\rm st})$ of a singularity $f \in \mathbb{C}[x, y]$, as introduced in Definition 2.4. That is, given an isolated plane curve singularity $f \in \mathbb{C}[x, y]$, the Legendrian link $\Lambda_f \subseteq (\mathbb{S}^3, \xi_{\rm st})$ is Legendrian isotopic to the Legendrian lift $\Lambda(D_{\tilde{f}}) \subseteq (\mathbb{S}^3, \xi_{\rm st})$ of a divide $D_{\tilde{f}} \subseteq \mathbb{R}^2$ of a real morsification, and thus we now directly focus on studying the Legendrian links $\Lambda(D_{\tilde{f}}) \subseteq (\mathbb{S}^3, \xi_{\rm st})$.

Let us now prove Theorem 1.1. For that, we use N. A'Campo's description [4] of the set of vanishing cycles associated to a divide of a real morsification. For each double point $p_i \in D$ in the divide $D := D_{\tilde{f}}$, there is a vanishing cycle ϑ_{p_i}. For each bounded region of $\mathbb{R}^2 \setminus D$, which we label by q_j, there is a vanishing cycle ϑ_{q_j}. These vanishing cycles are also naturally oriented by choosing the counter-clockwise orientation in the plane. First, we

[11] The orientation for the negative branch is reversed when considering the global link κ.

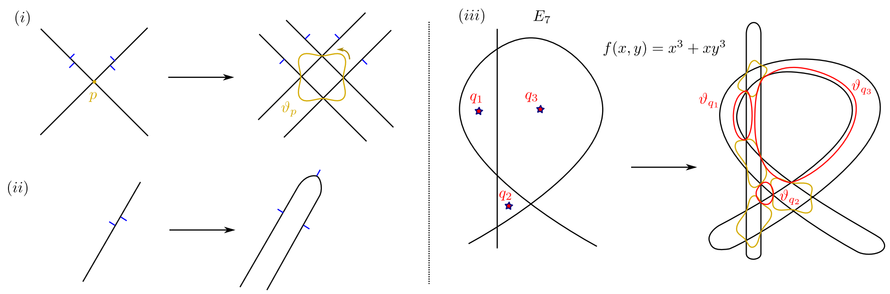


FIGURE 12. (Left) Two front homotopies from the pieces of a divide to a (generic) Legendrian front, in line with the local models in Fig. 9. The vanishing cycle ϑ_p is drawn in the Lagrangian base $\mathbb{R}^2$. (Right) A perturbation of a divide for the E_7-singularity. The vanishing cycles ϑ_p coming from the double points of the divide are drawn in yellow, and the vanishing cycles ϑ_q coming from each of the three bounded interior regions are drawn in red (colour figure online)

visualize those vanishing cycles by perturbing the divide $D \subseteq \mathbb{R}^2$ using the local models in Fig. 9, e.g., as depicted in Fig. 12.(i) and (ii). Let us denote this perturbed cooriented front by $D' \subseteq \mathbb{R}^2$, and note that D' only uses one conormal direction at a given point. This perturbation is a front homotopy from $\Lambda(D_{\tilde{f}})$ and thus produces a Legendrian isotopy of the associated Legendrian links $\Lambda(D_{\tilde{f}}) \cong \Lambda_f$ in $(\mathbb{S}^3, \xi_{\text{st}})$.

Once the perturbation has been performed, we can draw the curves $\vartheta_{p_i}, \vartheta_{q_j}$ as in Fig. 12. For instance, Fig. 12.(iii) depicts the case of the E_7-singularity with a particular choice of divide D and its perturbation D', with ϑ_{p_i} in yellow and ϑ_{q_j} in red. That is:

1. For each double point $p_i \in D$, i.e. a crossing, the curve ϑ_{p_i} is a closed simple curve through the four new double points in D',
2. For each closed region, ϑ_{q_j} is a simple closed curve which (exactly) passes through the double points at the perturbed boundary in D' of the region q_j.

The algorithm in [4] constructs a model for the topological Milnor fiber of f by using the real morsification $\tilde{f}$, as follows. First, start with the conical Lagrangian conormal $L(D') \subseteq (T^*\mathbb{R}^2, \lambda_{\text{st}})$ of the perturbed divide D'. This Lagrangian conormal intersects the unit cotangent bundle of $T^*\mathbb{R}^2$ at $\Lambda(D')$ and thus, being conical, the information of $L(D')$ is equivalent to the information of the Legendrian link $\Lambda(D') \subseteq (\partial(T^*\mathbb{R}^2), \lambda_{\text{st}}|_{\partial(T^*\mathbb{R}^2)})$ with its front $D' \subseteq \mathbb{R}^2$. The intersection $L(D') \cap \mathbb{R}^2 = D'$ with the zero section $\mathbb{R}^2 \subseteq T^*\mathbb{R}^2$ is the divide D'. Second, consider the bounded regions in $\mathbb{R}^2 \setminus D'$ which are *not* enclosed by either of the curves of type $\vartheta_{p_i}, \vartheta_{q_j}$, described in (1) and (2) above. These are the bounded regions in $\mathbb{R}^2 \setminus D'$ which do *not* come from a bounded square obtained by resolving a crossing (as in Fig. 9) nor from a bounded region in $\mathbb{R}^2 \setminus D$. Each of these regions is represented by an

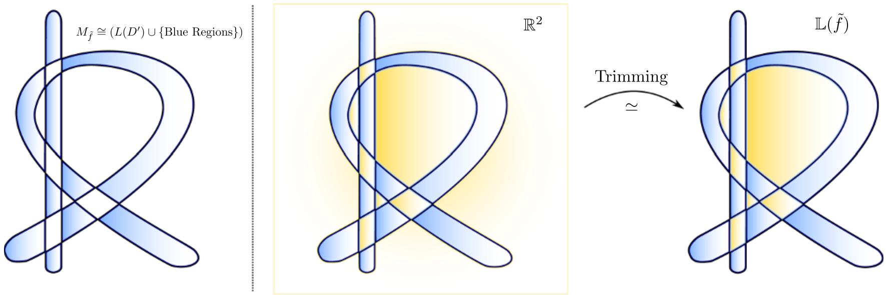


FIGURE 13. (Left) A Lagrangian model for the Milnor fiber of E_7 using the biconormal lift $L(D)'$ and some of the bounded regions in the zero section $\mathbb{R}^2 \subseteq (T^*\mathbb{R}^2, \lambda_{\text{st}})$, filled in blue. (Right) The Lagrangian skeleton $L(D') \cup \mathbb{R}^2$ previous to trimming the unbounded region (also depicted in yellow) and the result of applying a holonomy homotopy, where the unbounded region is trimmed to $\mathbb{L}(\tilde{f})$ (colour figure online)

embedded (exact) Lagrangian 2-disk, as they are contained in the Lagrangian zero section $(T^*\mathbb{R}^2, \lambda_{\text{st}})$. The topological surface obtained as the union of the Lagrangian conormal $L(D')$ with these Lagrangian 2-disks is a surface (with corners) which, upon smoothing, lies in the same smooth isotopy class of the Milnor fiber of f. This explains, following [4], that the union of the Lagrangian $L(D')$ with certain bounded Lagrangian regions in $\mathbb{R}^2 \setminus D'$ is a model for the topological Milnor fiber.

Remark 2.14. For instance, in the example depicted in Fig. 12 (right), there are 10 such regions in $\mathbb{R}^2 \setminus D'$ out of 17. We have depicted these regions in blue in Fig. 13 (left). Note that there are four crossings in D and three bounded regions in $\mathbb{R}^2 \setminus D$. The union of these 10 regions with $L(D')$ yields a topological surface of genus 4 and 2 boundary components – those of the 2-component link $\Lambda(D_f)$. Its first Betti number indeed matches $\mu(E_7) = 7$. $\square$

In addition to the above model for the Milnor fiber, the article [4] also guarantees that the curves $\vartheta_{p_i}, \vartheta_{q_j}$ are vanishing cycles for the real morsification $\tilde{f}$. At this stage, the key fact that we use from A'Campo's algorithm is that our choice of immersion of the divide $D' \subseteq \mathbb{R}^2$, given by the perturbation, exhibits Lagrangian 2-disks $\mathbb{D}^2_{p_i}, \mathbb{D}^2_{q_j} \subseteq \mathbb{R}^2$ such that $\partial\mathbb{D}^2_{p_i} = \vartheta_{p_i}$ and $\partial\mathbb{D}^2_{q_j} = \vartheta_{q_j}$. The union of all these Lagrangian 2-disks $\mathbb{D}^2_{p_i}, \mathbb{D}^2_{q_j}$ constitutes the set $\mathscr{T}(\vartheta_{\tilde{f}})$ of Lagrangian $\mathbb{D}^2$-thimbles in the statement of Theorem 1.1.

For the curves ϑ_{p_i}, this follows from Fig. 12.(i), or Fig. 9, where the 2-disk $\mathbb{D}^2_{p_i}$ is (a small extension of) the square given by the four double points in D' appearing in the perturbation of $p_i \in D$. For ϑ_{q_j}, the 2-disk $\mathbb{D}^2_{q_j}$ is chosen to be a small extension of the bounded region itself. These disks are (exact) Lagrangian because $\mathbb{R}^2 \subseteq (T^*\mathbb{R}^2, \lambda_{\text{st}})$ is exact Lagrangian. The Liouville vector field in $(T^*\mathbb{R}^2, \lambda_{\text{st}})$ vanishes at $\mathbb{R}^2$ and is tangent to $L(D')$. Hence, the inverse flow of the Liouville field retracts the Weinstein pair $(\mathbb{R}^4, \Lambda(D'))$ to

$L(D')$ union the zero section $\mathbb{R}^2$. This shows that $L(D')\cup\mathbb{R}^2$ is a Lagrangian skeleton of the Weinstein pair $(\mathbb{R}^4,\Lambda(D'))$. Figure 13 depicts this skeleton in its center, where the $\mathbb{R}^2$ is included in its entirety.

Now, the Lagrangian skeleton has an open piece at the unbounded part of $\mathbb{R}^2$. To complete our argument, it suffices to homotope the Lagrangian skeleton so that the unbounded part is trimmed to match the boundary $\mathscr{B}$ of the unbounded piece of $\mathbb{R}^2\setminus D'$. These skeletal modifications are explained in detail in [100, Section 3]. In a nutshell, one applies the holonomy modifications from [28, Section 12] to homotope the boundary at infinity of $\mathbb{R}^2$ until it coincides with $\mathscr{B}$, modifying the pseudo-gradient field accordingly and producing a Weinstein homotopy. In conclusion, the union of the conical Lagrangian $L(D')$, some bounded regions[12] of $\mathbb{R}^2\setminus D'$, and the Lagrangian 2-disks $\mathbb{D}^2_{p_i},\mathbb{D}^2_{q_j}\subseteq\mathbb{R}^2$ forms a Lagrangian skeleton of the Weinstein pair $(\mathbb{R}^4,\Lambda(D'))$, as required. □

Remark 2.15. The referee also suggested the following (equivalent) viewpoint to smoothly construct the Milnor fiber, which can also be helpful. Consider the bipartite vertices of the $A\Gamma$-diagram [47, Definition 3.1] associated to the divide D: by definition, this is a black vertex at each crossing and a white vertex for each bounded region. In the perturbed front diagram D', each black (resp. white) vertex yields a bounded region in the complement $\mathbb{R}^2\setminus D'$ whose boundary has all the conormals pointing outwards (resp. inwards). In the two types of curve in the proof above, the curves ϑ_{p_i} correspond to the black vertices and the curves ϑ_{q_j} correspond to the white vertices. A bounded region in the complement $\mathbb{R}^2\setminus D'$ whose boundary has all the conormals pointing outwards (resp. inwards) is called a source (resp. a sink); a region which is not a sink or a source is said to be *mixed*.

From this viewpoint, the smooth Milnor fiber for the morsification $\tilde{f}$ associated to $D=D_{\tilde{f}}$ can be constructed by consider a 2-disk for each bounded mixed region of $\mathbb{R}^2\setminus D'$ and attaching 1-handles connecting two such 2-disks for each intersection point of the pair of corresponding mixed regions.[13] It should be possible to make this construction in the embedded and exact Lagrangian context: the 2-disks coming from the bounded mixed regions of $\mathbb{R}^2\setminus D'$ are (embedded exact) Lagrangians by virtue of being contained in the zero section of the cotangent bundle $(T^*\mathbb{R}^2,\lambda_{\mathrm{st}})$, and one would just need to argue that the 1-handle attachment can be made an exact Lagrangian 1-handle attachment with boundaries as dictated by the fronts (i.e., that adding the conical Lagrangian piece $L(D')$ is tantamount to adding these Lagrangian 1-handles). □

2.4. Lagrangian skeleta

Arboreal Lagrangian skeleta $\mathbb{L}\subseteq(W,\lambda)$ for Weinstein 4-manifolds are defined in [79,100]. Given a Weinstein manifold $W=W(\Lambda)$, the arborealization

[12] Namely, the bounded regions in $\mathbb{R}^2\setminus D'$ which do *not* come from a bounded square obtained by resolving a crossing nor from a bounded region in $\mathbb{R}^2\setminus D$; i.e. the *blue* bounded regions, as depicted in Fig. 13.

[13] Some of these 1-handles might be attached between a region and itself.

procedure in [100] yields an arboreal Lagrangian skeleton $\mathbb{L} \subseteq (W, \lambda)$ with $\partial\mathbb{L} \neq \varnothing$. Intuitively, those Lagrangian skeleta are obtained by attaching 2-handles to $\mathbb{D}^2$ along a (modification of a) front for Λ, and thus roughly contain the same information as a front $\pi(\Lambda) \subseteq \mathbb{R}^2$ for Λ. Let $\Lambda \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$ be a Legendrian link and (W, λ) a Weinstein manifold.

Definition 2.16. A compact arboreal Lagrangian skeleton $\mathbb{L} \subseteq \mathbb{C}^2$ for a Weinstein pair $(\mathbb{C}^2, \Lambda)$ is said to be closed if $\partial\mathbb{L} = \Lambda$. A compact arboreal Lagrangian skeleton $\mathbb{L} \subseteq W$ for a Weinstein manifold (W, λ) is said to be closed if $\partial\mathbb{L} = \varnothing$.

The Lagrangian skeleta in Theorem 1.1 and Corollary 1.2 are arboreal and closed. For reference, we denote the two Cal-skeleta associated to a real morsification $\tilde{f}$ of an isolated plane curve singularity $f \in \mathbb{C}[x, y]$ by

$$\mathbb{L}(\tilde{f}) := M_f \cup_{\vartheta(\tilde{f})} \bigcup_{i=1}^{|\vartheta(\tilde{f})|} \mathbb{D}^2, \qquad \overline{\mathbb{L}}(\tilde{f}) := \overline{M}_f \cup_{\vartheta(\tilde{f})} \bigcup_{i=1}^{|\vartheta(\tilde{f})|} \mathbb{D}^2.$$

The former $\mathbb{L}(\tilde{f})$ is a Lagrangian skeleton for the Weinstein pair $(\mathbb{C}^2, \Lambda_f)$, and the latter for the Weinstein 4-manifold $W(\Lambda_f)$. The notation $\overline{M}_f$ stands for the surface obtained by capping each of the boundary components of the Milnor fiber M_f with a 2-disk. The notation $\mathbb{L}(f)$ and $\overline{\mathbb{L}}(f)$ will stand for any Cal-skeleton obtained from *a* real morsification $\tilde{f}$ as in Theorem 1.1 and Corollary 1.2. Similarly, we will denote by $\vartheta(f)$ a collection of vanishing cycles $\vartheta(\tilde{f})$ obtained from *a* real morsification $\tilde{f}$, without necessarily specifying $\tilde{f}$.

Remark 2.17. In the context of low-dimensional topology, the 2-complexes underlying these Lagrangian skeleta are often referred to as *Turaev's shadows*, following [103, Chapter 8]. In particular, it is known how to compute the signature of a (Weinstein) 4-manifold from any Cal-skeleton by using [103, Chapter 9]. Similarly, the $SU(2)$-Reshetikhin-Turaev-Witten invariant of the three-dimensional (contact) boundary can be computed with the state-sum formula in [103, Chapter 10]. It would be interesting to explore if such combinatorial invariants can be enhanced to detect information on the contact and symplectic structures. □

3. Augmentation stack and the cluster algebra of Fomin–Pylyavskyy–Shustin–Thurston

In the article [47], the authors develop a connection between the topology of an isolated singularity f and the theory of cluster algebras. In concrete terms, they associate a cluster algebra $A(f)$ to an isolated singularity. An initial cluster seed for $A(f)$ is given by a quiver $Q(D_{\tilde{f}})$ coming from the AΓ-diagrams of a divide $D_{\tilde{f}}$ of a real morsification $\tilde{f}$ of f. Equivalently, by [4,61], the quiver $Q(D_{\tilde{f}})$ is the intersection quiver for a set of vanishing cycles associated to a real morsification of f. The conjectural tenet in [47] is that different choices of Morsifications lead to mutation equivalent quivers and, conversely, two quivers associated to two real morsifications of the *same* complex topological singularity must be mutation equivalent.

There are two varieties associated to a cluster algebra, the $\mathcal{X}$-cluster variety and the $\mathcal{A}$-cluster variety [44,59,95]. In the case of the cluster algebra $A(f)$ from [47], one can ask whether either of these varieties has a particularly geometric meaning. Our suggestion is that either of these cluster varieties is the moduli space of *exact* Lagrangian fillings for the Legendrian knot $\Lambda_f \subseteq (\mathbb{R}^3, \xi_{\mathrm{st}})$, with the appropriate additional data (e.g. local systems). Equivalently, they are the moduli space of (certain) objects of a Fukaya category associated to the Weinstein pair $(\mathbb{C}^2, \Lambda_f)$; for instance, the partially wrapped Fukaya category of $\mathbb{C}^2$ stopped at Λ_f. In this sense, these cluster varieties are mirror to the Weinstein pair $(\mathbb{R}^4, \Lambda_f)$.[14] Focusing on the Legendrian link $\Lambda_f \subseteq (\mathbb{R}^3, \xi_{\mathrm{st}})$, let us then suggest an alternative route from a plane curve singularity $f \in \mathbb{C}[x, y]$ to a cluster algebra $\mathcal{A}(f)$, following Definition 2.4 and Proposition 2.2 and 2.13.

Starting with $f \in \mathbb{C}[x, y]$, consider the Legendrian,[15] $\Lambda_f \subseteq (\mathbb{R}^3, \xi_{\mathrm{st}})$, where $(\mathbb{R}^3, \xi_{\mathrm{st}})$ is identified as the complement of a point in $(\mathbb{S}^3, \xi_{\mathrm{st}})$ and the Legendrian DGA $\mathscr{A}(\Lambda_f)$, as defined by Y. Chekanov in [25] and see [36]. Then we define $\mathcal{A}(f)$ to be the coordinate ring of functions on the *augmentation* variety $\mathcal{A}(\Lambda_f)$ of the DGA $\mathscr{A}(\Lambda_f)$. Technically, the DGA $\mathscr{A}(\Lambda_f)$ allows for a choice of base points, and the augmentation variety depends on that. Thus, it is more accurate to define:

Definition 3.1. Let $f \in \mathbb{C}[x, y]$ define an isolated singularity, the augmentation algebra $\mathcal{A}(f)$ associated to f is the ring of k-regular functions on the moduli stack of objects $\mathrm{ob}(\mathrm{Aug}_+(\Lambda_f))$ of the augmentation category $\mathrm{Aug}_+(\Lambda_f)$. □

The $\mathrm{Aug}_+(\Lambda)$ augmentation category of a Legendrian link $\Lambda \subseteq (\mathbb{R}^3, \xi_{\mathrm{st}})$ is introduced in [84]. An exact Lagrangian filling[16] defines an object in the category $\mathrm{Aug}_+(\Lambda)$, and the morphisms between two such objects are given by (a linearized version of) Lagrangian Floer homology. In fact, there is a sense in which any object in $\mathrm{Aug}_+(\Lambda)$ comes from a Lagrangian filling [88,89], possibly immersed, and thus $\mathrm{ob}(\mathrm{Aug}_+(\Lambda))$ is a natural candidate for a moduli space of Lagrangian fillings. The algebra $\mathcal{A}(f)$ is known to be a cluster algebra [51] in characteristic two. The lift to characteristic zero can be obtained by combining [22] and [51].

By Proposition 2.2, $\mathcal{A}(f)$ is a well-defined invariant of the complex topological singularity. For these Legendrian links $\Lambda = \Lambda_f$, the Couture-Perron algorithm [30] implies that there exist a Legendrian front $\pi(\Lambda_f) \subseteq \mathbb{R}^2$ given by the (-1)-closure of a positive braid $\beta\Delta^2$, where Δ is the half-twist; equivalently the front is the rainbow closure of the positive braid β [20]. Hence,

[14] The difference between $\mathcal{X}$- and $\mathcal{A}$-varieties should be the *decorations* we require for the Lagrangian fillings.

[15] In the context of plabic graphs [47, Section 6] the zig-zag curves [59,91] also provide a front for the Legendrian link Λ_f.

[16] Throughout the text, exact Lagrangian fillings are, if needed, implicitely endowed with a $\mathbb{C}^*$-local system.

there is a set of non-negatively graded Reeb chords generating the DGA $\mathscr{A}(\Lambda_f)$ and $\text{ob}(\text{Aug}_+(\Lambda_f))$ coincides with the set of k-valued augmentations of $\mathscr{A}(\Lambda_f)$ where exactly *one* base point per component has been chosen, k a field. The articles [22,66] provide an explicit and computational model for $\text{ob}(\text{Aug}_+(\Lambda_f))$, and thus $\mathcal{A}(f)$, as follows.

First, suppose that $\Lambda = \Lambda_f$ is a knot. Then, $\mathcal{A}(f)$ is the algebra of regular functions of the affine variety

$$X(\beta) := \{\mathcal{B}(\beta\Delta^2) + \text{diag}_{i(\beta)}(t, 1, \ldots, 1) = 0\} \subseteq \mathbb{C}^{|\beta\Delta^2|+1},$$

where $\mathcal{B}$ are the $(i(\beta) \times i(\beta))$-matrices defined in [22, Section 3] and Computation 3.2 below, $i(\beta)$ is the number of strands of β, Δ, and $|\beta\Delta^2|$ is the number of crossings of $\beta\Delta^2$. In the case Λ_f is a *link* with l components, the space $\text{ob}(\text{Aug}_+(\Lambda_f))$ is a stack[17], with isotropy groups of the form $(\mathbb{C}^*)^k$. If the tenet [47, Conjecture 5.5] holds, the affine algebraic type of the augmentation stack $\text{ob}(\text{Aug}_+(\Lambda_f))$ of a Legendrian link should recover the Legendrian link Λ_f and the complex topological type of the singularity f. Here is how to compute $\text{ob}(\text{Aug}_+(\Lambda_f))$.

Computation 3.2. Let $\Lambda = \Lambda_f$ be an algebraic knot, we can find a set of equations for the affine variety $\text{ob}(Aug_+(\Lambda_f))$, essentially using [67], see also [22]. Consider a positive braid[18] $\beta^\circ \in \text{Br}_n^+$ such that the (-1)-closure of β° is a front for $\Lambda = \Lambda(\beta^\circ)$. For $k \in [1, n-1]$, define the following $n \times n$ matrix $P_k(z)$, with variable $z \in \mathbb{C}$:

$$(P_k(z))_{ij} = \begin{cases} 1 & i = j \text{ and } i \neq k, k+1 \\ 1 & (i,j) = (k, k+1) \text{ or } (k+1, k) \\ z & i = j = k+1 \\ 0 & \text{otherwise;} \end{cases}$$

Namely, $P_k(z)$ is the identity matrix except for the (2×2)-submatrix given by rows and columns k and $k+1$, where it is $\left(\begin{smallmatrix} 0 & 1 \\ 1 & z \end{smallmatrix}\right)$. Suppose that the crossings of β°, left to right, are $\sigma_{k_1}, \ldots, \sigma_{k_s}$, $s = |\beta^\circ| \in \mathbb{N}$, $\sigma_i \in \text{Br}_n^+$ the Artin generators. Then the augmentation stack $\text{ob}(\text{Aug}_+(\Lambda_f))$ is cut out in $\mathbb{C}^s \times \mathbb{C}^* = \text{Spec}[z_1, z_2, \ldots, z_s, t, t^{-1}]$ by the n^2 equations

$$\text{diag}_n(t, 1, 1, \ldots, 1) + P_{k_1}(z_1) P_{k_2}(z_2) \cdots P_{k_s}(z_s) = 0. \tag{3.1}$$

The matrix $P_{k_1}(z_1) P_{k_2}(z_2) \cdots P_{k_s}(z_s)$ is denoted by $\mathcal{B}(\beta^\circ)$. Equations 3.1 provide a computational mean to an explicit description of the affine varieties $\text{ob}(\text{Aug}_+(\Lambda_f))$ that yield the cluster algebra $\mathcal{A}(f)$. □

Example 3.3. Consider the plane curve singularity[19] described by

$$\begin{aligned} f(x,y) &= -12x^{10}y^2 - 4x^9y^2 - 2x^7y^4 + 6x^6y^4 - 4x^3y^6 + x^{14} - 2x^{13} + x^{12} + y^8 \\ &= \left(2x^3y^2 - 4x^5y + x^7 - x^6 - y^4\right)\left(2x^3y^2 + 4x^5y + x^7 - x^6 - y^4\right) \end{aligned}$$

[17] Namely, it is isomorphic to a quotient of $X(\beta) \times (\mathbb{C}^*)^l$ by a non-free $(\mathbb{C}^*)^{l-1}$-action.

[18] Note that β° can be written in the form $\beta^\circ = \beta\Delta^2$.

[19] We have chosen this example as a continuation of [30, Example 5.3] and [47, Figure 6].

The Puiseux expansion yields $y(x) = x^{3/2} + x^{7/4}$ and using the Couture-Perron algorithm [30], or [47, Definition 11.3], a positive braid word associated to this singularity is

$$\beta = (\sigma_2\sigma_1\sigma_3\sigma_2\sigma_1\sigma_3\sigma_2\sigma_1)\sigma_3(\sigma_1\sigma_2\sigma_3\sigma_1\sigma_2\sigma_3\sigma_1\sigma_2)\sigma_1\sigma_3$$

The Legendrian $\Lambda_f \subseteq (\mathbb{R}^3, \xi)$ is the rainbow closure of β, and the (-1)-framed closure of $\beta^\circ = \beta\Delta^2$. Note that Λ_f is a knot, and thus we will use one base point $t \in \mathbb{C}^*$ in the computation of $X(\beta) = \mathrm{ob}(\mathrm{Aug}_+(\Lambda_f))$. Following Computation 3.2 above, we can write equations for affine variety $X(\beta)$ as a subset $X(\beta) \subseteq \mathbb{C}^{31} \times \mathbb{C}^*$. We use coordinates $(z_1, z_2, \ldots, z_{31}; t) \in \mathbb{C}^{31} \times \mathbb{C}^*$, $(z_1, z_2, \ldots, z_{19})$ corresponding to the 19 crossings of β and $(z_{20}, \ldots, z_{31})$ account for the 12 crossings of $\Delta^2 \in \mathrm{Br}_3^+$. There are a total of 16 equations, the first two of which read as follows:

$$\begin{aligned}
&z_{11} + z_9z_{12} + (z_9 + (z_{11} + z_9z_{12})\,z_{18})\,z_{20} + (z_{13} + z_9z_{14} + (z_{11} + z_9z_{12})\,z_{15})\,z_{21}\\
&\quad + (z_9z_{16} + (z_{11} + z_9z_{12})\,z_{17} + (z_{13} + z_9z_{14} + (z_{11} + z_9z_{12})\,z_{15})\,z_{19} + 1)\,z_{23} = -t^{-1}\\
&z_7 + z_6z_9 + (z_8z_{10} + z_6z_{11} + (z_7 + z_6z_9)z_{12} + 1)z_{18} + (z_8 + z_6z_{13} + (z_7 + z_6z_9)z_{14}\\
&\quad +(z_8z_{10} + z_6z_{11} + (z_7 + z_6z_9)z_{12} + 1)z_{15})z_{22} + (z_6 + (z_7 + z_6z_9)z_{16}\\
&\quad +(z_8z_{10} + z_6z_{11} + (z_7 + z_6z_9)z_{12} + 1)z_{17}\\
&\quad +(z_8 + z_6z_{13} + (z_7 + z_6z_9)z_{14} + (z_8z_{10} + z_6z_{11}\\
&\quad +(z_7 + z_6z_9)z_{12} + 1)z_{15})z_{19})z_{24} + (z_8z_{10} + z_6z_{11} + (z_7 + z_6z_9)z_{12}\\
&\quad +(z_7 + z_6z_9 + (z_8z_{10} + z_6z_{11} + (z_7 + z_6z_9)z_{12} + 1)z_{18})z_{20}\\
&\quad +(z_8 + z_6z_{13} + (z_7 + z_6z_9)z_{14}\\
&\quad +(z_8z_{10} + z_6z_{11} + (z_7 + z_6z_9)z_{12} + 1)z_{15})z_{21} + (z_6 + (z_7 + z_6z_9)z_{16}\\
&\quad +(z_8z_{10} + z_6z_{11} + (z_7 + z_6z_9)z_{12} + 1)z_{17}\\
&\quad +(z_8 + z_6z_{13} + (z_7 + z_6z_9)z_{14} + (z_8z_{10} + z_6z_{11}\\
&\quad +(z_7 + z_6z_9)z_{12} + 1)z_{15})z_{19})z_{23} + 1)z_{31} = 0
\end{aligned}$$

The remaining 14 equations are longer, but can be readily obtained. This hopefully illustrates that the method is computationally immediate.[20] □

Remark 3.4. (i) One may consider the moduli stack $\mathrm{ob}(\mathrm{Sh}^1_{\Lambda_f}(\mathbb{R}^2))$ of sheaves with microlocal rank-1 along Λ_f, instead of $\mathrm{ob}(\mathrm{Aug}_+(\Lambda_f))$. By [84], there is an equivalence of categories $\mathrm{Aug}_+(\Lambda_f) \cong \mathrm{Sh}^1_{\Lambda_f}(\mathbb{R}^2)$. The stack $\mathrm{ob}(\mathrm{Sh}^1_{\Lambda_f}(\mathbb{R}^2))$ is a $\mathcal{X}$-cluster variety; the associated $\mathcal{A}$-cluster variety in the cluster ensemble is the moduli of *framed* sheaves [95].[21] In short, the cluster algebra $\mathcal{A}(f)$ could have been defined in terms of the moduli space of constructible sheaves microlocally supported in Λ, instead of Floer theory.

(ii) The Aug_+-category is Floer-theoretical in nature, e.g. its morphisms are certain Floer homology groups. It would have also been natural to consider the partially wrapped Fukaya category $W(\mathbb{C}^2, \Lambda_f)$, as defined [50, 101], or the infinitesimal Fukaya category $Fuk(\mathbb{C}^2, \Lambda)$ [77, 81]. These

[20] Even if the equations themselves, being rather long, may not be particularly enlightening.

[21] The cluster algebra structure for $\mathcal{A}(f)$ defined by [51] is obtained by pulling-back the cluster algebra structure of the open Bott-Samelson cell associated to β. There should exist a cluster algebra structure on $\mathcal{A}(f)$ defined strictly in Floer-theoretical terms.

are Floer-theoretical Legendrian invariants associated to Λ_f, and thus the singularity $f \in \mathbb{C}[x,y]$, which might be of interest on their own.

4. A few computations and remarks

Consider the derived dg-category $\mathrm{Sh}_\Lambda(M)$ of constructible sheaves in a closed smooth manifold M microlocally supported at a Legendrian link $\Lambda \subseteq (\partial(T^*M), \xi_{\mathrm{st}})$, e.g. as introduced in [97, Section 1]. Equivalently, one may consider a conical Lagrangian $L \subseteq T^*M$ instead of $\Lambda \subseteq (\partial(T^*M), \xi_{\mathrm{st}})$; in practice, the input data is a wavefront $\pi(\Lambda) \subseteq M$ [8]. Let $\mu\,\mathrm{sh}$ denote the sheaf of microlocal sheaves defined[22] in [80, Section 5]. There are two situations we consider, depending on whether the focus is on the Weinstein pair $(\mathbb{C}^2, \Lambda_f)$ or on the Weinstein 4-manifold $W(\Lambda_f)$:

(i) *Sheaf Invariants of the Weinstein pair* $(\mathbb{C}^2, \Lambda_f)$.[23] The category of microlocal sheaves $\mu\,\mathrm{sh}(\mathbb{L}(f))$ is an invariant of $(\mathbb{C}^2, \Lambda_f)$, as established in [60,80,97].[24] In this case, the global sections $\mu\,\mathrm{sh}(\mathbb{L}(f))$ is a category equivalent to the more familiar $\mathrm{Sh}_{\Lambda(f)}(\mathbb{R}^2)$. For simplicity, we focus on the moduli stack $\mathcal{S}(f) \subseteq \mathrm{ob}(\mathrm{Sh}_{\Lambda(f)}(\mathbb{R}^2))$ of sheaves whose microlocal support is rank one, microlocally supported in the Legendrian link of an isolated plane curve singularity $f : \mathbb{C}^2 \longrightarrow \mathbb{C}$. See [69, Section 7.5] or [60, Section 1.10] for a detailed discussion on these sheaves. In our case $\Lambda = \Lambda(f)$, $\mathcal{S}(f)$ is an Artin stack of finite type [97, Prop. 5.20], and typically is an algebraic variety or a G-quotient thereof, with $G = (\mathbb{C}^*)^k$ or $\mathrm{GL}(k, \mathbb{C})$. Note that $\mu\,\mathrm{sh}(\mathbb{L}(f))$ is equivalent to the wrapped Fukaya category of $\mathbb{C}^2$ stopped at Λ_f [49].

(ii) *Sheaf Invariants of the Weinstein 4-manifold* $W(\Lambda_f)$. The category $\mu\,\mathrm{sh}(\overline{\mathbb{L}}(f))$ of microlocal sheaves [80] on a Lagrangian skeleton $\overline{\mathbb{L}}(f) \subseteq W(\Lambda_f)$ is an invariant of $W(\Lambda_f)$, up to Weinstein homotopy [80] and up to symplectomorphism [49]. This category is[25] $\mathrm{Sh}_{\vartheta(f)}(\overline{M}_f)$, or $\mu loc(\overline{\mathbb{L}}(f))$, in the notation of [96], i.e. the global sections of the Kashiwara-Schapira sheaf of dg-categories [96, Prop. 3.5] on the Lagrangian skeleton $\overline{\mathbb{L}}(f)$. For simplicity, we focus on the moduli stack $\theta(f) \subseteq \mu\,\mathrm{sh}(\overline{\mathbb{L}}(f))$ of microlocal rank-1 sheaves as well. Note that $\mu\,\mathrm{sh}(\overline{\mathbb{L}}(f))$ is equivalent to the wrapped Fukaya category of $W(\Lambda_f)$ by [49].

The moduli stack $\mathcal{S}(f)$ in (i) is isomorphic to the stack of microlocal rank-1 sheaves in $\mathrm{ob}(\mathrm{Sh}_{\vartheta(f)}(M_f))$. This is because the union of $\mathbb{R}^2 \subseteq T^*\mathbb{R}^2$ and the Lagrangian cone of $\Lambda \subseteq (T^+\mathbb{R}^2, \xi_{\mathrm{st}})$ is a Lagrangian skeleton for the relative Weinstein pair $(\mathbb{C}^2, \Lambda)$, so is $\mathbb{L}(f)$ by Theorem 1.1, and

[22] Thanks go to V. Shende for helpful discussions on sheaf invariants.

[23] Invariance up to Weinstein homotopy [28], and also symplectomorphism of Liouville pairs.

[24] The category $\mu\,\mathrm{sh}(\mathbb{L}(f))$ is likely *not* an invariant of the Weinstein 4-manifold $W(\Lambda_f)$ itself.

[25] Recall that we denote by $\vartheta(f)$ a collection of vanishing cycles $\vartheta(\tilde{f})$ obtained from *a* real morsification $\tilde{f}$.

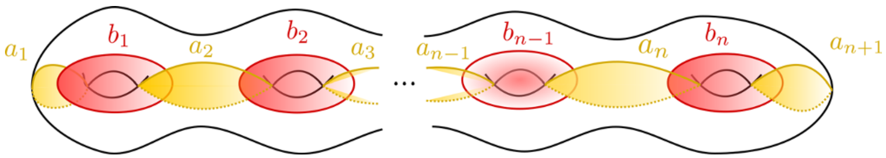


FIGURE 14. A Cal-skeleta $\overline{\mathbb{L}}(f_{2n+1})$ for the Weinstein 4-manifolds $W(\Lambda(A_{2n+1}))$

$\mathrm{ob}(\mathrm{Sh}_{\vartheta(f)}(M_f))$ is an invariant of the Weinstein pair $(\mathbb{C}^2,\Lambda)$, independent of the choice of Lagrangian skeleton. Thus, the difference between $\mathcal{S}(f)$ and $\theta(f)$ is at the boundary, which for $\mathcal{S}(f)$ might give monodromy contributions (and these become trivial on $\theta(f)$). In other words, since $\overline{\mathbb{L}}(f)$ is obtained from $\mathbb{L}(f)$ by attaching 2-disks (to close the boundary of the Milnor fiber M_f), the category $\mu\,\mathrm{sh}(\overline{\mathbb{L}}(f))$ is a homotopy pull-back of $\mu\,\mathrm{sh}(\mathbb{L}(f))$.

Remark 4.1. There are currently two methods for computing $\mathcal{S}(f)$: either by direct means, as exemplified in [97], or by using the equivalence of categories $Aug_+(\Lambda(f)) \cong \mathrm{Sh}^1_{\Lambda_f}(\mathbb{R}^2)$ from [84, Theorem 1.3], the latter being denoted by $\mathcal{C}_1(\Lambda_f)$ in [84]. Thanks to the computational techniques available for augmentation varieties, the moduli of objects $\mathrm{ob}(Aug_+(\Lambda(f)))$ is readily computable for (-1)-framed closures of positive braids as in Sect. 3 above, confer Computation 3.2. Similarly $\theta(f)$ could be computed directly, or by means of the isomorphism to the wrapped Fukaya category[26] of $W(\Lambda_f)$. $\square$

In this section, we take to opportunity to build on [80, 96] and perform an actual computation for a class of Cal-Skeleta coming from Theorem 1.1.

4.1. Cal-skeleta for A_n-singularities

Consider the A_n-singularity $f_n(x,y) = x^{n+1} + y^2$. The Legendrian $\Lambda(A_n) \subseteq (\mathbb{R}^3,\xi_{\mathrm{st}})$ associated to the singularity is the max-tb Legendrian $(2,n+1)$-torus link. By Theorem 1.1, a Lagrangian skeleton $\mathbb{L}(f_n)$ for the Weinstein pair $(\mathbb{C}^2,\Lambda_f)$ is obtained by attaching n 2-disks to a $(3/2-(-1)^n/2)$–punctured $\lfloor\frac{n-1}{2}\rfloor$–genus surface along an A_n-Dynkin chain of embedded curves. Similarly, Corollary 1.2 implies that a Lagrangian skeleton $\overline{\mathbb{L}}(f_n)$ for the Weinstein 4-manifold $W_n = W(\Lambda(A_n))$ is given by attaching n 2-disks to a $\lfloor\frac{n-1}{2}\rfloor$–genus surface along an A_n-Dynkin chain, as depicted in orange in Fig. 15, see also Fig. 14.

Let us compute $\theta(f_n)$ for $n \in \mathbb{N}$ even, so that $\Lambda(A_n)$ is a knot; the $n \in \mathbb{N}$ odd case is similar. The key technical tool is the Disk Lemma [68, Lemma 4.2.3]. The Disk Lemma explains, in precise terms, how to compute the category of microlocal sheaves on a two-dimensional Lagrangian skeleton $\mathbb{S}\cup_\gamma \mathbb{D}^2$ in terms of the category for the corresponding Lagrangian skeleton $\mathbb{S}$, where $\mathbb{D}^2$ is attached along an embedded smooth curve $\gamma \subseteq \mathbb{S}$. In brief, the Disk Lemma states that the microlocal sheaf category for $\mathbb{S}\cup_\gamma \mathbb{D}^2$ has as its

[26] Should the reader be willing to use the surgery formula, this wrapped Fukaya category may be presented as modules over the Legendrian DGA of Λ_f. (This is only informative and not needed for the present purposes.)

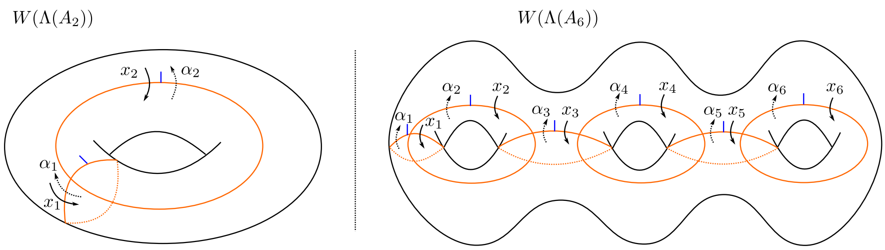


FIGURE 15. The Cal-skeleta $\overline{\mathbb{L}}(f)$ for the Weinstein 4-manifolds $W(\Lambda(A_2))$ and $W(\Lambda(A_6))$. The relative Cal-skeleta $\overline{\mathbb{L}}(f)$ for the corresponding Weinstein pairs $(\mathbb{C}^2, \Lambda(A_2))$ and $(\mathbb{C}^2, \Lambda(A_6))$ are obtained by introducing one puncture to the surfaces

objects pairs consisting of an object $\mathscr{F}_{\mathbb{S}}$ in the category for $\mathbb{S}$ *and* a (derived) trivialization of the microlocal monodromy of $\mathscr{F}_{\mathbb{S}}$ along γ, i.e. a homotopy from this microlocal monodromy to the identity.

The complement $\overline{M}_f \setminus \vartheta(f)$ of the vanishing cycles is a 2-disk, and the category of local systems is just $\mathbb{C}$-mod. Thus, the moduli of simple constructible sheaves on $\overline{M}_f$ microlocally supported on (the Legendrian lift of) the vanishing cycles $\vartheta(f)$ consists of a vector space $V = \mathbb{C}$ and maps $x_1, x_2, \ldots, x_n \in End(V)$, one associated to each vanishing cycle. This is depicted in Fig. 15 for $n = 2, 6$, and note that $n = |\vartheta(f)|$. Denote by $\overline{\mathbb{L}}(f_n)_0 \subseteq T^*\overline{M}_f$ the Lagrangian skeleton given by $\overline{M}_f$ union the conormal lifts of $\vartheta(f)$. These maps are *not* necessarily invertible in $\mu\,\mathrm{sh}(\overline{\mathbb{L}}(f_n)_0)$.

The skeleton $\overline{\mathbb{L}}(f_n)$ is obtained by attaching n Lagrangian 2-disks to $\overline{\mathbb{L}}(f_n)_0$, i.e. $\overline{\mathbb{L}}(f_n)$ is the homotopy push-out of $\overline{\mathbb{L}}(f_n)_0$ and the disjoint union of n 2-disks. In consequence, the category of microlocal sheaves on $\overline{\mathbb{L}}(f_n)$ is given by the homotopy pull-back of the category of microlocal sheaves on $\overline{\mathbb{L}}(f_n)_0$ and the category of microlocal sheaves on n disjoint 2-disks (which are just copies of $\mathbb{C}$-mod). Attaching a 2-disk along a vanishing V_i cycle in $\vartheta(f)$, $i \in [1, n]$, has the effect of trivializing the "monodromy" corresponding map x_i, by the Disk Lemman [68, Lemma 4.2.3] cited above; see [96, Section 4] and [68, Section 4.2] for the details. Here, the monodromy[27] is given by restricting a microlocal sheaf to (an arbitrarily small neighborhood of) V_i. Note that in this restriction, we land into a 1-dimensional Lagrangian skeleton given by a circle $V_i \cong S^1$ *union* conical segments coming from the adjacent vanishing cycles. Let us call γ_i the composition of maps from $cone(x_i)$ to itself obtained by going around V_i, each of the maps coming from traversing a segment. Then, the trivialization is a homotopy to the identity, and it translates into adding a map α_i such that $x_i\alpha_i - 1 = \gamma_i$.

Example 4.2. Consider the map x_1 in Fig. 15 (Left), which is depicted transversely to the vanishing cycle V_1. The restriction of a microlocal sheaf to a

[27]We had written "monodromy" in quotations because it is not a priori necessarily invertible.

neighborhood of V_1 gives a microlocal sheaf for the skeleton $\mathbb{S}^1 \cup T_p^{*,+}\mathbb{S}^1 \subseteq T^*\mathbb{S}^1$, where $T_p^{*,+}\mathbb{S}^1$ is the positive half of the cotangent fiber at a point $p \in \mathbb{S}^1$. Such a microlocal sheaf is described by a (complex of) vector space(s) and an endomorphism. In this case the vector space is $V = \mathbb{C}$ and this endomorphism is identified with $\gamma_1 = x_2$. Hence, trivializing along V_1 adds a map $\alpha_1 \in End(\mathbb{C})$, which we can think of as a variable $\alpha_1 \in \mathbb{C}$, such that $x_1\alpha_1 + 1 = -x_2$. Similarly, trivializing along V_2, with $\gamma_2 = -\alpha_1$, adds a variable $\alpha_2 \in \mathbb{C}$ such that $1 + x_2\alpha_2 = -\alpha_1$. Hence $\theta(f)$ is the affine variety

$$\theta(f_3) = \{(x, y, z) \in \mathbb{C}^3 : xyz + x - z - 1 = 0\}.$$

This affine variety appears in the study of isomonodromic deformations of the Painlevé I equation [105, Section 3.10], see also [18, Section 5]. □

The vanishing cycles V_1, V_n have simpler monodromies γ_1, γ_n, as they only intersect *one* other vanishing cycle. Adding the 2-disks to the skeleton $\overline{\mathbb{L}}(f_n)_0$ along V_1, V_n yields a category of microlocal sheaves whose moduli space of simple objects is described by that of $\overline{\mathbb{L}}(f_n)_0$ and the two equations $x_1\alpha_1 + 1 = -x_2$ and $x_n\alpha_n + 1 = -\alpha_{n-1}$. For each of the middle vanishing cycles V_i, $2 \le i \le n-1$, we have the monodromy $\gamma_i = \alpha_{i-1}x_{i+1}$. In consequence, attaching the n 2-disks $\overline{\mathbb{L}}(f_n)_0$ along all the curves V_i, $i \in [1, n]$, leads to the moduli space

$$\theta(f) \cong \{(x_i, \alpha_i) \in (\mathbb{C}^2)^n : x_1\alpha_1 + 1 = -x_2, x_n\alpha_n + 1 \\ = -\alpha_{n-1}, 1 + x_j\alpha_j = \alpha_{j-1}x_{j+1}, j \in [2, n-1]\}.$$

Remark 4.3. Consider $(n+3)$-tuples of vectors $(v_1, \ldots, v_{n+3}) \in \mathbb{C}^2$, modulo $\mathrm{GL}_2(\mathbb{C})$, the equations for $\theta(f)$ above can be read directly by writing the $(n+3)$-tuple as

$$\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}-1\\x_1\end{pmatrix}, \begin{pmatrix}\alpha_1\\x_2\end{pmatrix}, \begin{pmatrix}\alpha_2\\x_3\end{pmatrix}, \begin{pmatrix}\alpha_3\\x_4\end{pmatrix}, \begin{pmatrix}\alpha_4\\x_5\end{pmatrix}, \ldots, \begin{pmatrix}\alpha_{n-1}\\x_n\end{pmatrix}, \begin{pmatrix}\alpha_n\\-1\end{pmatrix},$$

and imposing $v_i \wedge v_{i+1} = 1$, where we have use the $GL_2(\mathbb{C})$ gauge group to trivialize the first two vectors, and one component of the third and last vectors. Boalch [18] names this moduli stack after Sibuya [99]. Note that [18, Section 5] points out that some of these equations were initially discovered by Euler [42]. In the context of open Bott–Samelson cells [95,98], these spaces appear as the open positroid varieties $\{p \in \mathrm{Gr}(2, n+3) : P_{i,i+1}(p) \neq 0\}$, where $P_{i,j}$ is the Plücker coordinate given by the minor at the i and j columns, and the index i is understood $\mathbb{Z}/(n+3)$-cyclically. □

Finally, we notice that the cohomology $H^*(\theta(f), \mathbb{C})$, or that of $H^*(\mathcal{S}(f), \mathbb{C})$, can be an interesting invariant [97, Section 6]. For the case of A_n-singularities, we can use the fact that these are actually cluster varieties of A_n-type in order to compute their cohomology using [72, Section 6.2]. For $n = 2m \in \mathbb{N}$ even, and removing any $\mathbb{C}^*$-factors coming from frozen variables, one obtains that the Abelian graded cohomology group is isomorphic to $\mathbb{Q}[t]/t^{m+1}$, $|t| = 2$. In general, the mixed Hodge structure for these moduli spaces can be nontrivial, but for singularities of A_n-type, these cohomologies are of Hodge–Tate type, and entirely concentrated in $\bigoplus_{k\geq 0} H^{k,(k,k)}$.

Remark 4.4. It would be valuable to understand the relation between sheaf invariants of a singularity $f \in \mathbb{C}[x,y]$, such as $\mu\,\mathrm{sh}(\mathbb{L}(f))$ and $\mu\,\mathrm{sh}(\overline{\mathbb{L}}(f))$, and classical invariants from singularity theory [3,9,10]. In particular, it could be valuable to develop more systematic methods to compute $\mu\,\mathrm{sh}(\mathbb{L}(f))$ and $\mu\,\mathrm{sh}(\overline{\mathbb{L}}(f))$ both directly and from a divide. □

5. Structural conjectures on Lagrangian fillings

Let $\Lambda \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$ be a max-tb Legendrian link. The classification of embedded exact Lagrangian fillings $L \subseteq (\mathbb{D}^4, \lambda_{\mathrm{st}})$ with fixed boundary Λ, up to Hamiltonian isotopy, is a central question. The only Legendrian Λ for which a complete classification exists is the standard unknot [33]. In this case, the standard Lagrangian flat disk is the unique filling: there is precisely *one* exact Lagrangian filling, up to Hamiltonian isotopy.

The recent developments [20,22,23] show that such finiteness is actually rare: e.g. the max-tb torus links (n,m) admit *infinitely* many exact Lagrangian filling, up to Hamiltonian isotopy, if $n, m \geq 4$. It is proven in [20] that Legendrian representatives of infinitely many types of either torus, satellite or hyperbolic knots admit infinitely many Hamiltonian isotopy classes of embedded exact Lagrangian fillings. This final section states and discusses Conjectures 5.1 and 5.4, which might help towards our understanding of the classification of exact Lagrangian fillings of Legendrian links.

Geometric strategy Given $\Lambda \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$, we would like to know whether it admits finitely many Lagrangian fillings or not, and in the finite case provide the exact count. Theorem 1.1 provides insight for the class of Legendrian links $\Lambda \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$ that are algebraic links and, more generally, arise from a divide. Indeed, Lagrangian fillings for Λ can be constructed by using the Lagrangian skeleta for the Weinstein pair $(\mathbb{C}^2, \Lambda)$ built in the statement. For instance, the inclusion of the Lagrangian Milnor fiber $L_{\tilde{f}} \subseteq \mathbb{L}_{\tilde{f}}$ provides an exact Lagrangian filling, and performing Lagrangian disk surgeries along the Lagrangian 2-disks in $\mathbb{L}_{\tilde{f}} \setminus L_{\tilde{f}}$, which bound vanishing cycles, will potentially yield new Lagrangian fillings. This strategy can be implemented in certain cases but, in general, one must be able to find an *embedded* Lagrangian disk in the new Lagrangian skeleton (with an embedded boundary curve), to perform the next Lagrangian disk surgery. Curves being immersed rather than embedded[28], might a priori represent a challenge.[29] This geometric scheme has the following algebraic incarnation.

Algebraic strategy Consider the intersection quiver $Q_{\vartheta(\tilde{f})}$ of vanishing cycles for a real morsification $\tilde{f}$, Lagrangian disk surgeries induce mutations of the quiver [96] and the (microlocal) monodromies of a local system serve as cluster $\mathcal{X}$-variables [23,98]. Thus, the cluster algebra $\mathcal{A}(Q(f))$ associated

[28]Equivalently, the existence of curves with zero algebraic intersection but non-empty geometric intersection.

[29]The vanishing cycles can be organized as a quiver Q, the additional data of a superpotential (Q, W) should be helpful in solving the disparity between *immersed* and *embedded* curves in the Milnor fiber.

to the quiver, as it appears in [47], governs *possible* exact Lagrangian fillings for the Legendrian link Λ. That is, a Lagrangian filling $L \subseteq (\mathbb{D}^4, \lambda_{\rm st})$ yields a cluster chart for this algebra [51,98], and the Lagrangian skeleta from Theorem 1.1 provide a geometric realization for the quiver in the form of an exact Lagrangian filling with ambient Lagrangian disks ending on it.

The recent developments [20,51,96,98] and the existence of the Lagrangian skeleta in Theorem 1.1 shyly hint towards the fact that, possibly, Lagrangian fillings *are* classified by the cluster algebra $\mathcal{A}(Q(f))$. That is, every cluster chart in $\mathcal{A}(Q(f))$ is induced by *precisely* one exact Lagrangian filling.[30] It should be emphasized that this is *not* known for any $\Lambda \subseteq (\mathbb{R}^3, \xi_{\rm st})$ except the standard Legendrian unknot. It is possible that the case of the Hopf link $\Lambda(A_1)$ can be solved by building on the techniques in [92], which classifies exact Lagrangian tori near the Whitney sphere;[31] this is currently work in progress.

Having presented the available evidence, we state the following conjectural guide:

Conjecture 5.1. (ADE Classification of Lagrangian Fillings) *Let $\Lambda \subseteq (\mathbb{R}^3, \xi_{st})$ be the Legendrian rainbow closure of a positive braid such that the mutable part of its brick quiver is connected. Then one of the following possibilities occur:*

1. *Λ is smoothly isotopic to the link of the A_n-singularity. Then Λ has precisely $\frac{1}{n+2}\binom{2n+2}{n+1}$ exact Lagrangian fillings.*
2. *Λ is smoothly isotopic to the link of the D_n-singularity. Then Λ has precisely $\frac{3n-2}{n}\binom{2n-2}{n-1}$ exact Lagrangian fillings.*
3. *Λ is smoothly isotopic to the link of the E_6, E_7 or the E_8-singularities. Then Λ has precisely 833, 4160, and 25080 exact Lagrangian fillings, respectively.*
4. *Λ has infinitely many exact Lagrangian fillings.*

The following comments are in order:

(i) In [45], Fomin and Zelevinsky classify cluster algebras of *finite* type. This is an ADE-classification, parallel to the classification of simple singularities [9], the Cartan–Killing classification of semisimple Lie algebras, finite crystallographic root systems (via Dynkin diagrams) and the like. Thus, Conjecture 5.1 first states that Λ will have *finitely* many exact Lagrangian fillings, up to Hamiltonian isotopy, if and only if the associated quiver is ADE.

(ii) The case of $\Lambda = \Lambda_f$ an algebraic link associated to a non-simple singularity $f \in \mathbb{C}[x,y]$ of a plane curve follows from [20], and the case of a Legendrian Λ with a non-ADE underlying quiver has recently been proven in [52]. These approaches are based on the following fact: if there exists an embedded exact Lagrangian cobordism from Λ_- to Λ_+ and Λ_- admits infinitely many Lagrangian fillings, then so does Λ_+. See [22,86]

[30] That is, two Lagrangian fillings inducing the same cluster chart in $\mathcal{A}(Q(f))$ are Hamiltonian isotopic *and* every cluster chart is induced by at least one Lagrangian filling.

[31] See also [29], which appeared during the writing of this manuscript.

and [20, Section 6]. This itself initiates the quest for finding the smallest Legendrian link which admits infinitely many exact Lagrangian fillings. At present, if we measure the size of a link Λ as $\pi_0(\Lambda)+2g(\Lambda)$, $g(\Lambda)$ the (minimal) genus of a (any) embedded Lagrangian filling, the smallest known Legendrian link has $g(\Lambda)=1$ and two components $\pi_0(\Lambda)=2$; it is built in [22]. Intuitively, it is the geometric link corresponding to the $\tilde{A}_2$ cluster algebra.

(iii) According to (ii) above, the missing ingredient for Conjecture 5.1 is showing that (1), (2) and (3) hold. For the A_n-case (1), it is known that there are *at least* the stated Catalan number worth of exact Lagrangian fillings, distinct up to Hamiltonian isotopy. This was originally proven by Pan [87] and subsequently understood in [98,102] from the perspective of microlocal sheaf theory. It remains to show that any exact Lagrangian filling of $\Lambda(A_n)$ is Hamiltonian isotopic to one of those; the first unsolved case is the Hopf link $\Lambda(A_1)$ having exactly two embedded exact Lagrangian fillings.[32] For the $\Lambda(D_n), \Lambda(E_6), \Lambda(E_7)$ and $\Lambda(E_8)$ cases in Conjecture 5.1, one needs to first find the corresponding number of distinct Lagrangian fillings, and then show these are all. The construction part should be relatively accessible, in the spirit of either [23,87,98], and it is reasonable to suspect that these many fillings can be distinguished using either augmentations or microlocal monodromies.[33]

(iv) The numbers appearing in Conjecture 5.1. (i)–(iii) are the number of cluster seeds for the corresponding cluster algebra. Precisely, consider a root system of Cartan-Killing type X_n, $e_1,\dots,e_n$ its exponents and h the Coxeter number. Then the numbers in Conjecture 5.1 are $N(X_n)=\prod_{i=1}^n(e_i+h+1)(e_i+1)^{-1}$ for $X_n=A_n,D_n,E_6,E_7,E_8$. It is natural to strengthen Conjecture 5.1. (iv) to: 4. *There exist a natural bijection between Hamiltonian isotopy classes of exact Lagrangian fillings of* Λ *and cluster seeds of the (natural) cluster* A*-structure on the augmentation variety of* Λ*, decorated with one marked point per component.* (The bijection assigns to a Lagrangian filling L the set $H^1(L,\mathbb{C}^*)$ of $\mathbb{C}^*$-local systems on L, which is naturally a subset of the augmentation variety and a cluster chart.)

Note that Conjecture 5.1 has a natural analogue for $W(\Lambda)$. Namely, the Hamiltonian isotopy classes of closed exact Lagrangians in $W(\Lambda)$ are precisely given by the numerics above. This aligns with the spirit of the nearby Lagrangian conjecture, now for surface skeleta: if we interpret $W(\Lambda)$ as "generalized cotangent bundle" $T^*\overline{\mathbb{L}}$, where $\mathbb{L}$ is a Cal-skeleton for Λ, then the conjecture would be that any closed exact Lagrangian is Hamiltonian isotopic to a Lagrangian which is either a subset of $\overline{\mathbb{L}}$ or can be obtained from it via (iterated) Lagrangian disk mutations.

[32] In particular, this would show that the two possible Polterovich surgeries [90] of a 2-dimensional Lagrangian node are the only two exact Lagrangian cylinders near the node, up to Hamiltonian isotopy.

[33] Showing these exhaust all fillings, up to Hamiltonian isotopy, is another matter, possibly much more challenging.

The brick graph of a positive braid is defined in [13,94], it can be enhanced to a quiver, which we call the brick quiver, following the algorithm in [95, Section 3.1] or [51, Section 4.2], which itself generalizes the wiring diagram construction in [16,43].

Remark 5.2. The hypothesis of the mutable part of its brick quiver being connected is necessary. We could otherwise add a meridian to any positive braid, which would create a disconnected quiver; the resulting cluster algebra would be a product with A_1, which preserves being of finite type. It stands to reason that adding a meridian to a Legendrian link Λ would yield a Legendrian link $\Lambda \cup \mu$ with exactly *twice* as many Lagrangian fillings. It is clear that there are at least twice as many Lagrangian fillings for $\Lambda \cup \mu$, as there are two distinct Lagrangian cobordisms from Λ to $\Lambda \cup \mu$. The simplest case is $\Lambda = \Lambda_0$ the standard Legendrian unknot and $\Lambda \cup \mu \cong \Lambda(A_1)$ the Hopf link, which should have $2 = 2 \cdot 1$ Lagrangian fillings, in accordance with Conjecture 5.1. The next case would be $\Lambda = \Lambda(A_1)$, so that $\Lambda(A_1) \cup \mu \cong \Lambda(D_2)$, in line with $\Lambda(D_2)$ conjecturally having $4 = 2 \cdot 2$ Lagrangian fillings. □

Note that the article [22] has provided the first examples of Legendrian links $\Lambda \subseteq (\mathbb{S}^3, \xi_{\mathrm{st}})$ which are *not* rainbow closures of positive braids and yet they admit infinitely many Lagrangian fillings, up to Hamiltonian isotopy. These Legendrian links have components which are stabilized, not max-tb, and thus they cannot be rainbow closures of any positive braid. It would be interesting to extend Conjecture 5.1 to a larger class of links, possibly including (-1)-framed closures of certain positive braids, e.g., those with Demazure product equal to a half-twist, as studied in [22].

Remark 5.3. To the author's knowledge, [33,87], Theorem 1.1, and the recent [20,22,23,51,52], constitute the current evidence towards Conjecture 5.1. Hints towards Conjecture 5.1 might have appeared in the symplectic folklore in one form or another: e.g. the advent of Symplectic Field Theory led to the mantra of "pseudoholomorphic curves or nothing"[34], the subsequent arrival of microlocal sheaf theory to symplectic topology led to "sheaves or nothing". In the current zeitgeist, cluster algebras provide a new algebraic invariant that one might hope to be complete.[35] □

In the line of Remark 5.3, a natural strengthening of Conjecture 5.1, under same the hypotheses, would be to speculate that there exists *precisely* one Hamiltonian isotopy class of Lagrangian fillings per each cluster seed in the augmentation variety associated to $\Lambda \subseteq (\mathbb{R}^3, \xi_{\mathrm{st}})$. Given our current understanding, this might as well be the case. The statement is correct for the unknot and current work in progress indicates that it is correct for the Hopf link.

[34] That is, if pseudoholomorphic invariants cannot distinguish two objects, they must be equal.

[35] As with the previous two cases, there is no particularly hard evidence for "cluster algebras or nothing".

Finally, an ADE-classification is often part of a larger classification,[36] involving a few additional families. For instance, simple Lie algebras are classified by connected Dynkin diagrams, which are A_n, D_n, E_6, E_7, E_8, known as the simply laced Lie algebras, and B_n, C_n, F_4 and G_2. These latter cases, B_n, C_n, F_4 and G_2, are interesting on their own right. For instance, simple singularities are classified according to A_n, D_n, E_6, E_7, E_8, and B_n, C_n, F_4 then arise in the classification of simple *boundary* singularities [9, Chapter 17.4], as shown in [10, Chapter 5.2]. (See also D. Bennequin's [15, Section 8] and [7].) In general, the tenet is that B_n, C_n, F_4 and G_2 arise when classifying the same objects as in the ADE-classification *with the additional* data of a symmetry.[37] This a perspective (and technique) called *folding*, ubiquitous in the study of B_n, C_n, F_4, G_2, which is developed in [46, Section 2.4] for the case of cluster algebras.

Let us consider a Legendrian $\Lambda \subseteq (\mathbb{R}^3, \xi_{\text{st}})$, a Lagrangian filling $L \subseteq (\mathbb{R}^4, \lambda_{\text{st}})$, $\partial L = \Lambda$, and a finite group G acting faithfully on $(\mathbb{R}^4, \lambda_{\text{st}})$ by exact symplectomorphisms, inducing an action on the boundary piece $(\mathbb{R}^3, \xi_{\text{st}})$ by contactomorphisms. For instance, $s : \mathbb{R}^4 \longrightarrow \mathbb{R}^4$, $s(x, y, z, w) = (-x, -y, z, w)$ is an involutive symplectomorphism which restricts to the contactomorphism $(x, y, z) \mapsto (-x, -y, z)$ on its boundary piece $(\mathbb{R}^3, \ker\{dz - ydx\})$. Let us define an exact Lagrangian G-filling of Λ to be an exact Lagrangian filling L of Λ such that $G(L) = L$ and $G(\Lambda) = \Lambda$ setwise. Also, by definition, we say $\Lambda \subseteq (\mathbb{R}^3, \xi_{\text{st}})$ admits a G-symmetry if there exists a faithful action of G by contactomorphisms on $(\mathbb{R}^3, \xi_{\text{st}})$ such that $G(\Lambda) = \Lambda$ setwise. Examples of such symmetries can be readily drawn in the front projection, as shown in Fig. 16 for $\Lambda(A_9), \Lambda(D_8), \Lambda(E_6)$ and $\Lambda(D_4)$. Following the tenet above, the following classification might be plausible:

Conjecture 5.4. (BCFG Classification of Lagrangian Fillings) *Let $\Lambda(\beta) \subseteq (\mathbb{S}^3, \xi_{st})$ the Legendrian rainbow closure of a positive braid β:*

1. (B_n) *If $\Lambda(\beta) = \Lambda(A_{2n-1})$, the $\mathbb{Z}_2$-symmetry $(x, z) \longrightarrow (-x, z)$ for the front depicted in* Fig. 16 *lifts to a $\mathbb{Z}_2$-symmetry of $\Lambda(A_{2n-1})$. Then $\Lambda(A_{2n-1})$ has precisely $\binom{2n}{n}$ exact Lagrangian $\mathbb{Z}_2$-fillings.*
2. (C_n) *If $\Lambda(\beta) = \Lambda(D_{n+1})$, the $\mathbb{Z}_2$-symmetry $(x, z) \longrightarrow (-x, z)$ for the front depicted in* Fig. 16 *lifts to a $\mathbb{Z}_2$-symmetry of $\Lambda(D_{n+1})$. Then $\Lambda(D_{n+1})$ has precisely $\binom{2n}{n}$ exact Lagrangian $\mathbb{Z}_2$-fillings.*
3. (F_4) *If $\Lambda(\beta) = \Lambda(E_6)$, the $\mathbb{Z}_2$-symmetry $(x, z) \longrightarrow (-x, z)$ in the front depicted in* Fig. 16 *lifts to a $\mathbb{Z}_2$-symmetry of $\Lambda(E_6)$. Then $\Lambda(E_6)$ has precisely* 105 *exact Lagrangian $\mathbb{Z}_2$-fillings.*
4. (G_2) *If $\Lambda(\beta) = \Lambda(D_4)$, the $\mathbb{Z}_3$-symmetry in the front depicted in* Fig. 16 *lifts to a $\mathbb{Z}_3$-symmetry of $\Lambda(D^4)$. Then $\Lambda(D_4)$ has precisely* 8 *exact Lagrangian $\mathbb{Z}_3$-fillings.*

[36] The larger classification is an ABCDEFG-classification, which admittedly does not roll off the tongue.

[37] The study of boundary singularities can be understood as the study of singularities taking into account a certain $\mathbb{Z}_2$-symmetry.

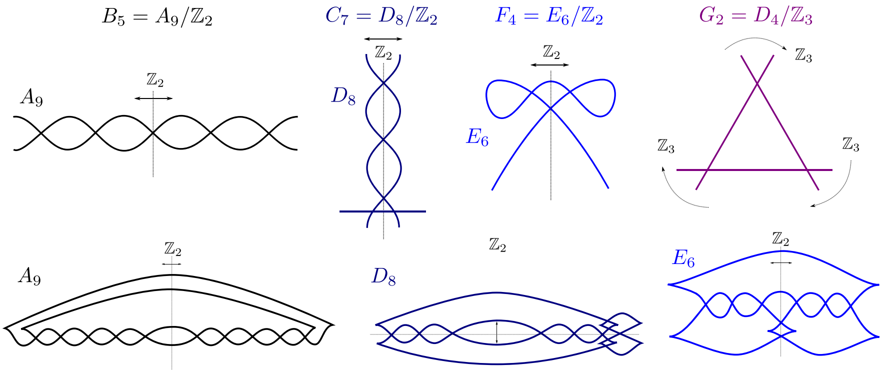


FIGURE 16. Legendrian fronts for $\Lambda(A_{2n-1}), \Lambda(D_{n+1}), \Lambda(E_6), \Lambda(D_4)$ with G-symmetries, $G = \mathbb{Z}_2, \mathbb{Z}_3$. The upper row exhibits these symmetric fronts as divides of the associated singularities, and the lower row depicts them in the standard front projection $(x, y, z) \mapsto (x, z)$ for a Darboux chart $(\mathbb{R}^3, \xi_{\text{st}})$

For the G_2-case in Conjecture 5.4.(4), it might be helpful to notice that the D_4-singularity is topologically equivalent to $f(x, y) = x^3 + y^3$. The $\mathbb{Z}_3$-symmetry cyclically interchanges the three linear branches of this singularity. In particular, we can draw a front for the Legendrian $\Lambda(D_4)$ as the $(3, 3)$-torus link, the rainbow closure of $\beta = (\sigma_1\sigma_2)^6$.[38]

For the B_n-case in Conjecture 5.4.(1), the construction of $\binom{2n}{n}$ distinct Lagrangian $\mathbb{Z}_2$-fillings likely follows from adapting [87]. Indeed, in the $\mathbb{Z}_2$-invariant front for $\Lambda(A_{2n-1})$, as depicted in Fig. 16, there are n crossing to the left, equivalently right, of the $\mathbb{Z}_2$-symmetry axis. We can construct a $\mathbb{Z}_2$-filling of $\Lambda(A_{2n-1})$ by opening those n crossings in any order, with the rule that we simultaneously open the corresponding $\mathbb{Z}_2$-symmetric crossing.[39] Should one distinguish these $\mathbb{Z}_2$-fillings via their augmentations, as in [87], an appropriate G-equivariant Floer theoretic invariant (e.g., G-equivariant DGA and its augmentations) needs to be defined. The perspective of microlocal sheaves [102] yields combinatorics closer to those of triangulations [45, Section 12.1], modeling A_n-cluster algebras, and thus might provide a simpler route to distinguish these fillings. In either case, Conjecture 5.4 calls for a G-equivariant theory of invariants for Legendrian submanifolds of contact manifolds.

5.1. Some questions

We finalize this section with a series of problems on Weinstein 4-manifolds and their Lagrangian skeleta. To my knowledge, there are several unanswered

[38]The $\mathbb{Z}_3$-action should coincide with the loop $\Xi_1 \circ (\delta^{-1} \circ \Xi_1 \circ \delta)$ from [20, Section 2].
[39]The naive count of 312-pattern avoiding permutations from [32,87] would indicate that there are $\frac{1}{n}\binom{2n}{n}$ such Lagrangian $\mathbb{Z}_2$-fillings, instead of $\binom{2n}{n}$. Thus, should Conjecture 5.4 hold, there must be an additional rule for $\mathbb{Z}_2$-fillings (not just those in [87, Lemma 3.10]), possibly related to the fact that the crossing closest to the $\mathbb{Z}_2$-axis is different from the rest.

questions at this stage, including checkable characterizations of Weinstein 4-manifolds of the form $W(\Lambda_f)$, where Λ_f is the Legendrian link of an isolated plane curve singularity. Here are some interesting, yet hopefully reasonable, problems:

Problem 1. Find a characterization of Legendrian links $\Lambda \subseteq (\mathbb{S}^3, \xi_{st})$ for which $(\mathbb{C}^2, \Lambda)$, or $W(\Lambda)$, admits a Cal-skeleton. (Ideally, a verifiable characterization.)

For instance, if $\Lambda \subseteq (\mathbb{S}^3, \xi_{st})$ is the rainbow closure of a positive braid, then $W(\Lambda)$ can be shown to admit a Cal-skeleton by methods similar to the ones presented in this manuscript. In contrast, if Λ is a stabilized Legendrian knot, then $W(\Lambda)$ does not admit a Cal-skeleton.

Problem 2. Find necessary and sufficient conditions for a Lagrangian skeleton $\mathbb{L} \subseteq (W, \lambda)$ to guarantee that the Stein manifold (W, λ) is an affine algebraic manifold. Similarly, characterize Legendrian links $\Lambda \subseteq (\mathbb{S}^3, \xi_{st})$ such that $W(\Lambda)$ is an affine algebraic variety.

Note that the standard Legendrian unknot $\Lambda_0 \cong \Lambda(A_0) \subseteq (\mathbb{S}^3, \xi_{st})$ and the max-tb Hopf link $\Lambda(A_1) \subseteq (\mathbb{S}^3, \xi_{st})$ yield *affine* Weinstein manifolds, as we have

$$W(\Lambda_0) \cong \{(x, y, z) \in \mathbb{C}^3 : x^2 + y^2 + z^2 = 1\},$$
$$W(\Lambda(A_1)) \cong \{(x, y, z) \in \mathbb{C}^3 : x^3 + y^2 + z^2 = 1\}.$$

By [21, Section 4.1], the trefoil $\Lambda(A_2)$ is also an example of such a Legendrian link, as

$$W(\Lambda(A_2)) \cong \{(x, y, z) \in \mathbb{C}^3 : xyz + x + z + 1 = 0\}.$$

Heuristic computations indicate that $\Lambda(A_3)$ and $\Lambda(D_4)$ also have this property. See [73, 74] for a source of necessary conditions, and [93] for (topological) skeleta of affine hypersurfaces.

Problem 3. Find necessary and sufficient conditions for a Lagrangian skeleton[40] $\mathbb{L} \subseteq (W, \lambda)$ to guarantee that the Stein manifold (W, λ) is flexible.[41] (Again, a verifiable characterization.) Similarly, characterize $\Lambda \subseteq (\mathbb{S}^3, \xi_{st})$ such that $W(\Lambda)$ is flexible.

Note that affine manifolds $W \subseteq \mathbb{C}^N$ might be flexible [21, Theorem 1.1]. In particular, it could be fruitful to compare Lagrangian skeleta of $X_m = \{(x, y, z) \in \mathbb{C}^3 : x^m y + z^2 = 1\}$ for $m = 1$ and $m \geq 2$, e.g. the ones provided in [93].

Problem 4. Suppose that a Weinstein 4-manifold $W = W(\Lambda)$ is obtained as a Lagrangian 2-handle attachment to $(\mathbb{D}^4, \omega_{st})$. Given a Cal-skeleton $\mathbb{L} \subseteq (W, \lambda)$, devise an algorithm to find one such possible Legendrian $\Lambda \subseteq (\partial\mathbb{D}^4, \xi_{st})$.

[40] Not closed in this case.

[41] See [28] for flexible Weinstein manifolds. In the 4-dimensional case above, we might just define flexible as being of the form $W = W(\Lambda)$ where Λ is a stabilized knot.

(Note that such a Legendrian Λ might not be unique, i.e. it could be possible that two non-isotopic Legendrian knots Λ_1, Λ_2 might have Weinstein isomorphic traces $W(\Lambda_1) \cong W(\Lambda_2)$.)

Problem 5. Let $L \subseteq (W, \lambda)$ be a closed exact Lagrangian surface. Study whether there exists a Cal-skeleton $\mathbb{L} \subseteq (W, \lambda)$ such that $L \subseteq \mathbb{L}$. In addition, study whether there exists a Legendrian handlebody $\Lambda \subseteq (\#^k \mathbb{S}^1 \times \mathbb{S}^2, \xi_{\mathrm{st}})$, so that $W = W(\Lambda)$, and L is obtained by capping a Lagrangian filling of a Legendrian sublink of Λ.

See [106] for an interesting construction in the case of Bohr-Sommerfeld Lagrangian submanifolds and see [34] for a general discussion on regular Lagrangians. The nearby Lagrangian conjecture holds for $W = T^*\mathbb{S}^2, T^*\mathbb{T}^2$, thus the answer is affirmative in these cases.

Problem 6. Characterize which cluster algebras A can arise as the ring of functions of the augmentation stack of a Legendrian link $\Lambda \subseteq (\mathbb{S}^3, \xi)$.

By using double-wiring diagrams [16], (generalized) double Bruhat cells satisfy this property [95]. It is proven in [22, 51] that the cluster algebras $A(\tilde{D}_n)$ of affine D_n-type have this property. Heuristic computations indicate that the affine types $\tilde{A}_{p,q}$ also verify this [22]. It might be reasonable to conjecture that cluster algebras of surface type all have this property.

Here is a variation on this problem. Suppose that a cluster algebra A arises, e.g., as an augmentation variety associated to a Legendrian link Λ. An interesting problem might be to characterize those elements of the cluster automorphism group of A which arise as Legendrian loops of Λ. In certain cases, this is known to be the case for Grassmannian braid symmetries [20, 48], the square of the Donaldson–Thomas transformation [52] and the Zamolodchikov operator [66].

In general, relating geometric properties of Lagrangian fillings to algebraic properties of cluster algebras should be fruitful. For instance, already in Type A, it would be interesting to *geometrically characterize* those Lagrangian fillings of the $(2, n)$-torus links that yield *positive* cluster seeds. More ambitiously, it would seem useful to be able to access geometrically, e.g. via holomorphic curve counts, the $\mathbb{Z}^t$-tropical structure, or the $\mathbb{R}^+$-positive structure, of the cluster varieties associated to some Legendrian links.

Problem 7. Let $a_3(\Lambda)$ be the number of A_3-arboreal singularities of a Cal-skeleton $\mathbb{L} \subseteq (W, \lambda)$. Find the number $a_3(W) := \min_{\mathbb{L} \subseteq W} a_3(\mathbb{L})$, where $\mathbb{L} \subseteq W$ runs amongst all possible Cal-skeleta. In particular, characterize Weinstein 4-manifolds (W, λ) with $a_3(W) = 0$.

Problem 8. Develop a combinatorial theory of symplectomorphisms in $\mathrm{Symp}(W, d\lambda)$ in terms of Cal-skeleta $\mathbb{L} \subseteq (W, \lambda)$.

This is being developed in the case $\dim(W) = 2$ by using A'Campo's tête-à-tête twists [5, Section 3], see also [6, Section 5]. A (symplectic) mapping class in $\mathrm{Symp}(W, d\lambda)$ is a composition of Dehn twists in this 2-dimensional case. This is no longer the case in $\dim(W) = 4$, e.g., due to the existence of

Biran-Giroux's fibered Dehn twists, confer [104, Section 3] and [107, Section 2]. Note that $\pi_0(\mathrm{Symp}(W))$ might be infinite even if W contains no exact Lagrangian 2-spheres [20].

Problem 9. Compare Cal-skeleta $\mathbb{L}_1 \subseteq (W_1, \lambda_1)$, $\mathbb{L}_2 \subseteq (W_2, \lambda_2)$ for exotic Stein pairs W_1, W_2. That is, W_1 is homeomorphic to W_2, but not diffeomorphic. In particular, investigate *skeletal corks*: combinatorial modifications on a Cal-skeleton that can produce exotic Stein pairs.

In [82], Naoe uses Bing's house [17] to study some such corks.

Problem 10. Find a contact analogue of Turaev's Shadow formula[42] [103, Chapter 10] for the contact 3-dimensional boundary in terms of the combinatorics of a Cal-skeleton $\mathbb{L} \subseteq (W, \lambda)$. That is, find a *contact* invariant[43] of $(\partial W, \lambda|_{\partial W})$ which can be computed in terms of the combinatorics of $\mathbb{L} \subseteq (W, \lambda)$.

Acknowledgements

The author is grateful to Patrick Popescu–Pampu and the referee for their insightful comments and suggestions on the manuscript. The author is supported by the NSF grant DMS-1841913, the NSF CAREER grant DMS-1942363, a BBVA Research Fellowship and the Alfred P. Sloan Foundation.

References

[1] A'Campo, N.: Sur la monodromie des singularités isolées d'hypersurfaces complexes. Invent. Math. **20**, 147–169 (1973)

[2] A'Campo, N.: Le groupe de monodromie du déploiement des singularités isolées de courbes planes. I. Math. Ann. **213**, 1–32 (1975)

[42] This expresses the $SU(2)$–Reshetikhin–Turaev–Witten quantum invariant of a 3-manifold in terms of a shadow as a (colored) multiplicative Euler characteristic.

[43] E.g. it would be interesting to describe the Ozsvath-Szabo contact class in Heegaard Floer homology, or M. Hutching's contact class in Embedded Contact homology, in terms of $\mathbb{L}$ as well.

[3] A'Campo, N.: Generic immersions of curves, knots, monodromy and Gordian number. Inst. Hautes Études Sci. Publ. Math. **88**, 151–169 (1999)

[4] A'Campo, N.: Real deformations and complex topology of plane curve singularities. Ann. Fac. Sci. Toulouse Math. **8**(1), 5–23 (1999)

[5] A'Campo, N.: Lagrangian spine and symplectic monodromy. In: Notes for the 6th Franco-Japanese-Vietnamese Symposium on Singularites (2018)

[6] A'Campo, N., Bobadilla, J.F. Pe, M.P., Pablo, P.C.: Tete-à-tete twists, monodromies and representation of elements of Mapping Class Group. ArXiv e-prints (2017)

[7] Arnold, V.I.: Critical points of functions on a manifold with boundary, the simple Lie groups B_k, C_k, F_4 and singularities of evolutes. Uspekhi Mat. Nauk **33**((5(203))), 91–105, 237 (1978)

[8] Arnol'd, V.I.: Singularities of Caustics and Wave Fronts, volume 62 of Mathematics and its Applications (Soviet Series). Kluwer Academic Publishers Group, Dordrecht (1990)

[9] Arnol'd, V.I., Guseĭn-Zade, S.M., Varchenko, A.N.: Singularities of Differentiable Maps. Vol. I, volume 82 of Monographs in Mathematics. Birkhäuser, Boston (1985) (**The classification of critical points, caustics and wave fronts, Translated from the Russian by Ian Porteous and Mark Reynolds**)

[10] Arnol'd, V.I., Guseĭn-Zade, S.M., Varchenko, A.N.: Singularities of Differentiable Maps. Vol. II, Volume 83 of Monographs in Mathematics. Birkhäuser, Boston (1988) (Monodromy and asymptotics of integrals, Translated from the Russian by Hugh Porteous, Translation revised by the authors and James Montaldi)

[11] Arnold, V.I., Goryunov, V.V., Lyashko, O.V., Vasil'ev, V.A.: Singularity Theory. I. Springer, Berlin (1998) (Translated from the 1988 Russian original by A. Iacob, Reprint of the original English edition from the series Encyclopaedia of Mathematical Sciences [ıt Dynamical systems. VI, Encyclopaedia Math. Sci., 6, Springer, Berlin (1993))

[12] Auroux, D.: A beginner's Introduction to Fukaya Categories. In: Contact and Symplectic Topology, Volume 26 of Bolyai Soc. Math. Stud, pp. 85–136. János Bolyai Math Soc, Budapest (2014)

[13] Baader, S., Lewark, L., Liechti, L.: Checkerboard graph monodromies. Enseign. Math. **64**(1–2), 65–88 (2018)

[14] Bennequin, D.: Entrelacements et équations de Pfaff. In: Third Schnepfenried Geometry Conference, Vol. 1 (Schnepfenried, 1982), Volume 107 of Astérisque. Soc. Math. France, Paris, pp. 87–161 (1983)

[15] Bennequin, D.: Caustique mystique (d'après Arnol' d et al.). Number 133–134. Seminar Bourbaki, vol. 1984/85, pp. 19–56 (1986)

[16] Berenstein, A., Fomin, S., Zelevinsky, A.: Cluster algebras. III. Upper bounds and double Bruhat cells. Duke Math. J. **126**(1), 1–52 (2005)

[17] Bing, R. H.: Some aspects of the topology of 3-manifolds related to the Poincaré conjecture. In: Lectures on Modern Mathematics, vol. II. Wiley, New York, pp. 93–128 (1964)

[18] Boalch, P.: Wild character varieties, points on the Riemann sphere and Calabi's examples. In: Representation Theory, Special Functions and Painlevé Equations—RIMS 2015, Volume 76 of Adv. Stud. Pure Math.. Math. Soc. Japan, Tokyo, pp. 67–94 (2018)

[19] Brauner, K.: Zur geometrie def funktionen zweier veränderlichen. Abh. Math. Sem. Hamburg **6**, 1–54 (1928)

[20] Casals, R., Gao, H.: Infinitely many Lagrangian fillings. Ann. Math. **195**(1), 207–249 (2022)

[21] Casals, R., Murphy, E.: Legendrian fronts for affine varieties. Duke Math. J. **168**(2), 225–323 (2019)

[22] Casals, R., Ng, L.L.: Braid Loops with infinite monodromy on the Legendrian contact DGA. (2020) ArXiv e-prints 2101.02318

[23] Casals, R., Zaslow, E.: Legendrian weaves. Geom. Topol. (2022)

[24] Chantraine, B.: Lagrangian concordance of Legendrian knots. Algebr. Geom. Topol. **10**(1), 63–85 (2010)

[25] Chekanov, Y.: Differential algebra of Legendrian links. Invent. Math. **150**(3), 441–483 (2002)

[26] Chmutov, S.: Diagrams of divide links. Proc. Am. Math. Soc. **131**(5), 1623–1627 (2003)

[27] Chongchitmate, W., Ng, L.: An atlas of Legendrian knots. Exp. Math. **22**(1), 26–37 (2013)

[28] Cieliebak, K., Eliashberg, Y.: From Stein to Weinstein and back, Volume 59 of American Mathematical Society Colloquium Publications. American Mathematical Society, Providence (Symplectic geometry of affine complex manifolds) (2012)

[29] Coté, L., Rizell, G.D.: Symplectic Rigidity of Fibers in Cotangent Bundles of Riemann Surfaces (2020) ArXiv e-prints

[30] Couture, O., Perron, B.: Representative braids for links associated to plane immersed curves. J. Knot Theory Ramif. **9**(1), 1–30 (2000)

[31] Eisenbud, D., Neumann, W.: Three-Dimensional Link Theory and Invariants of Plane Curve Singularities. Annals of Mathematics Studies, vol. 110. Princeton University Press, Princeton (1985)

[32] Ekholm, T., Honda, K., Kálmán, T.: Legendrian knots and exact Lagrangian cobordisms. J. Eur. Math. Soc. (JEMS) **18**(11), 2627–2689 (2016)

[33] Eliashberg, Y., Polterovich, L.: Local Lagrangian 2-knots are trivial. Ann. Math. (2) **144**(1), 61–76 (1996)

[34] Eliashberg, Y., Ganatra, S., Lazarev, O.: Flexible Lagrangians. Int. Math. Res. Not. IMRN **8**, 2408–2435 (2020)

[35] Epstein, J., Fuchs, D., Meyer, M.: Chekanov-Eliashberg invariants and transverse approximations of Legendrian knots. Pac. J. Math. **201**(1), 89–106 (2001)

[36] Etnyre, J.B., Ng, L.: Legendrian contact homology in $\mathbb{R}^3$ (2018). arXiv e-prints

[37] Etnyre, J.B.: Introductory lectures on contact geometry. In *Topology and geometry of manifolds (Athens, GA, 2001)*, Volume 71 of Proc. Sympos. Pure Math.. Amer. Math. Soc., Providence, pp. –107 (2003)

[38] Etnyre, J.B.: Legendrian and transversal knots. In: Handbook of Knot Theory. Elsevier B. V., Amsterdam, pp. 105–185 (2005)

[39] Etnyre, J.B.: Lectures on open book decompositions and contact structures. In: Floer Homology, Gauge Theory, and Low-Dimensional Topology, Volume 5 of Clay Math. Proc.. Amer. Math. Soc., Providence, pp. 103–141 (2006)

[40] Etnyre, J.B., Honda, K.: Cabling and transverse simplicity. Ann. Math. (2) **162**(3), 1305–1333 (2005)

[41] Etnyre, J.B., LaFountain, D.J., Tosun, B.: Legendrian and transverse cables of positive torus knots. Geom. Topol. **16**(3), 1639–1689 (2012)

[42] Euler, L.: Specimen algorithmi singularis. Novi Commentarii academiae scientiarum Petropolitanae **9**, 53–69 (1764)

[43] Fock, V.V., Goncharov, A.B.: Cluster x-varieties, amalgamation, and Poisson-Lie groups. In: Algebraic Geometry and Number Theory, Volume 253 of Progr. Math.. Birkhäuser Boston, pp. 27–68 (2006)

[44] Fock, V., Goncharov, A.: Moduli spaces of local systems and higher Teichmüller theory. Publ. Math. Inst. Hautes Études Sci. **103**, 1–211 (2006)

[45] Fomin, S., Zelevinsky, A.: Cluster algebras. II. Finite type classification. Invent. Math. **154**(1), 63–121 (2003)

[46] Fomin, S., Zelevinsky, A.: Y-systems and generalized associahedra. Ann. Math. (2) **158**(3), 977–1018 (2003)

[47] Fomin, S., Pylyavskyy, P., Shustin, E., Thurston, D.: Morsifications and mutations (2020). ArXiv e-prints

[48] Fraser, C.: Braid group symmetries of Grassmannian cluster algebras (2018). ArXiv e-prints

[49] Ganatra, S., Pardon, J., Shende, V.: Microlocal Morse theory of wrapped Fukaya categories (2018). ArXiv e-prints

[50] Ganatra, S., Pardon, J., Shende, V.: Covariantly functorial wrapped Floer theory on Liouville sectors. Publ. Math. Inst. Hautes Études Sci. **131**, 73–200 (2020)

[51] Gao, H., Shen, L., Weng, D.: Augmentations, fillings, and clusters (2020). ArXiv e-prints

[52] Gao, H., Shen, L., Weng, D.: Positive braid links with infinitely many fillings (2020). ArXiv e-prints

[53] Geiges, H.: An introduction to Contact Topology. Cambridge Studies in Advanced Mathematics, vol. 109. Cambridge University Press, Cambridge (2008)

[54] Ghys, É.: A Singular Mathematical Promenade. ENS Éditions, Lyon (2017)

[55] Gibson, W., Ishikawa, M.: Links of oriented divides and fibrations in link exteriors. Osaka J. Math. **39**(3), 681–703 (2002)

[56] Giroux, E.: Géométrie de contact: de la dimension trois vers les dimensions supérieures. In: Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002), Beijing. Higher Ed. Press, pp. 405–414 (2002)

[57] Giroux, E.: Ideal Liouville domains—a cool gadget (2017). ArXiv e-prints: 1708.08855

[58] Gompf, R.E.: Handlebody construction of Stein surfaces. Ann. Math. (2) **148**(2), 619–693 (1998)

[59] Goncharov, A.B.: Ideal webs, moduli spaces of local systems, and 3d Calabi-Yau categories. In: Algebra, Geometry, and Physics in the 21st Century, Volume 324 of Progr. Math.. Birkhäuser/Springer, Cham, pp. 31–97 (2017)

[60] Guillermou, S., Kashiwara, M., Schapira, P.: Sheaf quantization of Hamiltonian isotopies and applications to nondisplaceability problems. Duke Math. J. **161**(2), 201–245 (2012)

[61] Guseĭn-Zade, S.M.: Intersection matrices for certain singularities of functions of two variables. Funkcional. Anal. i Priložen. **8**(1), 11–15 (1974)

[62] Hirasawa, M.: Visualization of A'Campo's fibered links and unknotting operation. In: Proceedings of the First Joint Japan-Mexico Meeting in Topology (Morelia, 1999) **121**, pp. 287–304 (2002)

[63] Ishikawa, M.: Tangent circle bundles admit positive open book decompositions along arbitrary links. Topology **43**(1), 215–232 (2004)

[64] Ishikawa, M., Naoe, H.: Milnor fibration. A'Campo's divide and Turaev's shadow (2020). ArXiv e-prints

[65] Kähler, E.: Mathematische Werke/Mathematical works. Walter de Gruyter & Co., Berlin (2003) (Edited by Rolf Berndt and Oswald Riemenschneider)

[66] Kálmán, T.: Contact homology and one parameter families of Legendrian knots. Geom. Topol. **9**, 2013–2078 (2005)

[67] Kálmán, T.: Braid-positive Legendrian links. Int. Math. Res. Not. **29**, 14874 (2006)

[68] Karabas, D.: Microlocal Sheaves on Pinwheels (2018). ArXiv e-prints

[69] Kashiwara, M., Schapira, P.: Sheaves on Manifolds, Volume 292 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Berlin (With a chapter in French by Christian Houzel) (1990)

[70] Kawamura, T.: Quasipositivity of links of divides and free divides. Topol. Appl. **125**(1), 111–123 (2002)

[71] LaFountain, D.: Studying uniform thickness i: Legendrian simple torus knots (2009). e-print at arxiv: 0905.2760

[72] Lam, T., Speyer, D.E.: Cohomology of cluster varieties. I. Locally acyclic case (2018). ArXiv e-prints

[73] McLean, M.: The growth rate of symplectic homology and affine varieties. Geom. Funct. Anal. **22**(2), 369–442 (2012)

[74] McLean, M.: Affine varieties, singularities and the growth rate of wrapped Floer cohomology. J. Topol. Anal. **10**(3), 493–530 (2018)

[75] Milnor, J.W.: Topology from the Differentiable Viewpoint. Based on notes by David W. Weaver. The University Press of Virginia, Charlottesville (1965)

[76] Milnor, J.: Singular Points of Complex Hypersurfaces. Annals of Mathematics Studies, No. 61. Princeton University Press, Princeton (1968)

[77] Nadler, D.: Microlocal branes are constructible sheaves. Selecta Math. (N.S.) **15**(4), 563–619 (2009)

[78] Nadler, D.: Non-characteristic expansions of Legendrian singularities (2015). ArXiv e-prints

[79] Nadler, D.: Arboreal singularities. Geom. Topol. **21**(2), 1231–1274 (2017)

[80] Nadler, D., Shende, V.: Sheaf quantization in Weinstein symplectic manifolds (2020). ArXiv e-prints

[81] Nadler, D., Zaslow, E.: Constructible sheaves and the Fukaya category. J. Am. Math. Soc. **22**(1), 233–286 (2009)

[82] Naoe, H.: Mazur manifolds and corks with small shadow complexities. Osaka J. Math. **55**(3), 479–498 (2018)

[83] Neuwirth, L.P.: editor. Knots, Groups, and 3-manifolds. Princeton University Press, Princeton. University of Tokyo Press, Tokyo. Papers dedicated to the memory of R. H. Fox, Annals of Mathematics Studies, No. 84 (1975)

[84] Ng, L., Rutherford, D., Sivek, S., Zaslow, E.: Augmentations are Sheaves. ArXiv e-prints, Vivek Shende (2015)

[85] Ozbagci, B., Stipsicz, A.I.: Surgery on Contact 3-Manifolds and Stein Surfaces. Bolyai Society Mathematical Studies, vol. 13. Springer, Berlin (2004)

[86] Pan, Yu.: The augmentation category map induced by exact Lagrangian cobordisms. Algebr. Geom. Topol. **17**(3), 1813–1870 (2017)

[87] Pan, Yu.: Exact Lagr. fillings of Legendrian $(2, n)$ torus links. Pac. J.M. **289**(2), 417–441 (2017)

[88] Pan, Y., Rutherford, D.: Functorial LCH for immersed Lagr. cobordisms (2019). ArXiv e-prints

[89] Pan, Y., Rutherford, D.: Augmentations and immersed Lagrangian fillings (2020). ArXiv e-prints

[90] Polterovich, L.: The surgery of Lagrange submanifolds. Geom. Funct. Anal. **1**(2), 198–210 (1991)

[91] Postnikov, A.: Total positivity, Grassmannians, and networks (2006). ArXiv e-prints

[92] Rizell, G.D.: The classification of Lagrangians nearby the Whitney immersion. Geom. Topol. **23**(7), 3367–3458 (2019)

[93] Ruddat, H., Sibilla, N., Treumann, D., Zaslow, E.: Skeleta of affine hypersurfaces. Geom. Topol. **18**(3), 1343–1395 (2014)

[94] Rudolph, L.: Quasipositive annuli (Constructions of quasipositive knots and links. IV). J. Knot Theory Ramif. **1**(4), 451–466 (1992)

[95] Shen, L., Weng, D.: Cluster structures on double Bott–Samelson cells (2015). ArXiv e-prints

[96] Shende, V., Treumann, D., Williams, H.: On the combinatorics of exact Lagrangian surfaces (2016). ArXiv e-prints

[97] Shende, V., Treumann, D., Zaslow, E.: Legendrian knots and constructible sheaves. Invent. Math. **207**(3), 1031–1133 (2017)

[98] Shende, V., Treumann, D., Williams, H., Zaslow, E.: Cluster varieties from Legendrian knots. Duke Math. J. **168**(15), 2801–2871 (2019)

[99] Sibuya, Y.: Global Theory of a Second Order Linear Ordinary Differential Equation with a Polynomial Coefficient. North-Holland Publishing Co., Amsterdam-Oxford, American Elsevier Publishing Co., Inc., New York. North-Holland Mathematics Studies, vol. 18 (1975)

[100] Starkston, L.: Arboreal singularities in weinstein skeleta. Selecta Math. (to appear)

[101] Sylvan, Z.: On partially wrapped Fukaya categories. J. Topol. **12**(2), 372–441 (2019)

[102] Treumann, D., Zaslow, E.: Cubic planar graphs and legendrian surface theory (2016). ArXiv e-prints

[103] Turaev, V.G.: Quantum Invariants of Knots and 3-Manifolds. De Gruyter Studies in Mathematics, vol. 18. Walter de Gruyter & Co., Berlin (1994)

[104] Uljarevic, I.: Floer homology of automorphisms of Liouville domains. J. Symplect. Geom. **15**(3), 861–903 (2017)

[105] van der Put, M., Saito, M.-H.: Moduli spaces for linear differential equations and the Painlevé equations. Ann. Inst. Fourier (Grenoble) **59**(7), 2611–2667 (2009)

[106] Vérine, A.: Bohr–Sommerfeld Lagrangian submanifolds as minima of convex functions. J. Symplect. Geom. **18**(1), 333–353 (2020)

[107] Wehrheim, K., Woodward, C.T.: Exact triangle for fibered Dehn twists. Res. Math. Sci. **3**, 75 (2016)

[108] Weinstein, A.: Contact surgery and symplectic handlebodies. Hokkaido Math. J. **20**(2), 241–251 (1991)

[109] Yau, M.-L.: Surgery and isotopy of Lagrangian surfaces. In: Proceedings of the Sixth International Congress of Chinese Mathematicians. Vol. II, Volume 37 of Adv. Lect. Math. (ALM). Int. Press, Somerville, pp. 143–162 (2017)

Roger Casals
Department of Mathematics, Shields Avenue
University of California Davis
Davis CA95616
USA
e-mail: casals@math.ucdavis.edu

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:26
https://doi.org/10.1007/s11784-022-00940-1
Published online April 6, 2022

Journal of Fixed Point Theory and Applications

Reeb chords of Lagrangian slices

Baptiste Chantraine

Abstract. In this short note, we observe that the boundary of a properly embedded compact exact Lagrangian sub-manifold in a subcritical Weinstein domain X necessarily admits Reeb chords. The existence of a Reeb chord either follows from an obstruction to the deformation of the boundary to a cylinder over a Legendrian sub-manifold or from the fact that the wrapped Floer homology of the Lagrangian vanishes once this boundary has been "collared".

Mathematics Subject Classification. Primary 53D12, Secondary 57R17.

Keywords. Lagrangian sub-manifolds, reeb chords.

1. Introduction

Let Y be a contact manifold with contact form α, and a *Lagrangian slice* in Y is a sub-manifold $i : \Lambda \hookrightarrow Y$, such that their exists a Lagrangian embedding

$$C : (1-\varepsilon, 1+\varepsilon) \times \Lambda \hookrightarrow ((1-\epsilon, 1+\epsilon) \times Y, d(t\alpha))$$

satisfying

$$C\left((1-\varepsilon, 1+\varepsilon) \times \Lambda\right) \pitchfork \{1\} \times Y = C\left(\{1\} \times \Lambda\right).$$

It is the opinion of the author that such objects are interesting and the aim of this note is to attract attention to those. Note that the definition depends on the contact form as the transverse intersection condition depends on the identification of the symplectisation with $\mathbb{R}_+ \times Y$. Any Legendrian sub-manifold of Y is a Lagrangian slice, since the trivial cylinder of such a sub-manifold is Lagrangian (that this notion does not depend on the choice of the contact forms comes from the fact that those are the slices giving Lagrangians tangent to the Liouville vector field, they are therefore transverse to any hyper-surface transverse to this vector field). Naturally Lagrangian slices appear in the following manner : let L be a Lagrangian sub-manifold of a symplectic manifold (M, ω). Let Y be a contact hyper-surface of M, such

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

that $L \pitchfork Y$. Then, $\Lambda = Y \pitchfork L$ is a Lagrangian slice. A *Reeb chord* of a Lagrangian slice Λ is a trajectory of the Reeb vector field of α whose start and end points are on Λ.

There is a seemingly related notion to Lagrangian slices which are prelagrangian sub-manifolds (studied for instance in [12]). A *prelagrangian* sub-manifold of $(Y^{2n-1}, \ker\alpha)$ is an n-dimensional sub-manifold L transverse to $\ker\alpha$, such that there is a function f on Y, such that the pullback of $f\alpha$ to L is closed. In this situation, any preimage of regular values of $f|_L$ is a Lagrangian slice. As embedded projections of Lagrangians in $\mathbb{R}_+ \times Y$ to Y are prelagrangian, the correspondence goes the other way around (note that one can always arrange a transverse slice to have embedded image in Y playing a bit with the Reeb flow).

A Lagrangian slice Λ is said to be *fillable* if there exists a filling X of Y and a properly embedded Lagrangian sub-manifold L of X, such that $\partial L = L \pitchfork Y = \Lambda$. Those are the slices that appear when intersecting closed Lagrangians (or Lagrangians cylindrical at infinity in Liouville manifolds) with a *separating* contact hyper-surface.

If one can deform the Liouville structure on $(1-\varepsilon, 1+\varepsilon) \times Y$ keeping Y a contact hyper-surface and so that the Liouville vector field is tangent to $(1-\varepsilon, 1+\varepsilon)\times\Lambda$, then we say that the slice is *collarable*. In [4], we gave examples of non-collarable fillable slices. Other examples of non-collarable slices appear in [18] and [13] as a slice of a Lagrangian cap cannot be collarable near its maxima.

Remark 1.1. One might find this definition of collarable unsatisfying from the contact point of view: the contact structure on Y changes. In the compact (or local) case, Moser–Gray type theorem implies that this is the same as asking that there is an isotopy of Λ through Lagrangian slices to a Legendrian sub-manifold of Y. We keep the definition this way as it is easier to manipulate, it is more convenient to study Reeb chords of slices, and it relates more easily to the notion of regular Lagrangian.

The question of collarability of slices is implicit in the question of decomposability of cobordisms (see [11] and [8]) which appears prominently in the literature about Lagrangian cobordisms in low dimension. In all dimensions, the collarability question is a sub-question of the question of regularity of Lagrangian sub-manifold from [14] (a Lagrangian is regular if one can modify the Liouville vector field so that it is tangent to the Lagrangian) as in low-dimension regular cobordisms are decomposable (as observed for instance in [15]). The author believes that the quantitative study of Lagrangian slice, particularely Inequality (2) from Sect. 4, could provide some tools to study the question of regularity of cobordism.

There is a whole hierarchy of fillability that we can imagine and it is not the purpose of this note to give a comprehensive overview of it. Our aim is to suggest that even if such objects form a larger class than Legendrian sub-manifolds, they still admits Reeb chords in a similar fashion as Legendrians do (we do not claim here that all Legendrians admits Reeb chords). For instance, we prove the following:

Theorem 1.2. *Let Λ be a compact Lagrangian slice in a compact contact manifold Y, such that there exist a subcritical Weinstein manifold W of Y and a proper exact Lagrangian embedding $L \hookrightarrow W$, such that $\partial L = \Lambda$. Then, Λ admits a Reeb chord.*

In particular, if its sub-level is subcritical, the transverse intersection of a compact exact Lagrangian with the level set of a convex Hamiltonian always has trajectories of the Hamiltonian vector field starting and ending on the Lagrangian on that level set.

In Sect. 5, we provide a proof of this theorem and it appears to be ridiculously easy (it almost appears in [3, Section 5.1]), so the interest of the paper lies somewhere else (if it lies anywhere). As mentioned before, the only purpose of this note is to raise some interesting points about such objects and discuss some of their Hamiltonian dynamics properties.

2. What does a Lagrangian slice look like?

Since they form a larger class than Legendrian sub-manifolds, it is a natural question to ask what properties a Lagrangian slice must satisfy.

Let $i : \Lambda \to Y$ be a sub-manifold. If it is a Lagrangian slice (i.e., a transverse intersection of a Lagrangian in the symplectisation with Y), then the general (linear) theory of symplectic reduction tells you that the Reeb vector field is not tangent to Λ and that Λ projects to a (smooth) Lagrangian immersion in the (singular) reduced symplectic space Y/R_α. Therefore, if it is a Lagrangian slice, then we must have $d(i^*\alpha) = 0$ and $di(T\Lambda) \cap \ker d\alpha = \{0\}$.

Conversely, let $i : \Lambda \hookrightarrow Y$ a sub-manifold, such that the pull-back of α is closed and $di(T\Lambda) \cap \ker d\alpha = \{0\}$. Let V be a copy of a neighborhood of the 0-section in $T^*\Lambda$, such that Λ corresponds to the 0-section and V is transverse to R_α. The characteristic foliation of the hyper-surface $(1 - \varepsilon, 1 + \epsilon) \times V$ in $(1 - \varepsilon, 1 + \epsilon) \times Y$ is transverse to $\{1\} \times V$ and crossing Λ with this foliation realises Λ as a Lagrangian slice. We have proved

Lemma 2.1. *Let $i : \Lambda^{n-1} \to Y^{2n-1}$ be an embedding. The Λ is a Lagrangian slice iff $di(T\Lambda) \cap \ker d\alpha = \{0\}$ and $i^*\alpha$ is closed.*

Remark 2.2. Observe that when Y is of dimension 3, then these conditions reduce to asking that Λ is not tangent to the Reeb foliation which is a generic condition.

Since for a Lagrangian slice $i^*\alpha$ is closed, one can define the notion of *exact* Lagrangian slice by requiring $i^*\alpha$ to be exact (for instance, the non-collarable examples from [4], [18] and [13] are exact). This implies that the Lagrangian $(1 - \epsilon, 1 + \epsilon) \times \Lambda$ is exact. This notion is invariant under modifications of the Liouville structure by Hamiltonian vector fields.

Explicit occurence of this construction appears in [21] and [18] in the particular case of the standard symplectic 3-dimensional case where we see that up to translation in the Reeb direction the shape of a Lagrangian is determined by the symplectic reduction of its slices.

3. Deformation of Liouville vector fields near the boundary

In this section, we study how to deform Liouville vector fields, so that they remain transverse to a given contact hyper-surface, and such computations are standard and appear in different forms in the literature; see for instance [6, Lemma 12.1].

Let $(W, \lambda; f)$ be a Weinstein cobordism with positive boundary $i : Y \hookrightarrow W$, such that the contact form is given by $i^*\lambda$. Let $(1-\epsilon, 1] \times Y$ be a collar neighbourhood of Y, such that the Liouville flow structure is given by $t \cdot \alpha$. Let H be a function of the form $H(t, x) = \rho(t)h(x)$ for a function h on Y and a compactly supported increasing function ρ on $(1-\epsilon, 1]$, such that $\rho(1) = 1$. We denote the forms $\lambda_H := \lambda + dH$, it is still a Liouville form for the symplectic structure on W the Liouville vector field for λ_H is $V_\lambda + X_H$. The contact hyper-surface Y is still transverse to X_{λ_H} iff at $t = 1$

$$dt(V_\lambda + X_H) > 0.$$

Since $dt(V_\lambda) = t$, this condition becomes

$$dt(X_H) > -1.$$

The symplectic form writes as $dt \wedge \alpha + t \cdot d\alpha$; thus, we have that

$$dH = h \cdot \rho' dt + \rho \cdot dh = dt(X_H)\alpha + t(X_h \iota d\alpha) - \alpha(X_H)dt.$$

Evaluating on R_α (the Reeb vector field of α) at $t = 1$, we obtain that Y is transverse to V_{λ_H} iff

$$dh(R_\alpha) > -1. \tag{1}$$

4. Collarability of Lagrangian slices

Let $i : \Lambda \to Y$ be an exact Lagrangian slice. We want now to make a deformation similar to the preceding section, such that V_{λ_H} is tangent to the Lagrangian $j : (1-\epsilon, 1+\epsilon) \times \Lambda \hookrightarrow X$ near $\{1\} \times \Lambda$. Let f be a function, such that $df = i^*\lambda$ and let γ be a Reeb chord of Λ. We define the *action* of γ to be (observe the order of 0 and 1 here)

$$a(\gamma) = f(\gamma(0)) - f(\gamma(1)).$$

Remark 4.1. This is a well-defined notion whenever $\gamma(0)$ and $\gamma(1)$ are on the same connected component of Λ (we refer to such chords as *pure*, the other type being called *mixed*). In the Legendrian case (i.e., when Λ is collarable), the action of a chord is 0 (when it is pure); therefore, the action is *not* the action of the corresponding intersection point when taking the Lagrangian projection (when it make sense), and this last quantity is what we refer in the present note as the *length* of the Reeb chord.

To have V_{λ_H} tangent to the Lagrangian (i.e., symplectically orthogonal to it), we need to have $(j^*\lambda_H) = 0$ near $\{1\} \times \Lambda$ which gives

$$j^*\lambda + d(H \circ j) = 0.$$

This prescribes the value of H near $\{1\}\times\Lambda$, in particular up to a constant $h = -f$ on $\{1\} \times \Lambda$. To extend h to a function on Y satisfying Inequality (1), we need to have that any pure chord γ of Λ satisfies

$$l(\gamma) > a(\gamma). \tag{2}$$

Any Reeb chord that violates this inequality will be called *small*, if it is not small a Reeb chord will be called *long*. We have proved

Lemma 4.2. *If a Lagrangian slice is not collarable then it has a small (pure) Reeb chords.*

Remark 4.3. Observe that for any of the deformation λ_H of Λ as in Sect. 3, the Reeb flow of the new contact form $i^*(\lambda + dH)$ is given by a reparametrisation of the original Reeb flow, since $d\alpha$ does not change (an explicit computation give another manifestation of Inequality (1)). This implies that the question of existence of Reeb chord is unchanged.

Remark 4.4. When the Lagrangian slice is not exact, one can still try to deform the Liouville form $t\alpha$ adding a *closed* 1-form. This closed form must extend the restriction of α to the slices, so there is first a cohomological condition on the embedding of Λ in Y. If this condition is satisfied, then one can find a cover of Y, such that α restricted to the lift of Λ is exact and proceed similarly as in the exact case.

5. Reeb chords of fillable slices

We are now ready to prove our main theorem.

Proof of Theorem 1.2. Let L be an exact filling of the slice Λ. If there are no small Reeb chords, then from Sect. 4, we know that we can modify the Liouville vector field, so that L becomes an exact filling of a Legendrian sub-manifold. Assuming Λ has no chords at all (from Remarks 4.3, it means that for either of the contact forms α and α_H has no Reeb chords), and then, we can positively wrap a small deformation of L without introducing any intersection point outside the domain X. We can use this wrapped copy to compute Wrapped Floer homology of L from [1] (either wrapping all at once, or using a direct limit depending what is our favourite definition of $WF(L)$). This implies that no high energy intersection points are involved in the computation of the wrapped Floer homology; therefore, it is isomorphic to the infinitesimally wrapped Floer homology, that is

$$WF(L) = H^*(L).$$

However, since W is subcritical, it has vanishing symplectic homology, $SH(W) = 0$ (see [22]). This is a contradiction as $WF(L)$ is a module over $SH(W)$ (see [20, Theorem 6.17]), and thus, $WF(L) = 0$.

This covers the collarable case, the other case is covered by Lemma 4.2, and therefore, the proof is complete. □

6. Concluding remarks

The proof of Theorem 1.2 applies to more general context depending on the tools one wants to use and the information we have on the slice or the filling to conclude existence of Reeb chords.

Obviously, one can replace the subcritical condition of W by the algebraic condition that $SH(W) = 0$. Also if $SH(W) \neq 0$, one can instead require that the Lagrangian filling L does not intersect the Lagrangian skeleton of W, since such a condition guarantees that $WF(L) = 0$ by Viterbo's functoriality (see [7, Theorem 9.11] or [1, Section 5]).

We can also hope to relax the vanishing of $SH(W)$ to some more specific case, as long as one can prove that $WF(L) \neq H^*(L)$, we can deduce the existence of the Reeb chords. For that purpose, the structural results from [7] can reveal to be useful.

In another direction, we can relax the fillability condition using [5] by only requiring that the slice is at the top of a Lagrangian cobordism with bottom being a Legendrian knot that admits an augmentation

It would be interesting if one could use considerations from Sects. 3 and 4 to investigate the regular Lagrangian question (and therefore the question of decomposability of Lagrangian cobordisms).

In [2], Arnol'd conjectured that any Legendrian in the standard 3-sphere admits a Reeb chord for any Reeb vector field. This version was proved in [19] (for all standard contact spheres) and in [16] [17] (for any Legendrian in any 3-dimensional contact manifold). In the discussion following his conjecture, he also discusses that some homological bounds might exist for the number of such Reeb chords, and some instance of this can be found in [10] and [9] where some estimate is given in terms of the topology of the Legendrian or some of its Lagrangian fillings. The existence result from Theorem 1.2 falls in between: the collarable case indeed gives the usual estimate coming from wrapped Floer homology or linearised contact homology, but the obstruction to collarability only provides one chord. Whether or not more can be found in the non-collarable case is unknown to the author.

Acknowledgements

Despite the short length of this note, I have been casually talking about this observation for quite a while. I want to acknowledge discussions with Vincent Colin, Kyle Hayden, Emmy Murphy, and Egor Shelukhin that lead me to write it. I also want to thank an anonymous referee whose comments helped to clarify and improve the paper in many places. I am partially supported by the ANR projects QUANTACT (ANR-16-CE40-0017) and ENUMGEOM (ANR-18-CE40-0009).

References

[1] Abouzaid, M., Seidel, P.: An open string analogue of Viterbo functoriality. Geom. Topol. **14**(2), 627–718 (2010)

[2] Arnold, V.: First steps of symplectic topology. In: VIIIth International Congress on Mathematical Physics (Marseille, 1986), pp. 1–16. World Scientific Publishing, Singapore (1987)

[3] Chantraine, B.: Lagrangian concordance of Legendrian knots. Algebr. Geom. Topol. **10**(1), 63–85 (2010)

[4] Chantraine, B.: Some non-collarable slices of Lagrangian surfaces. Bull. Lond. Math. Soc. **44**(5), 981–987 (2012)

[5] Chantraine, B., Dimitroglou Rizell, G., Ghiggini, P., Golovko, R.: Floer theory for Lagrangian cobordisms. J. Differ. Geom. **114**(3), 393–465 (2020)

[6] Cieliebak, K., Eliashberg, Y.: From Stein to Weinstein and back. Symplectic geometry of affine complex manifolds. Colloquium Publications American Mathematical Society, vol. 59, p. 354. American Mathematical Society (AMS), Providence (2012)

[7] Cieliebak, K., Oancea, A.: Symplectic homology and the Eilenberg-Steenrod axioms. Algebr. Geom. Topol. **18**(4), 1953–2130 (2018)

[8] Cornwell, C., Ng, L., Sivek, S.: Obstructions to Lagrangian concordance. Algebr. Geom. Topol. **16**(2), 797–824 (2016)

[9] Dimitroglou, G.R., Golovko, R.: On homological rigidity and flexibility of exact Lagrangian endocobordisms. Int. J. Math. **25**, 1450098 (2014)

[10] Ekholm, T., Etnyre, J., Sullivan, M.: The contact homology of Legendrian submanifolds in $\mathbb{R}^{2n+1}$. J. Differ. Geom. **71**(2), 177–305 (2005)

[11] Ekholm, T., Honda, K., Kálmán, T.: Legendrian knots and exact Lagrangian cobordisms. J. Eur. Math. Soc. (JEMS) **18**(11), 2627–2689 (2016)

[12] Eliashberg, Y., Hofer, H., Salamon, D.: Lagrangian intersections in contact geometry. Geom. Funct. Anal. **5**(2), 244–269 (1995)

[13] Eliashberg, Y., Murphy, E.: Lagrangian caps. Geom. Funct. Anal. **23**(5), 1483–1514 (2013)

[14] Eliashberg, Y., Ganatra, S., Lazarev, O.: Flexible Lagrangians. Int. Math. Res. Not. IMRN **8**, 2408–2435 (2020)

[15] Golla, M., Juhász, A.: Functoriality of the EH class and the LOSS invariant under Lagrangian concordances. Algebr. Geom. Topol. **19**(7), 3683–3699 (2019)

[16] Hutchings, M., Taubes, C.H.: Proof of the Arnold chord conjecture in three dimensions 1. Math. Res. Lett. **18**(2), 295–313 (2011)

[17] Hutchings, M., Taubes, C.H.: Proof of the Arnold chord conjecture in three dimensions. II. Geom. Topol. **17**(5), 2601–2688 (2013)

[18] Lin, F.: Exact Lagrangian caps of Legendrian knots. J. Symplectic Geom. **14**(1), 269–295 (2016)

[19] Mohnke, K.: Holomorphic disks and the chord conjecture. Ann. Math. **154**(1), 219–222 (2001)

[20] Ritter, A.: Topological quantum field theory structure on symplectic cohomology. J. Topol. **6**(2), 391–489 (2013)

[21] Sauvaget, D.: Curiosités lagrangiennes en dimension 4. Ann. Inst. Fourier (Grenoble) **54**(6), 1997–2020 (2005)

[22] Viterbo, C.: Functors and computations in Floer homology with applications. I. Geom. Funct. Anal. **9**(5), 985–1033 (1999)

Baptiste Chantraine
CNRS, Laboratoire de Mathématiques Jean Leray, LMJL
UMR 6629, Nantes Université
F-44000 Nantes
France
e-mail: `baptiste.chantraine@univ-nantes.fr`

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:31
https://doi.org/10.1007/s11784-022-00941-0
Published online April 8, 2022

Journal of Fixed Point Theory and Applications

Basic facts and naive questions

Marc Chaperon

To Claude Viterbo for his 60th birthday.

Abstract. We try to find geometric reasons for KAM theorems and such.

Mathematics Subject Classification. Primary 99Z99, Secondary 00A00.

Keywords. V.I. Arnold, unfoldings, implicit functions.

I first met Claude at the seminar on contact and symplectic geometry organised from 1982 on by Daniel Bennequin at the École normale supérieure. It was much oriented towards the beautiful conjectures V.I. Arnold had stated in the mid-60s, inspired by Poincaré's "last geometric theorem." What made the seminar seminal[1] is that its beginning coincided with the first breakthrough in that direction: at the end of 1982, Charles Conley and Eduard Zehnder proved [21] the conjecture on fixed points of Hamiltonian transformations of the standard symplectic $2n$-torus stated in [2, Appendix 9].[2]

It so happened that, the summer before, I had thought about this conjecture, seen how to deduce it from another statement about exact Lagrangian isotopies of the zero section in $T^*\mathbb{T}^n$ and proved a symplectic isotopy extension lemma [7] implying that such an isotopy extends to a compactly supported Hamiltonian isotopy of the ambient space. Almost immediately after reading the preprint of [21], I adapted the Conley–Zehnder proof to get [7] the more general statement, of which I had just learned that a slightly less precise form had also been conjectured by Arnold [1,3].

The two weeks spent on the proof of this Arnold conjecture brought me more recognition than the two years of very hard work on my 1980 *thèse*

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

[1]Besides its audience, that comprised Michèle Audin, Abbas Bahri, Alain Chenciner, Nicole Desolneux-Moulis, Ivar Ekeland, Albert Fathi, Michel Herman, Misha Gromov, François Laudenbach, Jean-Claude Sikorav—plus the author and, soon, Claude Viterbo...

[2]Their simple, functional-analytic proof did not make our seminar unanimously happy: not only had the topologists been dreaming of something more geometric, but Conley and Zehnder had proved the conjecture without knowing that it existed: when telling John Mather about their recent work, they had mentioned tori as a side remark, and Mather had informed them that they had solved a famous problem.

d'État [10,11,15]: soon after the Bourbaki seminar [7], I lectured on the Arnold conjectures in Marseille–Luminy and published with Edi Zehnder an expanded version [18] of these lectures. Having felt ill at ease when teaching the fact (established by Amann and Zehnder) that the *non* C^2 infinite dimensional action functional, once reduced to finite dimensions, is as smooth as the Hamiltonian of the isotopy, I found it urgent to design a purely finite dimensional proof of "my" Arnold conjecture; this again took me two weeks [8,9] and brought me much more recognition.

François Laudenbach liked this new proof; he had an extremely bright (and nice) student of his, Jean-Claude Sikorav, work on its generalisations and consequences. Jean-Claude first proved with François [27] what had been the true aim of [8,9], namely, the extension of my result from the cotangent bundle of the torus to that of an arbitrary closed manifold.[3] He then noticed that formula (7) in [8] means that my discretised action is a generating phase for the deformed Lagrangian submanifold, and extended this to arbitrary closed manifolds [36]. In [37], he generalised the result to Lagrangian immersions and gave an easy proof of the Arnold conjecture on fixed points in situations including surfaces, first obtained more painfully in [35].

At about the same time, the 25-year-old Claude, who had been the student of Laudenbach and Ekeland, solved [39] a big problem: the Weinstein conjecture in $\mathbb{R}^{2n}$. Oddly enough, he did not use Jean Claude's generating phases, with which he would soon do wonders [40].

The last of my favourite Arnold conjectures had been proved [25] via holomorphic disks, and Floer theory[4] had taken off [24], leaving me with my fear of flying. My belief that some room was left for earthlier methods[5] now rested mostly on Claude's shoulders. He did not disappoint me.

1. Subharmonic bifurcations in real or complex dimension one

We first recall the simplest case of the most basic fact.

1.1. The period doubling bifurcation

Let $h : (u, x) \mapsto h_u(x)$ be a C^k local map $(\mathbb{R}^2, 0) \to (\mathbb{R}, 0)$, $k \geq 2$, such that the derivative $h_0'(0) = \partial_x h(0,0)$ equals -1.

The fixed point 0 of h_0 is *robust*, meaning that every h_u with u small enough has a fixed point $\varphi(u)$ nearby, depending C^k on the parameter u:

Proposition 1.1. *The fixed points of the* unfolding $\tilde{h} : (u, x) \mapsto \big(u, h_u(x)\big)$ *form near* 0 *the graph of a* C^k *function* $\varphi : (\mathbb{R}, 0) \to (\mathbb{R}, 0)$.

Proof. This follows from the implicit function theorem applied to the C^k equation $F(u, x) := x - h_u(x) = 0$, as $F(0,0) = 0$ and $\partial_x F(0,0) = 2 \neq 0$. □

Of course 0 is a 2-*periodic point of* h_0, i.e. a fixed point of $h_0^2 := h_0 \circ h_0$.

[3] A little sooner, Helmut Hofer had done this [26] in an infinite dimensional framework. Whatever its proof, I thought the result would be more central than it turned out to be.
[4] Whose idea owes as much to Charlie Conley as to Edi Zehnder, see the foreword of [12].
[5] A similar belief is at the origin of the present article.

Proposition 1.2. *Assume that* $\alpha(u) := {h_u}'\big(\varphi(u)\big)$ *satisfies* $\alpha'(0) \neq 0$*. Then, near* 0*, the* 2*-periodic points of* $\tilde{h}$ (solutions of $h_u^2(x) = x$) *form the union of two curves intersecting transversally at* 0*: of course* $\operatorname{graph}\varphi$*, and a* C^{k-1} *curve* W *of which* $\tilde{h}|_W$ *is an involution, implying that* T_0W *is the* x*-axis.*

Proof. Conjugating $\tilde{h}$ by the local diffeomorphism $(u,x) \mapsto \big(u, x-\varphi(u)\big)$, we may assume $\varphi = 0$—the new ${h_u}'(0)$ is the old ${h_u}'\big(\varphi(u)\big)$.

By Taylor's formula, $h_u(x) = xg_u(x)$ near 0, where $g_u(x) = \int_0^1 {h_u}'(tx)\,dt$, hence $g_u(0) = \alpha(u)$, and therefore, $g_0(0) = -1$; the map $g : (u,x) \mapsto g_u(x)$ is C^{k-1} and the equation $h_u^2(x) = x$ writes $xg_u(x)g_u\big(h_u(x)\big) = x$, which means $x = 0$ (the fixed points) or $G(u,x) := g_u(x)g_u\big(h_u(x)\big) - 1 = 0$. As $G(0,0) = 0$ and $\partial_u G(0,0) = \frac{d}{du}\big|_{u=0} g_u(0)^2 = -2\alpha'(0) \neq 0$, there exist open neighbourhoods U, V of 0 in $\mathbb{R}$ such that the zeros of $G|_{U\times V}$ form the "graph" $W = \{u = \psi(x)\}$ of a C^{k-1} implicit function $\psi : V \to U$.

The map $\tilde{h}$ is by definition an involution of its set of 2-periodic points, of which $W\backslash\{0\}$ is an open subset, which becomes $\tilde{h}$-invariant if W is replaced by $W \cap \tilde{h}(W)$ (this means restricting conveniently the open subset V). Invariance writes $\psi(x) = \psi\big(h_{\psi(x)}(x)\big)$, hence $\psi'(0) = \lim_{x\to 0} \frac{\psi\left(h_{\psi(x)}(x)\right) - \psi(x)}{h_{\psi(x)}(x) - x} = 0$ since the only fixed point of $\tilde{h}$ lying in W is 0; thus, T_0W is the x-axis. $\square$

Example. The curve W can have various positions with respect to T_0W:

- If $h_u(x) = \alpha(u)x$, where $\alpha : (\mathbb{R}, 0) \to (\mathbb{R}, -1)$ is a C^k function with $\pm\alpha'(0) > 0$, then W is the x-axis; the fixed point 0 of h_u is attracting for $\pm u < 0$, repulsing for $\pm u > 0$, and this cannot be called a bifurcation.
- If $h_u(x) = -(1 + u - x^2)x$, then the fixed point 0 of h_u is attracting for $u < 0$, repulsing for $u > 0$, and W is the parabola $u = x^2$; for $u > 0$, the attracting 2-periodic orbit $\{-\sqrt{u}, \sqrt{u}\}$, born for $u = 0$, takes the place of 0 as an attractor of h_u, a genuine *bifurcation.*
- If $h_u(x) = -(1 + u + x^2)x$ then, for $u < 0$, the repulsing 2-periodic orbit $\{-\sqrt{-u}, \sqrt{-u}\}$ gradually "throttles" the attracting fixed point 0, so that for $u \geq 0$ no attractor of h_u persists near 0, a true *catastrophe.*

The generic situations look like the last two examples (Fig. 1).

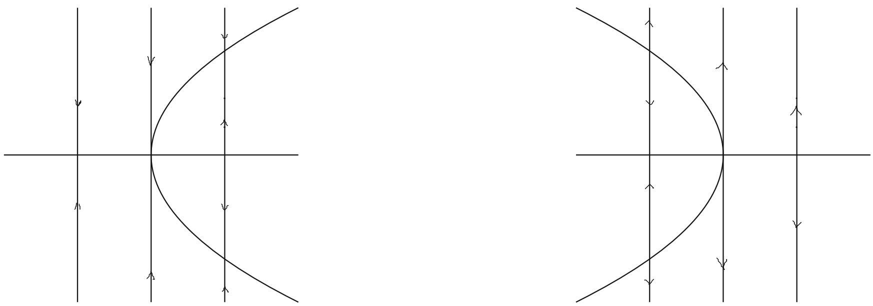


Figure 1. Bifurcation and catastrophe

1.2. Subharmonic bifurcations, holomorphic case

Proposition 1.3. *Let $h : (u, z) \mapsto h_u(z)$ be a local map $(\mathbb{C}^2, 0) \to (\mathbb{C}, 0)$, holomorphic and such that $h_0'(0) = \partial_z h(0, 0)$ is a qth root of unity $\rho = e^{2\pi i \frac{p}{q}}$, $0 < p < q$. Then,*

(i) *The fixed points of the unfolding $\tilde{h} : (u, z) \mapsto \big(u, h_u(z)\big)$ form near 0 the graph of a holomorphic function $\varphi : (\mathbb{C}, 0) \to (\mathbb{C}, 0)$.*

(ii) *Assume that $\alpha(u) := {h_u}'\big(\varphi(u)\big)$ satisfies $\alpha'(0) \neq 0$. Then, near 0, the q-periodic points of $\tilde{h}$* (solutions (u, z) of the equation $h_u^q(z) = z$) *form the union of two holomorphic curves intersecting transversally at 0: the fixed point set* graph φ *and a curve W on which $\tilde{h}$ induces a $\mathbb{Z}/q\mathbb{Z}$-action.*

(iii) *When ρ is a* primitive *qth root of unity, the curve W and the z-plane are tangent to order $q - 1$ at 0.*

Proof. (i) As $h_0'(0) = \rho \neq 1$, just apply the holomorphic implicit function theorem to the equation $z - h(u, z) = 0$.

(ii) As in the proof of Proposition 1.2, one can assume $\varphi = 0$, hence $h_u(z) = z\, g(u, z)$ with g holomorphic this time and $g(u, 0) = \alpha(u)$. The equation $h_u^q(z) = z$ writes $z g_u(z) g_u\big(h_u(z)\big) \cdots g_u\big(h_u^{q-1}(z)\big) = z$, which means either $z = 0$ (the fixed points) or

$$G(u, z) := g_u(z) g_u\big(h_u(z)\big) \cdots g_u\big(h_u^{q-1}(z)\big) - 1 = 0.$$

As $G(u, 0) = \alpha(u)^q - 1$ and $\alpha(0) = \rho$, one has $\partial_u G(0) = q\rho^{q-1}\alpha'(0) \neq 0$ and $G(0) = 0$, hence there exist open neighbourhoods U, V of 0 in $\mathbb{C}$ such that the zeros of $G|_{U\times V}$ form the "graph" $W = \{u = \psi(z)\}$ of a holomorphic implicit function $\psi : V \to U$.

The map $\tilde{h}$ induces by definition an action of $\mathbb{Z}/q\mathbb{Z}$ on its set of q-periodic points, of which $W \setminus \{0\}$ is an open subset, that becomes $\tilde{h}$-invariant if W is replaced by $W \cap \tilde{h}^{-1}(W) \cap \cdots \cap \tilde{h}^{1-q}(W)$ (as before, this means restricting conveniently the open subset V). Invariance writes

$$\psi(z) = \psi\Big(h\big(\psi(z), z\big)\Big). \tag{1.1}$$

(iii) Still assuming $\varphi = 0$, if we derivate (1.1), we get

$$\psi'(z) = \psi'\Big(h\big(\psi(z), z\big)\Big)\Big(\partial_1 h\big(\psi(z), z\big)\psi'(z) + {h_{\psi(z)}}'(z)\Big).$$

For $z = 0$, as the identity $h(u, 0) = 0$ implies that $\partial_1 h(u, 0) = 0$, this reads

$$\psi'(0) = \psi'(0)\, h_0'(0), \text{ that is, } (\rho - 1)\psi'(0) = 0, \text{ hence } \psi'(0) = 0,$$

which proves our result if $q = 2$. Otherwise assuming inductively that ψ vanishes to order $k-1$ at 0 for some $k \in \{2, \ldots, q-1\}$ and derivating k times (1.1) at 0, the Faà di Bruno formula and the identity $\partial_1 h(u, 0) = 0$ yield

$$\psi^{(k)}(0) = \psi^{(k)}(0)\, h_0'(0)^k, \quad \text{that is,} \quad (\rho^k - 1)\, \psi^{(k)}(0) = 0,$$

hence $\psi^{(k)}(0) = 0$ as ρ is a *primitive* qth root of unity. □

Example. If $h_u(z) = \alpha(u)z$, where $\alpha : (\mathbb{C}, 0) \to (\mathbb{C}, \rho)$ is a holomorphic function such that $\alpha'(0) \neq 0$, then W is the z-plane.

If $h_u(z) = (\rho + u - z^q)z$, then W is the curve $u = z^q$.

1.3. Opening Pandora's box

Under the hypotheses of Proposition 1.3 (ii)–(iii), α is a holomorphic local diffeomorphism $(\mathbb{C},0) \to (\mathbb{C},\rho)$. Viewing it as a local parameter change and performing the variable changes in the proof of Proposition 1.3, the following hypotheses are verified with $u_0 = \rho$:

Hypotheses Given $u_0 \in \mathbb{S}^1$, set $\tilde{u}_0 := (u_0, 0) \in \mathbb{C}^2$ and let $h : (u,z) \mapsto h_u(z)$ be a holomorphic local map $(\mathbb{C}^2, \tilde{u}_0) \to (\mathbb{C},0)$ such that $h_u(0) = 0$ and ${h_u}'(0) = u$. Proposition 1.3 now reads as follows:

Proposition 1.4. *If $u_0 = e^{2\pi i \frac{p}{q}}$, $0 < p < q$, $\gcd(p,q) = 1$, then the q-periodic points of $\tilde{h}$ near $\tilde{u}_0$ form the union of $\{z = 0\}$ and the $\tilde{h}$-invariant "graph" $W_{p/q} = \{u = \psi_{p/q}(z)\}$ of a holomorphic $\psi_{p/q} : (\mathbb{C},0) \to (\mathbb{C},u_0)$ such that ${\psi_{p/q}}^{(j)}(0) = 0$ for $1 \leq j < q$. The function $\psi := \psi_{p/q}$ verifies (1.1), and $\tilde{h}$ generates a $\mathbb{Z}/q\mathbb{Z}$-action on $W_{p/q}$, namely*

$$\big(m, \big(\psi(z), z\big)\big) \longmapsto \tilde{h}^m\big(\psi(z), z\big) = \big(\psi(z), {h_{\psi(z)}}^m(z)\big)\,,$$

induced by the $\mathbb{Z}/q\mathbb{Z}$-action $(m,z) \mapsto {h_{\psi_{p/q}(z)}}^m(z)$ on $\operatorname{Dom}\psi_{p/q}$. □

When u_0 is not a root of unity, the following result can apply to $f = h_{u_0}$:

Theorem 1.5. (Brjuno [5,6], Yoccoz [41]) *If $u_0 = e^{2\pi i\omega}$ with $\omega \in [0,1] \setminus \mathbb{Q}$, the following two conditions are equivalent:*

(i) *ω is a* Brjuno number, *meaning that the convergents $\frac{p_n}{q_n}$ of its continued fraction expansion verify $\sum \frac{\log q_{n+1}}{q_n} < \infty$.*

(ii) *Every holomorphic germ $f : (\mathbb{C},0) \to (\mathbb{C},0)$ such that $f'(0) = u_0$ is holomorphically linearisable.* □

Notes. The implication (i)⇒(ii) is Brjuno's. In 1942, Siegel [33] had proved (ii) under the stronger condition $\sup \frac{\log q_{n+1}}{\log q_n} < \infty$. This already defines a full measure set of numbers $u_0 \in \mathbb{S}^1$, but Theorem 1.5 provides the *optimal* set.

Back to families, in the trivial case $h_u(z) = uz$, every h_u is linear(isable). However, in general, h_{u_0} is linearisable if $u_0 = e^{2\pi i\omega}$ with ω Brjuno.

In that case, linearisability means that there exists a holomorphic local coordinate (conjugacy) $Z_\omega : (\mathbb{C},0) \to (\mathbb{C},0)$ such that $Z_\omega \circ h_{u_0} = u_0 Z_\omega$; as the rotation $z \mapsto u_0 z$ preserves each circle $S_r = \{|z| = r\}$, every closed curve $C_r = Z_\omega^{-1}(S_r)$ with $r > 0$ small enough is h_{u_0}-invariant and, of course, $Z_\omega|_{C_r}$ conjugates $h_{u_0}|_{C_r}$ to the rotation $z \mapsto e^{2\pi i\omega}z$ restricted to S_r.

Question 1.6. *Is this the limit of what happens near $u = e^{2\pi i p_n/q_n}$? Do the holomorphic functions ψ_{p_n/q_n} tend to the constant $\psi_\omega = u_0 = e^{2\pi i\omega}$ in some uniform neighbourhood of 0 and, for $z \in \mathbb{C}$ close to 0, do the periodic orbits $\Big\{\Big(\psi_{p_n/q_n}(z), {h_{\psi_{p_n/q_n(z)}}}^k(z)\Big) : 0 \leq k < q_n\Big\}$ of $\tilde{h}$ tend to the closed $\tilde{h}$-invariant curve $\{u_0\} \times C_r$ such that $z \in C_r$?*

More precisely, does the (holomorphic) standard linearisation[6]

$$Z_{p_n/q_n}(z) = \frac{1}{q_n}\sum_{k=0}^{q_n-1} e^{-2\pi i k p_n/q_n} h_{\psi_{p_n/q_n}(z)}{}^k(z)$$

of the $\mathbb{Z}/q_n\mathbb{Z}$*-action* $(m,z) \mapsto h_{\psi_{p_n/q_n}(z)}{}^m(z)$ *tend to* Z_ω *when* $n \to \infty$*?*

Notes. If $h_u(z) = uz$, the answer is trivially positive even when ω is not Brjuno. The question is whether this holds for arbitrary families h.

My hope would be to deduce the Siegel–Brjuno theorem from the uniform convergence of ψ_{p_n/q_n} and maybe Z_{p_n/q_n} in a uniform neighbourhood of 0, at least for some well-chosen family h. One might get invariant fractals at the limit when ω is irrational but not Brjuno, as in [20]—the Pérez-Marco hedgehogs [29], independent of any arithmetic conditions, might be obtained in this fashion.

2. Subharmonic bifurcations, Arnold tongues and KAM circles

Here, *smooth* means real analytic or C^∞.

2.1. Subharmonic bifurcations in real dimension two

Let $h : (u,z) \mapsto h_u(z)$ be a smooth local map $(\mathbb{R}^2 \times \mathbb{R}^2, 0) \to (\mathbb{R}^2, 0)$ such that the eigenvalues of the derivative $Dh_0(0) = \partial_z h(0,0)$ are primitive qth roots of unity $\rho = e^{2\pi i \frac{p}{q}}, \bar\rho = e^{-2\pi i \frac{p}{q}}$, $1 \le p < q$, $q \ge 3$.

Proposition 2.1. (i) *The fixed points of the unfolding* $\tilde h : (u,z) \mapsto \big(u, h_u(z)\big)$ *form near* 0 *the graph of a smooth function* $\varphi : (\mathbb{R}^2, 0) \to (\mathbb{R}^2, 0)$*.*

(ii) *There is a smooth local function* $\alpha : (\mathbb{R}^2, 0) \to (\mathbb{C}, \rho)$ *such that the eigenvalues of* $Dh_u\big(\varphi(u)\big)$ *are* $\alpha(u), \overline{\alpha(u)}$*.*

(iii) *If* $D\alpha(0) : \mathbb{R}^2 \to \mathbb{C}$ *is bijective then, near* 0*, the* q*-periodic points of* $\tilde h$ *form the union of two surfaces intersecting transversally at* 0*: of course* graph φ*, plus a* C^{q-3} *surface* W *on which* $\tilde h|_W$ *induces a* $\mathbb{Z}/q\mathbb{Z}$*-action.*

Proof. (i) follows from the implicit mapping theorem applied to the smooth equation $F(u,z) := z - h(u,z) = 0$, as $\partial_z F(0,0) : \mathbb{R}^2 \to \mathbb{R}^2$ is invertible.
(ii) follows from the formula for the eigenvalues of a real 2×2 matrix with no real eigenvalue.
(iii) We may assume $\varphi = 0$, and the new $Dh_u(0)$ is the old $Dh_u\big(\varphi(u)\big)$.

Lemma 2.2. (a) *An* $\mathbb{R}$*-linear change of variables* $J(u) : \mathbb{R}^2 \to \mathbb{C}$*, depending smoothly on* u*, yields* $h : (\mathbb{R}^2 \times \mathbb{C}, 0) \to (\mathbb{C}, 0)$ *and* $Dh_u(0)z = \alpha(u)z$*.*

(b) *Modulo a change of variables, polynomial of degree* $q-1$ *with respect to* $z, \bar z$*, whose coefficients are smooth functions of* u*, the Taylor polynomial* Q_u *of* h_u *to order* $q-1$ *at* 0 *for small* u *is of the form*

$$Q_u(z) = z\left(\alpha(u) + \sum_{k=1}^{[\frac{q-1}{2}]} b_k(u)|z|^{2k}\right) + \beta(u)\bar z^{q-1}.$$

[6]The other holomorphic local linearisations Z satisfy $Z \circ Z^{-1}_{p_n/q_n}(z) = za(z^{q_n})$, $a(0) \neq 0$.

Proof of the lemma. (a) The isomorphism $J(u) \in L(\mathbb{R}^2, \mathbb{C})$ is an eigenvector of $Dh_u(0)^T : \lambda \mapsto \lambda \circ Dh_u(0)$ associated to the eigenvalue $\alpha(u)$. Under the condition, e.g. $J(u)(1,0) = 1$, it is unique and depends smoothly on u.

(b) By normal form theory or direct computation, one can assume that $Q_u(z) - \alpha(u)z$ is a $\mathbb{C}$-linear combination (depending smoothly on u) of monomials $z^j \bar{z}^k$ with $1 < j + k \leq q-1$ and $u_0^j \bar{u}_0^k = u_0$, that is, $e^{2\pi i (j-k-1)p/q} = 1$, which writes $(j-k-1)p = \ell q$ with $\ell \in \mathbb{Z}$. As $\gcd(p,q) = 1$, one has $\ell = mp$, $m \in \mathbb{Z}$, hence $j - k - 1 = mq$ and either $m = 0$, hence $z^j \bar{z}^k = z|z|^{2k}$, or $m = -1$ and $j = 0$, yielding $z^j \bar{z}^k = \bar{z}^{q-1}$. $\diamond$

By Taylor's formula,

$$
\begin{aligned}
h_u(z) &= Q_u(z) + \sum_{j=0}^{q} z^j \bar{z}^{q-j} \int_0^1 \frac{(1-t)^{q-1}}{(q-1)!} \binom{q}{j} \partial_z^j \partial_{\bar{z}}^{q-j} h_u(tz)dt \\
&= z\left(a(u,z) + b(u,z)\frac{\bar{z}^{q-1}}{z}\right)
\end{aligned}
$$

where a, b are smooth, $a(u,0) = \alpha(u)$ and $b(u,0) = \beta(u)$. It follows that $h_u(z) = zg_u(z)$ with $g : (u,z) \mapsto g_u(z)$ only C^{q-3} in general and $g_u(0) = \alpha(u)$.

- For $q > 3$, the same arguments as in the proof of Proposition 1.3 yield a C^{q-3} implicit function $\psi : (\mathbb{C}, 0) \to (\mathbb{R}^2, 0)$ whose graph W has the required properties near the origin—in particular, (1.1) holds.
- If $q = 3$, then $h_u(z) = A_u(z)z$ near 0, where $A_u(z) = \int_0^1 Dh_u(tz)\,dt$ (hence $A_u(0)z = \alpha(u)z$), and one can similarly apply the implicit map theorem along $r = 0$ after dividing by r the equation $h_u^3(re^{i\theta}) = re^{i\theta}$.[7]

The details are left to the reader. □

Example. If $h_u(z) = (\rho + u)z - \bar{z}^{q-1}$, $u, z \in \mathbb{C}$, then W is the surface $u = \bar{z}^{q-1}/z$, which is C^{q-3} but not C^{q-2}. Thus, our bound for the differentiability of W is sharp. No such problem arised in the holomorphic case.

2.2. Arnold tongues

Under the hypotheses of Proposition 2.1 (iii), α is a smooth local diffeomorphism $(\mathbb{R}^2, 0) \to (\mathbb{C}, \rho)$. Viewing it as a local parameter change, the following hypotheses are verified with $u_0 = \rho$, modulo the variable changes in the proof of Proposition 2.1:

Hypotheses. For $u_0 \in \mathbb{S}^1$, setting $\tilde{u}_0 := (u_0, 0) \in \mathbb{C}^2$, let $h : (u,z) \mapsto h_u(z)$ be a smooth local map $(\mathbb{C}^2, \tilde{u}_0) \to (\mathbb{C}, 0)$ such that $h_u(0) = 0$ and $Dh_u(0)z = uz$. Proposition 2.1 now reads as follows:

Proposition 2.3. *If $u_0 = e^{2\pi i \frac{p}{q}}$, $0 < p < q$, $\gcd(p,q) = 1$, then, near $\tilde{u}_0$, the q-periodic points of $\tilde{h}$ form the union of $\{z = 0\}$ and the $\tilde{h}$-invariant "graph" $W_{p/q} = \{u = \psi_{p/q}(z)\}$ of a C^{q-3} function $\psi_{p/q} : (\mathbb{C}, 0) \to (\mathbb{C}, u_0)$. The function $\psi := \psi_{p/q}$ verifies (1.1), and $\tilde{h}|_{W_{p/q}}$ generates a $\mathbb{Z}/q\mathbb{Z}$-action on $W_{p/q}$ as in Proposition 1.4.* □

[7] The "blown-up" surface $\breve{W} = \mathrm{graph}\,\breve{\psi}$ in polar coordinates is smooth, see Sect. 3.

The functions $\psi = \psi_{p/q}$ of Proposition 1.4, being holomorphic, are either constant or *open.* Thus, the invariant manifold $W_{p/q}$, projected into parameter space, is either $\{e^{2\pi ip/q}\}$ or (in general) an open neighbourhood of $e^{2\pi ip/q}$. The non-holomorphic case is altogether different:

Proposition 2.4. *If u_0 is a primitive qth root of unity $e^{2\pi ip/q}$, $0 < p < q$, $q > 4$, then:*

(i) *Up to a smooth local change of variables $\big(\mathbb{C}^2, \tilde u_0\big) \to \big(\mathbb{C}^2, \tilde u_0\big)$, of the form $(u,z) \mapsto (u, Z_u(z))$ with Z_u (real) polynomial of degree $q-2$, one has the following: near $\tilde u_0$, the unfolding $\tilde h$ is tangent to order $q-2$ along $\mathbb{C}\times\{0\}$ to a smooth unfolding $\tilde P(u,z) = (u, P_u(z))$ of the form*

$$P_u(z) = z\left(u - \sum_{k=1}^{\left[\frac{q-3}{2}\right]} b_k(u)|z|^{2k}\right).$$

(ii) *For $b_1(u_0) \neq 0$, the principal part of $\psi_{p/q}(z)$ is $u_0 + b_1(u_0)|z|^2$. Thus, for $\Re\big(\bar u_0 b_1(u_0)\big) \neq 0$, the set $\operatorname{Im}\psi_{p/q}$ of those u near u_0 for which h_u has a q-periodic orbit lies on* one *side of* $\mathbb{S}^1$.

(iii) *The function $\psi_{p/q}$ is tangent to order $q-3$ at $\tilde u_0$ to a normal form*

$$\hat\psi_{p/q}(z) = u_0 + \sum_{k=1}^{\left[\frac{q-3}{2}\right]} a_k|z|^{2k} =: \chi_{p/q}(|z|), \qquad a_k \in \mathbb{C}, \quad a_1 = b_1(u_0).$$

Thus, when the first Birkhoff invariant $b_1(u_0)$ *is nonzero, restricting* $\operatorname{Dom}\psi_{p/q}$ *if necessary, the set $\operatorname{Im}\psi_{p/q}$ is contained near u_0 in an "Arnold tongue" $\bigcup_{0\le t\le\varepsilon}\big\{u\in\mathbb{C} : |u-\chi_{p/q}(t)| \le \delta_\varepsilon t^{q-3}\big\}$ along the curve $\chi_{p/q}([0,\varepsilon])$, with $\varepsilon > 0$ small and $\lim_{\varepsilon\to 0}\delta_\varepsilon = 0$.*

Proof. (i) follows from Lemma 2.2 (b).
(ii)–(iii) As $\psi_{p/q}(z) = \psi_{p/q}\Big(h\big(\psi_{p/q}(z), z\big)\Big)$ by (1.1), the Taylor polynomial

$$\hat\psi_{p/q}(z) = u_0 + \sum_{1\le j+\ell\le q-3} c_{j\ell} z^j \bar z^\ell =: u_0 + \hat c(z)$$

satisfies $\hat\psi_{p/q}(z) = \hat\psi_{p/q}\big(P_{\hat\psi_{p/q}(z)}(z)\big)$ up to terms of degree greater than $q-3$. Denoting the Taylor expansion of $b_k(u_0+v)$ at $v=0$ by

$$\hat b_k(v) = \sum_{m\ge 0} b_{kmn} v^m \bar v^n,$$

this means that, up to terms of degree greater than $q-3$,

$$\hat c(z) = \hat c\left(z\Big(u_0 + \hat c(z) - \sum_{k=1}^{\left[\frac{q-3}{2}\right]} \hat b_k\big(\hat c(z)\big)|z|^{2k}\Big)\right).$$

(ii) It follows that $c_{10} = u_0 c_{10} = 0$, $c_{01} = \bar u_0 c_{01} = 0$, $c_{20} = u_0^2 c_{20} = 0$, $c_{02} = \bar u_0^2 c_{02} = 0$; thus, the first $c_{j\ell}$ that can be nonzero is $c_{11} =: a_1$, and it is equal to $b_{100} = \hat b_1(0) = b_1(u_0)$.;

(iii) Inductively, one can see that $\hat{c}(z) = \hat{c}(u_0 z)$, hence $\hat{c}(z) = \sum_{k=1}^{\left[\frac{q-3}{2}\right]} a_k |z|^{2k}$. The reader can fill in the details as an exercise. □

Example. If $h_u(z) = z\left(u - \sum_{k=1}^{\left[\frac{q-3}{2}\right]} a_k |z|^{2k}\right)$ then $\psi_{p/q} = \hat{\psi}_{p/q}$; thus, near u_0, $\operatorname{Im}\psi_{p/q}$ is the curve $\chi_{p/q}([0,\varepsilon])$.

2.3. Opening Pandora's box wider

Question 2.5. *For diophantine ω with convergents p_n/q_n, if $b_1(u_0) \neq 0$, one can wonder as in the holomorphic case whether one has the following:*

- *The smooth functions ψ_{p_n/q_n} tend to some $\psi_\omega : (\mathbb{C}, 0) \to (\mathbb{C}, u_0)$ in a uniform neighbourhood of* 0*; thus, the $\tilde{h}$-invariant surfaces W_{p_n/q_n} tend to the $\tilde{h}$-invariant surface $W_\omega = \{u = \psi_\omega(z)\}$.*
- *For small z, the periodic orbits $\left\{\left(\psi_{p_n/q_n}(z), {h_{\psi_{p_n/q_n}(z)}}^k(z)\right) : 0 \leq k < q_n\right\}$ of $\tilde{h}$ tend to a closed $\tilde{h}$-invariant curve $\{\psi_\omega(z)\} \times C_{\omega z}$ such that $z \in C_{\omega z}$ and that the rotation number of $h_{\psi(z)}|_{C_{\omega z}}$ is ω.*
- *The standard linearisation of the $\mathbb{Z}/q_n\mathbb{Z}$-action $(m, z) \mapsto {h_{\psi_{p_n/q_n}(z)}}^m(z)$ tends to a local transformation Z_ω linearising the local diffeomorphism $z \mapsto h_{\psi_\omega(z)}(z)$. Hence, the $\mathbb{T}$-action $(\theta, z) \mapsto Z_\omega^{-1}\big(e^{2\pi i\theta} Z_\omega(z)\big)$ leaves ψ_ω invariant,*[8] *implying that* $\operatorname{Im}\psi_\omega$ *is* a curve *(with boundary), limit of the narrower and narrower Arnold tongues* $\operatorname{Im}\psi_{p_n/q_n}$.

Example (KAM invariant curves). Assume that h possesses the following properties near some $u_0 = e^{2\pi i\omega_0}$ with $\omega_0 \in \mathbb{R} \setminus \mathbb{Q}$:

(i) If $|u| = 1$, the transformation h_u preserves the area.
(ii) $h_u = |u|\, h_{u/|u|}$, hence h_u multiplies the area by $|u|^2$.
(iii) One has $b_1(u_0) \neq 0$.

By (ii), no h_u with $|u| \neq 1$ can have a closed invariant curve near 0. Thus, if the answer to Question 2.5 is positive, then every ψ_ω has modulus one, hence $b_1(u_0) = i\lambda u_0$, $\lambda \in \mathbb{R}$—which already follows from (i). Figure 2 shows what happens for $(u, z) = (e^{2\pi i\omega}, z) \in \mathbb{S}^1 \times \mathbb{C}$ close to $\tilde{u}_0$, in local coordinates (ω, z). The ω-axis is in red and the "paraboloids" are the surfaces W_ω with ω Diophantine, which do lie in $\mathbb{S}^1 \times \mathbb{C}$ as $|\psi_\omega(z)| = 1$. These surfaces intersect the slice $u = u_0$ at the h_{u_0}-invariant closed curves ("KAM circles") $C_{\omega z}$ with $\psi_\omega(z) = u_0$, which occupy most of the room near $z = 0$, with maybe complicated dynamics in between.

Notes. The limit surfaces W_ω and the linearisations Z_ω in Question 2.5 might be obtained as in [19] (where, however, the typical situation is $b_1(e^{2\pi i\omega}) \notin i\mathbb{R}$, yielding normally hyperbolic invariant circles). Figure 2, which I like a lot, most probably follows from standard KAM theory [23].

[8] Conjugating everything by Z_ω, one can assume $h_{\psi_\omega(z)}(z) = e^{2\pi i\omega} z$, hence (1.1) reads $\psi_\omega(z) = \psi_\omega(e^{2\pi i\omega} z)$, which yields $\psi_\omega(z) = \psi_\omega(e^{2\pi ik\omega} z)$ for every integer k and, therefore, $\psi_\omega(z) = \psi_\omega(e^{2\pi i\theta} z)$ for all $\theta \in \mathbb{T}$ by density.

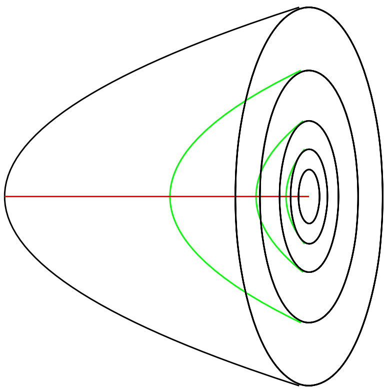


FIGURE 2. Geometry of the KAM theorem

3. Higher dimensions

3.1. Statement of the hypotheses

Hypotheses. Given $u_0 \in \mathbb{C}^d$, $d > 1$, whose components are nonzero and all different, set $\tilde{u}_0 := (u_0, 0) \in \mathbb{C}^d \times \mathbb{C}^d$ and let $h : (\mathbb{C}^d \times \mathbb{C}^d, \tilde{u}_0) \to (\mathbb{C}^d, 0)$ be a smooth local map $(u, x) \mapsto h_u(x)$ such that

$$h_u(0) = 0 \qquad \text{and} \qquad Dh_u(0) = \operatorname{diag} u : z \mapsto \big(u_1 z_1, \ldots, u_d z_d\big).$$

The case where h is holomorphic will be referred to as *the holomorphic case.*
Note. A general situation reduces to these hypotheses. Let $h : (u, x) \mapsto h_u(x)$ be a smooth local map $(\mathbb{R}^{2d} \times \mathbb{R}^{2d}, 0) \to (\mathbb{R}^{2d}, 0)$ such that the eigenvalues of $Dh_0(0)$ are simple and not real. Near 0, the fixed points of $\tilde{h}$ form the graph $x = \varphi(u)$ of a smooth implicit function, *which we may assume to be* 0.

There is [13,16] a smooth local map J of $\mathbb{R}^{2d}$ into the space of $\mathbb{R}$-linear isomorphisms $\mathbb{R}^{2d} \to \mathbb{C}^d$, defined near 0, such that each $J(u)Dh_u(0)J(u)^{-1}$ is a diagonal automorphism $\operatorname{diag} \alpha(u) : z \mapsto \big(\alpha_1(u)z_1, \ldots, \alpha_d(u)z_d\big)$ of $\mathbb{C}^d$ (thus, the eigenvalues $\alpha_j(u), \overline{\alpha_j(u)}$ of $Dh_u(0)$, $1 \le j \le d$, depend smoothly on u). *Via the identification* $(u, x) \mapsto \big(u, J(u)x\big)$*, we can view h as a local map* $(\mathbb{R}^{2d} \times \mathbb{C}^d, 0) \to (\mathbb{C}^d, 0)$ *such that* $Dh_u(0) = \operatorname{diag} \alpha(u)$.

Setting $\alpha(u) := \big(\alpha_1(u), \ldots, \alpha_d(u)\big)$ and *assuming* $D\alpha(0) : \mathbb{R}^{2d} \to \mathbb{C}^d$ *invertible*, the smooth local map $\alpha : (\mathbb{R}^{2d}, 0) \to \mathbb{C}^d$ is a local diffeomorphism. *If we view it as an identification, then* $u_0 := \alpha(0)$ *satisfies our hypotheses.*

3.2. Periodic orbits

Proposition 3.1. *Assume that* $u_0 = \rho = (\rho_1, \ldots, \rho_d)$*, where* $\rho_j = e^{2\pi i p_j/q}$*,* $0 < p_j < q$*. Let* $\pi : \big(\mathbb{R}_+ \times \mathbb{S}^{2d-1}, \{0\} \times \mathbb{S}^{2d-1}\big) \to (\mathbb{C}^d, 0)$ *be the oriented blowup* $\pi(r, y) := ry$ *("polar coordinates"). Then, setting* $\breve{u}_0 := (u_0, 0) \in \mathbb{C}^d \times \mathbb{R}_+$ *and denoting by* $\mathring{\mathbb{S}}^{2d-1}$ *the complement of the coordinate hyperplanes in* $\mathbb{S}^{2d-1}$*:*

(i) *Near* $\tilde{u}_0$*, the map h lifts to a smooth local map* $\breve{h} : (u, r, y) \mapsto \breve{h}_u(r, y)$ *of* $\big(\mathbb{C}^d \times \mathbb{R}_+ \times \mathbb{S}^{2d-1}, \{\breve{u}_0\} \times \mathbb{S}^{2d-1}\big)$ *into* $\big(\mathbb{R}_+ \times \mathbb{S}^{2d-1}, \{0\} \times \mathbb{S}^{2d-1}\big)$ *such that* $\pi \circ \breve{h}_u = h_u \circ \pi$ *and* $\breve{h}(\breve{u}_0, y) = \big(0, (\operatorname{diag} u_0)y\big)$ *for all* $y \in \mathbb{S}^{2d-1}$.

(ii) *The q-periodic points of the unfolding* $\breve{\tilde{h}} : (u, r, y) \mapsto \big(u, \breve{h}_u(r, y)\big)$ *contain* $\{\breve{u}_0\} \times \mathbb{S}^{2d-1}$ *and the* $\breve{\tilde{h}}$*-invariant "graph"* $\breve{W} = \{u = \breve{\psi}(r, y)\}$ *of a smooth local map* $\breve{\psi} = \breve{\psi}_{p/q} : \big(\mathbb{R}_+ \times \mathring{\mathbb{S}}^{2d-1}, \{0\} \times \mathring{\mathbb{S}}^{2d-1}\big) \to (\mathbb{C}^n, u_0)$.

(iii) *The non-fixed q-periodic points of* $\tilde{h}$ *contain the* $\tilde{h}$*-invariant "graph"* $W = \{u = \psi(z)\}$ *of a smooth* $\psi = \psi_{p/q} : \pi(\mathrm{Dom}\, \breve{\psi}) \setminus \{0\} \to \mathbb{C}^n \setminus \{u_0\}$.

(iv) *In the holomorphic case, W is holomorphic.*

Proof. (i) The relation $\pi \circ \breve{h}_u(r, y) = h_u \circ \pi(r, y)$ reads $\breve{h}_u(r, y) = (R_u, Y_u)(r, y)$ with $R_u(r, y) = |h_u(ry)|$ and $Y_u(r, y) = h_u(ry)/|h_u(ry)|$ for $r > 0$; now, by Taylor's formula, $h_u(ry) = rA_u(ry)y$, where $A_u(ry) := \int_0^1 Dh_u(try)\, dt$, hence $Y_u(r, y) = A_u(ry)y/|A_u(ry)y|$ wherever $A_u(ry)y \neq 0$, including $r = 0$ near $u = u_0$ since $A_u(0) = \mathrm{diag}\, u$.
(ii) One has that $h_u^q(ry) = ry$ if and only if $rG_u(r, y) = 0$, where $G_u(r, y) = G(u, r, y) = A_u\big(h_u^{q-1}(ry)\big) \cdots A_u(ry)y - y$, hence in particular $G(u, 0, y) = (\mathrm{diag}\, u)^q y - y$. Forgetting the fixed points $r = 0$, the equation $h_u^q(ry) = ry$ reads $G(u, r, y) = 0$. Now, all $y \in \mathbb{S}^{2d-1}$ verify $G(u_0, 0, y) = 0$ and $\partial_u G(u_0, 0, y) = q\, \mathrm{diag}(\bar{\rho}_1 y_1, \ldots, \bar{\rho}_d y_d)$, invertible if and only if $y_1 \cdots y_d \neq 0$, i.e. $y \in \mathring{\mathbb{S}}^{2d-1}$. Hence, there exist open neighbourhoods U of u_0 in $\mathbb{C}^d$ and $\breve{V}$ of $\{0\} \times \mathring{\mathbb{S}}^{2d-1}$ in $\mathbb{R}_+ \times \mathring{\mathbb{S}}^{2d-1}$ such that the zeros of $G|_{U \times \breve{V}}$ form the "graph" of a smooth implicit map $\breve{\psi} : \breve{V} \to U$; as before, this graph $\breve{W}$ becomes $\breve{\tilde{h}}$-invariant if it is replaced by $\breve{W} \cap \breve{\tilde{h}}^{-1}(\breve{W}) \cap \cdots \cap \breve{\tilde{h}}^{1-q}(\breve{W})$.
(iii) Recall that π is a diffeomorphism off the "exceptional divisor" $\pi^{-1}(0)$.
(iv) We can, therefore, "read" the equation $G_u(r, y) = 0$ via this diffeomorphism, that is, write it $g_u(z) := h_u^q(z) - z = 0$ for $z \neq 0$; as the unfolding $(u, r, y) \mapsto \big(u, G_u(r, y)\big)$ is a local diffeomorphism at every point of $\breve{W}$, so is $(u, r, y) \mapsto \big(u, rG_u(r, y)\big)$, hence the unfolding $\tilde{g} : (u, z) \mapsto \big(u, g_u(z)\big)$ is a diffeomorphism at every point of W; the map $\tilde{g}$ being holomorphic, its local inverses are, implying that W is holomorphic. □

Note. A nicer way to prove iv) is to use the complex blowup $\pi_{\mathbb{C}} : (D, z) \mapsto z$, $z \in D$, $D \subset \mathbb{C}^d$ complex line through 0;[9] the implicit function theorem yields a holomorphic $\breve{\psi}_{\mathbb{C}}$ "upstairs", defined on an open subset of the complement of the closure of $\{(D, z) : z \neq 0, z_1 \cdots z_d = 0\}$ and equal to u_0 on $\pi_{\mathbb{C}}^{-1}(0)$.

Example. If $h_u(z) = \mathrm{diag}\,\big(u + \chi(z_1^{q_1}, \ldots, z_d^{q_d})\big)z$, where q_j is the denominator of p_j/q in irreducible form and $\chi : (\mathbb{C}^d, 0) \to (\mathbb{C}^d, 0)$ is holomorphic, then $\psi(z) = \rho - \chi(z_1^{q_1}, \ldots, z_d^{q_d})$, which has contact of order at least $\min q_k$ with the constant ρ at 0.

Proposition 3.2. *The automorphism* $\mathrm{diag}\, \rho$ *lifts via π to the diffeomorphism* $\breve{\mathrm{diag}}\, \rho : (r, y) \mapsto \big(r, (\mathrm{diag}\, \rho)y\big)$.

[9] In the standard jth local chart of $\mathbb{CP}^{d-1}$, this blowup reads $\big(z_j, (w_k)_{k \neq j}\big) \mapsto z$ with $z_k = z_j w_k$ for $k \neq j$; the "forbidden" closed subset is the union of the hyperplanes $w_k = 0$.

(i) *Restricting* $\breve{W}$ *if required, there is a smooth diffeomorphism* $\breve{Z} = \breve{Z}_{p/q}$ *of* $\breve{W}$ *onto an open* $\breve{\operatorname{diag}}\,\rho$*-invariant subset* $\breve{\Omega} \supset \{0\} \times \mathring{\mathbb{S}}^{2d-1}$ *of* $\mathbb{R}_+ \times \mathring{\mathbb{S}}^{2d-1}$, *conjugating* $\breve{\tilde{h}}|_{\breve{W}}$ *to* $(\breve{\operatorname{diag}}\,\rho)|_{\breve{\Omega}}$, *with* $\breve{Z}(\breve{u}_0, y) = (0, y)$ *for all* $y \in \mathring{\mathbb{S}}^{2d-1}$.

(ii) *The map* $\breve{Z}$ *induces a smooth diffeomorphism* $Z = Z_{p/q}$ *of* W *onto the open* $\operatorname{diag}\rho$*-invariant "trefoil"* $\Omega := \pi(\breve{\Omega}) \setminus \{0\}$, *conjugating* $\tilde{h}|_W$ *to* $\operatorname{diag}\rho|_\Omega$ *and tending to* 0 *when the variable in* W *tends to* $\tilde{u}_0$.

(iii) *If* h *is holomorphic, so is* Z.

Proof. (i) The conjugacy $\breve{Z}_{p/q}$ is as in Question 1.6, but in polar coordinates:

$$\breve{Z}_{p/q}\big(\breve{\psi}(r,y),(r,y)\big) = \Big(r\,|C(r,y)|, \frac{C(r,y)}{|C(r,y)|}\Big),$$

where $$C(r,y) = \frac{1}{q}\sum_{k=0}^{q-1}(\operatorname{diag}\rho)^{-k} A_{\breve{\psi}(r,y)}\big(h^{k-1}_{\breve{\psi}(r,y)}(ry)\big)\cdots A_{\breve{\psi}(r,y)}(ry)y.$$

For all $y \in \mathring{\mathbb{S}}^{2d-1}$, one has that $C(0,y) = y$, hence $\breve{Z}(\breve{u}_0, y) = (0, y)$ and

$$D\breve{Z}_{p/q}(0,y) = \begin{pmatrix} 1 & 0 \\ * & \operatorname{id}_{y^\perp}\end{pmatrix} : \mathbb{R} \times y^\perp \to \mathbb{R} \times y^\perp$$

is invertible. It follows that $\breve{Z}_{p/q}$ is a smooth local diffeomorphism, whose domain can be made $\breve{\tilde{h}}$-invariant as usual. It is not difficult to check that it is a conjugacy, see equation (3.1) hereafter.
(ii) is obvious; by definition, the conjugacy $Z_{p/q}$ is as in Question 1.6:

$$Z_{p/q}\big(\psi_{p/q}(z), z\big) = \frac{1}{q}\sum_{k=0}^{q-1}(\operatorname{diag}\rho)^{-k} {h_{\psi_{p/q}(z)}}^k(z). \tag{3.1}$$

(iii) follows at once. □

Note. The diagonal action $e : (t, z) \mapsto e^{2\pi i \operatorname{diag} t} z$ of $\mathbb{T}^d$ on $\mathbb{C}^d$ preserves $\operatorname{diag}\rho$ and lifts to the action $\breve{e} : (t, r, y) \mapsto (r, e^{2\pi i \operatorname{diag} t} y) =: \breve{e}_t(r, y)$ of $\mathbb{T}^d$ on $\mathbb{R}_+ \times \mathbb{S}^{2d-1}$, which preserves $\breve{\operatorname{diag}}\,\rho$. The open subset $\breve{\Omega}$ becomes $\breve{e}$-invariant (and still $\breve{\operatorname{diag}}\,\rho$-invariant) if it is replaced by $\bigcap_{t\in\mathbb{T}^d} \breve{e}_t(\breve{\Omega})$, which contains $\{0\} \times \mathring{\mathbb{S}}^{2d-1}$ and is open because $\mathbb{T}^d$ is compact.

Hence, denoting again by $\breve{W}$ the inverse image of this new $\breve{\Omega}$ by $\breve{Z}$, the map $\breve{\tilde{h}}|_{\breve{W}}$ is invariant under the $\mathbb{T}^d$-action $\breve{Z}^*\breve{e} : \big(t, \breve{Z}^{-1}(r, y)\big) \mapsto \breve{Z}^{-1}\breve{e}(t, r, y)$; in particular, it preserves each orbit, which orbits constitute a foliation of $\breve{W}$ by d-tori $\breve{Z}^{-1}\big(\{r\} \times (x_1\mathbb{S}^1 \times \cdots \times x_d\mathbb{S}^1)\big)$ with $x_j > 0$ and $x_1^2 + \cdots + x_d^2 = 1$.

In general, these tori of course do not lie each in a slice $u =$ constant like the orbits of $\breve{\tilde{h}}|_{\breve{W}}$. The foliation, like the new $\breve{W}$, depends on the choice of $\breve{Z}$, which is far from unique since *the set of* $\breve{\operatorname{diag}}\,\rho$*-invariant smooth diffeomorphism germs* $\big(\mathbb{R}_+ \times \mathbb{S}^{2d-1}, \{0\} \times \mathbb{S}^{2d-1}\big) \to \big(\mathbb{R}_+ \times \mathbb{S}^{2d-1}, \{0\} \times \mathbb{S}^{2d-1}\big)$ *is infinite dimensional.*[10]

[10] Indeed, *the set of* $\operatorname{diag}\rho$*-invariant smooth diffeomorphism germs* $(\mathbb{C}^d, 0) \to (\mathbb{C}^d, 0)$ is *infinite dimensional*, as *any* smooth germ $\eta : (\mathbb{C}^d, 0) \to (\mathbb{C}^d, 0)$ yields the $\operatorname{diag}\rho$-invariant germ $\frac{1}{q}\sum_1^q(\operatorname{diag}\rho)^{-k} \circ \eta \circ (\operatorname{diag}\rho)^k$.

However, when p/q tends to some diophantine $\omega \in [0,1]^d$, the orbits of $\tilde{\breve{h}}|_{\breve{W}_{p/q}}$ should "become denser and denser in such invariant tori":

3.3. Passing to the limit in the holomorphic case?

In the holomorphic case, if $u_0 = (e^{2\pi i\omega_1}, \ldots, e^{2\pi i\omega_d})$, $\omega = (\omega_1, \ldots, \omega_d) \in \mathbb{T}^d$, the following result may apply to h_{u_0}:

Theorem 3.3. *Assume* ω diophantine *in the sense that, for some large* τ,

$$\inf_{1\le j\le d} \inf_{|k|\ge 2} |k|^\tau \left| e^{2\pi i k\omega} - e^{2\pi i\omega_j} \right| > 0,$$

where $k \in \mathbb{N}^d$, $|k| = k_1 + \cdots + k_d$ *and* $k\omega = k_1\omega_1 + \cdots + k_d\omega_d$. *Then, every holomorphic germ* $f : (\mathbb{C}^d, 0) \to (\mathbb{C}^d, 0)$ *such that* $Df(0) = \operatorname{diag} u_0$ *is holomorphically linearisable: there exists a holomorphic local diffeomorphism* $Z_\omega : (\mathbb{C}^d, 0) \to (\mathbb{C}^d, 0)$ *such that* $Z_\omega \circ h_{u_0} = (\operatorname{diag} u_0) Z_\omega$.[11] □

As the rotation $z \mapsto (\operatorname{diag} u_0)z = (e^{2\pi i\omega_j} z_j)_{1\le j\le d}$ preserves each d-torus $T_r = \{|z_1| = r_1, \ldots, |z_d| = r_d\}$, every embedded torus $T_{\omega r} = Z_\omega^{-1}(T_r)$ with $r_j > 0$ small enough is h_{u_0}-invariant and, of course, $Z_\omega|_{T_{\omega r}}$ conjugates $h_{u_0}|_{T_{\omega r}}$ to the rotation $z \mapsto (\operatorname{diag} u_0)z$ restricted to T_r.

Question 3.4. *Applied to* $f = h_{u_0}$, *is this the limit of what happens near* $u = (e^{2\pi i p_j/q})_{1\le j\le d}$ *when* $p/q \in \mathbb{Q}^d$ *tends to* ω?[12] *Do the maps* $\psi_{p/q}$ *tend to* $\psi_\omega = u_0$ *in some uniform neighbourhood of* 0? *For* $z \in \mathbb{C}^d$ *close to* 0, *does the periodic orbit* $\left\{ \left(\psi_{p/q}(z), {h_{\psi_{p/q(z)}}}^k(z) \right) : 0 \le k < q \right\}$ *of* $\tilde{h}$ *tend to the closed* $\tilde{h}$*-invariant torus* $\{u_0\} \times T_{\omega r}$ *such that* $z \in T_{\omega r}$? *More precisely, does the holomorphic linearisation* (3.1) *of* $\tilde{h}|_{W_{p/q}}$ *tend to* Z_ω *when* $n \to \infty$?

Note. This is not as simple as Question 1.6: indeed, unless I am mistaken, the maps $\psi_{p/q}$ are not *a priori* defined in a neighbourhood of 0, so that part of the question is whether $\operatorname{Dom} \psi_{p/q}$ tends to such a neighbourhood. On the other hand, it follows from normal form theory that, as in the case $d = 1$, the map $\psi = \psi_{p/q}$ has more and more contact with u_0 at 0 when $p/q \to \omega$.[13]

3.4. Passing to the limit in the smooth case?

If $u_0 = (e^{2\pi i\omega_1}, \ldots, e^{2\pi i\omega_d})$, where $\omega = (\omega_1, \ldots, \omega_d) \in \mathbb{T}^d$ is *non-resonant*, meaning that $\omega_1, \ldots, \omega_d \in \mathbb{T}$ are independent over $\mathbb{Z}$, then, by normal form theory, one has the following: for each positive integer N, up to smooth local conjugacy $(u, z) \mapsto \big(u, Z_u(z)\big)$, every h_u with $u - u_0$ small enough is tangent to order $2N + 1$ at 0 to a polynomial map

$$P_u(z) = \operatorname{diag}\left(u + \sum_{\ell=1}^{N} b_\ell(u)\big(|z_1|^2, \ldots, |z_d|^2\big) \right) z$$

[11] One can assume $DZ_\omega(0) = \mathrm{Id}$. Pöschel [31] attributes Theorem 3.3 to Siegel, who certainly proved its analogue for vector fields [34]. The same applies to its improvement by Brjuno. This "Siegel-Brjuno" theorem for maps and much more is proved in [30–32,38].

[12] For example, $p_j/q = p_{jn}/q_n$ can be the nth convergent of ω_j.

[13] If one prefers, $\breve{\psi}_{\mathbb{C}}$ has more and more contact with u_0 at points of $\pi_{\mathbb{C}}^{-1}(0)$.

with $b_\ell(u) : \mathbb{R}^d \to \mathbb{R}^d$ homogeneous of degree ℓ, depending smoothly on u. As for $d = 1$, it follows that when p/q tends to ω the map $\psi_{p/q}$ of Proposition 3.1 is tangent to higher and higher order at 0 to a polynomial normal form[14]

$$\hat{\psi}_{p/q}(z) = \chi_{p/q}\big(|z_1|^2, \dots, |z_d|^2\big), \qquad \chi_{p/q}(0) = \rho, \quad D\chi_{p/q}(0) = b_1(\rho).$$

Thus, if $b_1(u_0)$ (and, therefore, $b_1(\rho)$ for small $\frac{p}{q} - \omega$) is invertible then, restricting $\psi_{p/q}$, the set $\operatorname{Im}\psi_{p/q}$ lies near ρ in a thinner and thinner "Arnold tongue" along the smooth d-fold with corner $\chi_{p/q}([0,\varepsilon)^d)$ for small $\varepsilon > 0$.

Question 3.5. *Assume ω* diophantine *in the sense that, for some large τ,*

$$\inf_{m \neq 0} |m|^\tau \left|e^{2\pi i m\omega} - 1\right| > 0,$$

where $m \in \mathbb{Z}^d$, $|m| = m_1 + \cdots + m_d$ and $m\omega = m_1\omega_1 + \cdots + m_d\omega_d$. If $b_1(u_0)$ is invertible, one can wonder as in the holomorphic and one-dimensional cases whether one has the following when p/q tends to ω:

- *The $\breve{\psi}_{p/q}$'s tend to a map $\breve{\psi}_\omega$ of $\big(\mathbb{R}_+ \times \mathring{\mathbb{S}}^{2d-1}, \{0\} \times \mathring{\mathbb{S}}^{2d-1}\big)$ into $(\mathbb{C}^n, u_0)$ in a uniform neighbourhood of $\{0\} \times \mathring{\mathbb{S}}^{2d-1}$; thus, the $\tilde{\breve{h}}$-invariant surfaces $\breve{W}_{p/q}$ tend to the $\tilde{\breve{h}}$-invariant $2d$-fold $\breve{W}_\omega = \{u = \breve{\psi}_\omega(r, y)\}$.*
- *For each (r, y), the periodic orbits $\Big\{\Big(\breve{\psi}_{p/q}(r,y), \breve{h}^k_{\breve{\psi}_{p/q}(r,y)}(r,y)\Big)\Big\}_{0 \leq k < q}$ of $\tilde{\breve{h}}$ tend to a $\tilde{\breve{h}}$-invariant embedded d-torus $\{\breve{\psi}_\omega(r,y)\} \times T_{\omega r y}$ such that $(r, y) \in T_{\omega r y}$.*
- *The "linearisation" $\breve{Z}_{p/q}$ of Proposition 3.2 (i) tends to a local transformation $\breve{Z}_\omega$ "linearising" the local diffeomorphism $(r, y) \mapsto \breve{h}_{\breve{\psi}_\omega(r,y)}(r, y)$. Hence, the $\mathbb{T}^d$-action $\big(\theta, \breve{Z}_\omega^{-1}(r, y)\big) \mapsto \breve{Z}_\omega^{-1}\big(r, e^{2\pi i \operatorname{diag}\theta} y\big)$ leaves $\breve{\psi}_\omega$ invariant, implying that $\operatorname{Im}\breve{\psi}_\omega$ is a d-fold (with corner), limit of the narrower and narrower subsets $\operatorname{Im}\breve{\psi}_{p/q}$.*

Example. (KAM invariant tori) Assume that h possesses the following properties near some $u_0 = e^{2\pi i \omega_0}$ with ω_0 non-resonant:

(i) If $|u_1| = \cdots = |u_d| = 1$, the transformation h_u preserves the standard symplectic form $\sigma = \frac{1}{2i}(d\bar{z}_1 \wedge dz_1 + \cdots + d\bar{z}_d \wedge dz_d)$.

(ii) One has that $h_u = h_{(u_1/|u_1|, \dots, u_d/|u_d|)} \circ \operatorname{diag}(|u_1|, \dots, |u_d|)$, and therefore, $h_u^*\sigma = \frac{1}{2i}(|u_1|^2 d\bar{z}_1 \wedge dz_1 + \cdots + |u_d|^2 d\bar{z}_d \wedge dz_d)$.

(iii) The linear map $b_1(u_0)$ is an isomorphism.

Then, if the answer to question 3.5 is positive, it follows from (ii) that every ψ_ω takes its values in $\{|u_1| = \cdots = |u_d| = 1\}$, yielding a $3d$-dimensional analogue of Figure 2, see [23,28].[15]

[14] If one prefers, $\breve{\psi}$ has more and more contact with $\hat{\psi}_{p/q} \circ \pi$ at points of $\pi^{-1}(0)$.

[15] Note that the slices with $\omega \in \mathbb{Q}^d$ close to ω_0 contain invariant tori "far" away from 0.

4. Comments and references

My interest in this part of the program sketched in [13] awoke when I heard Abed Bounemoura talk about [4].

The dimension of both parameter and phase space, minimal here, can be much higher[16]. Proposition 1.3 has been known (at least) to me for thirty years, as well as the "blown-up" version of Proposition 2.1.[17] I have no reference for the higher dimensional results in Sect. 3.4. The excision of the coordinate hyperplanes in Propositions 3.1–3.2 corresponds to the closure of manifolds of periodic orbits of lower period, which might tend to (manifolds of) lower dimensional KAM tori à la Eliasson [22,23].

It is known that "good" periodic orbits accumulate on KAM tori. My naive hope is to do it the other way round and get the mysterious objects as limits of obvious ones, which would clarify a very intricate situation.

One of the sources of this article is an awfully biased reading of the two papers [19,20] by Alain Chenciner, to whom my debt cannot be overestimated, though he certainly does not share my viewpoint that conservative systems are essentially meant to deny the existence of death (and birth. . .).

Last but not least, I thank Jacques Féjoz and Laurent Stolovitch for very useful discussions and comments.

References

[1] Arnold, V.I.: Sur une propriété topologique des applications globalement canoniques de la mécanique classique. C. R. Acad. Sc. Paris **261**, Groupe 1, 3719–3722 (1965)

[2] Arnold, V.I.: Méthodes mathématiques de la mécanique classique. Mir, Moscou (1976), translated from the Russian original edition (1974) English version: Mathematical Methods of Classical Mechanics. Graduate Texts in Mathematics, vol. 60. Springer, New York (1989)

[3] Arnold, V.I., Avez, A.: Problèmes ergodiques de la mécanique classique. Gauthier-Villars, Paris (1967)

[4] Bounemoura, A., Fischler, S.: A Diophantine duality applied to the KAM and Nekhoroshev theorems. Math. Z. **275**(3), 1135–1167 (2013)

[5] Brjuno, A.D.: Analytical form of differential equations. Trans. Moscow Math. Soc. **25**, 131–288 (1971)

[6] Brjuno, A.D.: Analytical form of differential equations. Trans. Moscow Math. Soc. **26**, 199–239 (1972)

[7] Chaperon, M.: Quelques questions de géométrie symplectique [d'après, entre autres, Poincaré, Arnold, Conley et Zehnder]. Séminaire Bourbaki. Astérisque, 105–106, 231–249 (1983)

[16] In [13] one considers the tautological families $(f, x) \mapsto f(x)$, the parameter space being that of all maps f.

[17] As far as I know, the general version is due to Lino Samaniego [17].

[8] Chaperon, M.: Une idée du type "géodésiques brisées" pour les systèmes hamiltoniens. C. R. Acad. Sc. Paris **298**, série A, 293–296 (1984)

[9] Chaperon, M.: An elementary proof of the Conley-Zehnder theorem in symplectic geometry. In: Braaksma, B.L.J., Broer, H., Takens, F. (eds.) Dynamical Systems and Bifurcations, 1–8. Lecture Notes in Mathematics, vol. 1125. Springer, Berlin (1985)

[10] Chaperon, M.: Géométrie différentielle et singularités de systèmes dynamiques. Astérisque, 138–139, Société Mathématique de France, Paris (1986)

[11] Chaperon, M.: C^k-conjugacy of holomorphic flows near a singularity. Publ. Math. I.H.E.S. **64**, 143–183 (1986)

[12] Chaperon, M.: On generating families. In: Hofer, H., Taubes, C. H., Weinstein, A., Zehnder, E. (eds) The Floer Memorial Volume, 283–296. Progress in Mathematics, vol. 133. Birkhäuser, Basel (1995)

[13] Chaperon, M.: Singularities in dynamics: a catastrophic viewpoint. Proc. Steklov Inst. **267**, 8–39 (2009)

[14] Chaperon, M.: Generalized Hopf bifurcations: a birth lemma. Moscow Math. J. **11**(3), 413–438 (2011)

[15] Chaperon, M.: A forgotten theorem on $\mathbb{Z}^k \times \mathbb{R}^m$-action germs and related questions. Regul. Chaotic Dyn. **18**(6), 742–775 (2013)

[16] Chaperon, M., López de Medrano, S.: Regularities and singularities appearing in the study of polynomials and linear operators. Astérisque **323**, 123–160 (2009)

[17] Chaperon, M., Samaniego Mendoza, J.L.: On sub-harmonic bifurcations. C. R. Acad. Sci. Paris Sér. I **340**, 827–832 (2005)

[18] Chaperon, M., Zehnder, E.: Quelques résultats globaux en géométrie symplectique. In: Dazord, P., Desolneux-Moulis, N. (eds.) Géométrie symplectique et de contact: autour du théorème de Poincaré-Birkhoff, 71–121. Travaux en cours, Hermann, Paris (1984)

[19] Chenciner, A.: Bifurcations de points fixes elliptiques. I. Courbes invariantes. Publ. Math. I.H.E.S. **61**, 67–127 (1985)

[20] Chenciner, A.: Bifurcations de points fixes elliptiques. II. Orbites périodiques et ensembles de Cantor invariants. Invent. Math. **80**, 81–106 (1985)

[21] Conley, C.C., Zehnder, E.: The Birkhoff–Lewis fixed point theorem and a conjecture of V.I. Arnol'd. Invent. Math. **73**, 33–49 (1983)

[22] Eliasson, L.H.: Perturbations of stable invariant tori for Hamiltonian systems. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **15**(1), 115–147 (1988)

[23] Féjoz, J.: Introduction to KAM theory with a view to celestial mechanics. In: Bergounioux, M., Peyré, G., Schnörr, C., Caillau, J.-B., Haberkorn, T. (eds.) Variational Methods, pp. 387–433. W. de Gruyter, Berlin (2016)

[24] Floer, A.: Morse theory for Lagrangian intersections. J. Differ. Geom. **28**, 513–547 (1988)

[25] Gromov, M.: Pseudo holomorphic curves in symplectic manifolds. Invent. Math. **82**, 307–347 (1985)

[26] Hofer, H.: Lagrangian embeddings and critical point theory. Ann. Inst. Henri Poincaré, Analyse non linéaire **2**, 407–462 (1985)

[27] Laudenbach, F., Sikorav, J.-C.: Persistance d'intersection avec la section nulle au cours d'une isotopie hamiltonienne dans un fibré cotangent. Invent. Math. **82**, 349–357 (1985)

[28] Massetti, J.E.: A normal form à la Moser for diffeomorphisms and a generalization of Rüssmann's translated curve theorem to higher dimensions. Anal. PDE **11**(1), 149–170 (2018)

[29] Pérez-Marco, R.: Fixed points and circle maps. Acta Math. **179**, 243–294 (1997)

[30] Pöschel, J.: On invariant manifolds of complex analytic mappings near fixed points. Expo. Math. **4**, 97–109 (1986)

[31] Pöschel, J.: On the Siegel–Sternberg linearization theorem. arXiv:1702.03691v2 (2017)

[32] Rüssmann, H.: On the convergence of power series transformations of analytic mappings near a fixed point into a normal form. I.H.E.S, Preprint (1977)

[33] Siegel, C.L.: Iteration of analytic functions. Ann. Math. **43**, 607–612 (1942)

[34] Siegel, C.L.: Über die Normalform analytischer Differentialgleichungen in der Nähe einer Gleichgewichtslösung. (German) Nachr. Akad. Wiss. Göttingen. Math.-Phys. Kl. Math.-Phys.-Chem. Abt. **1952**, 21–30 (1952)

[35] Sikorav, J.-C.: Points fixes d'une application symplectique homologue l'identité. J. Differ. Geom. **22**, 49–79 (1985)

[36] Sikorav, J.-C.: Sur les immersions lagrangiennes dans un fibré cotangent. C.R. Acad. Sci. Paris t.302, série I, **3**, 119–122 (1986)

[37] Sikorav, J.-C.: Problèmes d'intersections et de points fixes en géométrie hamiltonienne. Comment. Math. Helv. **62**, 62–73 (1987)

[38] Stolovitch, L.: Family of intersecting totally real manifolds of $(\mathbb{C}^n, 0)$ and germs of holomorphic diffeomorphisms. Bull. Soc. Math. Fr. **143**(2), 247–263 (2015)

[39] Viterbo, C.: A proof of Weinstein's conjecture in $\mathbb{R}^{2n}$. Ann. Inst. Henri Poincaré, Analyse non linéaire **4**, 407–462 (1987)

[40] Viterbo, C.: Symplectic topology as the geometry of generating functions. Math. Ann. **292**, 685–710 (1992)

[41] Yoccoz, J.-C.: Petits diviseurs en dimension 1. Astérisque, 231, Société mathématique de France (1995)

Marc Chaperon
Université Paris Cité & Sorbonne Université, CNRS, IMJ-PRG
75006 Paris
France
e-mail: marc.chaperon@imj-prg.fr

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:53
https://doi.org/10.1007/s11784-022-00937-w
Published online June 28, 2022

Journal of Fixed Point Theory and Applications

Check for updates

Another look at the Hofer–Zehnder conjecture

Erman Çineli, Viktor L. Ginzburg and Başak Z. Gürel

Dedicated to Claude Viterbo on the occasion of his 60th birthday.

Abstract. We give a different and simpler proof of a slightly modified (and weaker) variant of a recent theorem of Shelukhin extending Franks' "two-or-infinitely-many" theorem to Hamiltonian diffeomorphisms in higher dimensions and establishing a sufficiently general case of the Hofer–Zehnder conjecture. A few ingredients of our proof are common with Shelukhin's original argument, the key of which is Seidel's equivariant pair-of-pants product, but the new proof highlights a different aspect of the periodic orbit dynamics of Hamiltonian diffeomorphisms.

Mathematics Subject Classification. 53D40, 37J12, 37J39.

Keywords. Periodic orbits, Hamiltonian diffeomorphisms, Frank's theorem, equivariant Floer cohomology, pseudo-rotations.

Contents

The work is partially supported by NSF CAREER award DMS-1454342 (BG) and by Simons Foundation Collaboration Grant 581382 (VG).

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

1. Introduction and main results

1.1. Introduction

In this paper, we give a different and simpler proof of a slightly modified and weaker version of a recent theorem of Shelukhin [40], extending Franks' "two-or-infinitely-many" theorem [14,15], to higher dimensions.

This celebrated theorem of Franks asserts that every area preserving diffeomorphism of S^2 has either exactly two or infinitely many periodic points. (Moreover, in the setting of Franks' theorem, there are also strong growth rate results; see, e.g., [16,30,32].) A generalization of Franks' theorem conjectured in [29, p. 263] is that a Hamiltonian diffeomorphism φ of a closed symplectic manifold has infinitely many periodic points whenever φ has "more than absolutely necessary" fixed points. (Hence, the title of [40] and of this paper.) The vaguely stated lower bound "more than absolutely necessary" is usually interpreted as a lower bound arising from some version of the Arnold conjecture, e.g., as the sum of the Betti numbers. For $\mathbb{CP}^n$, the expected threshold is $n + 1$ regardless of the non-degeneracy assumption and, in particular, it is 2 for $S^2 = \mathbb{CP}^1$ as in Franks' theorem. A slightly different interpretation of the conjecture, not directly involving the count of fixed points, is that the presence of a fixed or periodic point that is unnecessary from a homological or geometrical perspective is already sufficient to force the existence of infinitely many periodic points. We refer the reader to [20,27,28] for some results in this direction.

We note that whenever φ has finitely many periodic points, by passing to an iterate, one can assume them to be fixed points. Furthermore, φ has infinitely many periodic points if and only if it has infinitely many simple, i.e., uniterated, periodic orbits and the results are often stated in these terms. It is also worth keeping in mind that all known Hamiltonian diffeomorphisms φ with finitely many periodic orbits are strongly non-degenerate, i.e., φ^k is non-degenerate for all $k \in \mathbb{N}$.

Volume preserving diffeomorphisms or flows with finitely many simple periodic orbits play an important role in dynamics; see, e.g., [13] and references therein. In the Hamiltonian setting, they are sometimes referred to as pseudo-rotations. Recently, symplectic topological methods have been employed to study the dynamics of pseudo-rotations and its connections with symplectic topological properties of the underlying manifold in all dimensions; see [1,2,4,5,9,10,23,33,41,42].

The original proof of Franks' theorem utilized methods from low-dimensional dynamics, and the first purely symplectic topological proof was given in [11]. However, that proof and also a different approach from [6] were still strictly low-dimensional, and Shelukhin's theorem [40, Thm. A], is the first sufficiently general higher dimensional variant of Franks' theorem. (Strictly speaking [40, Thm. A] and our Theorem 1.1 and Corollary 1.2, which are overall slightly weaker, still fall short of completely reproving Franks' theorem in dimension two; we will discuss and compare these results in Sect. 1.2.) Similarly to [40], the key ingredient of our proof is Seidel's $\mathbb{Z}_2$-equivariant pair-of-pants product, [39]. (While we use the original version of the product, [40] relies on its $\mathbb{Z}_p$-equivariant version from [43].) Our proof also uses several simple ingredients from persistent homology theory in the form developed in [45] (see also [36]), although to a much lesser degree than [40].

Finally, it is worth pointing out that Hamiltonian pseudo-rotations are extremely rare and most of the manifolds do not admit such maps. This statement is known as the Conley conjecture. The state-of-the-art result is that the Conley conjecture holds for a manifold (M, ω) unless there exists $A \in \pi_2(M)$, such that $\langle c_1(TM), A\rangle > 0$ and $\langle \omega, A\rangle > 0$; see [7,22]. For example, the Conley conjecture holds when $c_1(TM)|_{\pi_2(M)} = 0$ or when M is negative monotone. For many manifolds, the conjecture is also known to hold C^∞-generically (see [18]); we refer the reader to [21] for a detailed survey and further references.

1.2. Shelukhin's theorem

Let φ be a Hamiltonian diffeomorphism of a closed monotone symplectic manifold M. We view φ as the time-one map in a time-dependent Hamiltonian flow and denote by $\mathcal{P}_k(\varphi)$ the set of its k-periodic points, arising from contractible k-periodic orbits. The Hamiltonian diffeomorphism φ is said to be *k-perfect* if $\mathcal{P}_k(\varphi) = \mathcal{P}_1(\varphi)$ and *perfect* if φ is k-perfect for all $k \in \mathbb{N}$. (We refer the reader to Sect. 2 for further notation and definitions used here.) We call φ a non-degenerate *pseudo-rotation* over a field $\mathbb{F}$ if it is non-degenerate, perfect, and the differential in the Floer complex of φ over $\mathbb{F}$ vanishes. This condition is independent of the choice of an almost complex structure and, by Arnold's conjecture, equivalent to that the number of 1-periodic orbits $|\mathcal{P}_1(\varphi)|$ is equal to the sum of Betti numbers of M over $\mathbb{F}$. Denote by $\beta(\varphi)$ the *boundary depth* of φ over $\mathbb{F}$, i.e., the length of the maximal finite bar in the *barcode* of φ; see [44,45] and also Sect. 4.

One of the goals of this paper is to give a simple proof to the following theorem proved in a slightly different form in [40].

Theorem 1.1. (Shelukhin's Theorem [40]) *Assume that φ is strongly non-degenerate and perfect and that $\beta(\psi)$ over $\mathbb{F}_2 := \mathbb{Z}_2$ is bounded from above for all Hamiltonian diffeomorphisms ψ of M or at least for all iterates $\psi = \varphi^{2^k}$ (e.g., $M = \mathbb{CP}^n$). Then, φ is a pseudo-rotation.*

Applying this to the iterates φ^{2^k}, we obtain

Corollary 1.2. [40] *Assume that φ is strongly non-degenerate, $\beta\big(\varphi^{2^k}\big)$ over $\mathbb{F}_2$ is bounded from above (e.g., $M = \mathbb{CP}^n$), and $|\mathcal{P}_1(\varphi)|$ is strictly greater than the sum of Betti numbers of M over $\mathbb{F}_2$. Then, $\big|\mathcal{P}_{2^k}(\varphi)\big| \to \infty$ as $k \to \infty$.*

This theorem is proved in Sect. 3.2 as an easy consequence of Theorem 3.1, a new result in this paper. (However, at least on the conceptual level, our proof of that theorem is also a subset of Shelukhin's argument, although the inclusion is rather implicit.)

In the rest of this section, we discuss the conditions of Theorem 1.1 and also some of the differences between Corollary 1.2 and the original Shelukhin's theorem [40, Thm. A], which is in several ways more general and more precise.

First of all, in the most recent version of [40, Thm. A], there are no restrictions on the ground coefficient field $\mathbb{F}$, while here $\mathbb{F} = \mathbb{F}_2$. When $\mathbb{F}$ is $\mathbb{Q}$, the assertion is that $\mathcal{P}_p(\varphi)$ contains a simple periodic orbit for every large prime p. As a consequence, one obtains the growth of order at least $O(k/\log k)$ for the number of simple periodic orbits of period up to k. This difference stems from the fact that the main tool used in [40] is the $\mathbb{Z}_p$-equivariant pair-of-pants product introduced in [43], while we rely on a somewhat simpler $\mathbb{Z}_2$-equivariant pair-of-pants product defined in [39]. We touch upon the p-iterated analogues of Theorem 1.1 and Corollary 1.2 in Remark 5.5.

Secondly, [40, Thm. A] allows for some degeneracy of φ. Namely, in the setting of Corollary 1.2, the number of 1-periodic orbits $|\mathcal{P}_1(\varphi)|$ in the condition that $|\mathcal{P}_1(\varphi)|$ is strictly greater than the sum of Betti numbers is replaced by

$$\sum_{x \in \mathcal{P}_1(\varphi)} \dim_{\mathbb{F}} \mathrm{HF}(x; \mathbb{F}), \tag{1.1}$$

where $\mathrm{HF}(x; \mathbb{F})$ is the local Floer (co)homology of x with coefficients in a field $\mathbb{F}$ (see, e.g., [19]). Note that, as a consequence, Corollary 1.2 still holds without the non-degeneracy assumption, provided that the number of 1-periodic orbits with $\mathrm{HF}(x; \mathbb{F}) \neq 0$ is greater than the sum of Betti numbers. In the setting of this paper, one should take $\mathbb{F} = \mathbb{F}_2$ and we will further discuss the degenerate case of Theorem 1.1 and Corollary 1.2 in Sect. 5.2. Overall, the role of the condition that $\mathrm{HF}(x; \mathbb{F}) \neq 0$ is unclear to us beyond the case of S^2. Franks' theorem has an analogue for a certain class of symplectomorphisms of surfaces, and then, interestingly, this condition becomes essential; see [3,18].

However, from our perspective, the most important difference lies in the proofs, which highlight different aspects of the dynamics and Floer theory of φ. Our proof focuses on the behavior of the shortest bar $\beta_{\min}$ in the barcode of φ (rather than the longest finite bar, a.k.a. the boundary depth, $\beta \geq \beta_{\min}$ [45]) or, to be more precise, of the shortest Floer arrow under the iteration from φ to φ^2; see Sect. 3.1. In particular, we show in Theorem 3.1 that when φ is 2-perfect the shortest arrow persists under such an iteration, although it may migrate into the equivariant domain for φ^2, and the length of the arrow doubles. The shortest non-equivariant arrow for φ^2 is at least as long as the equivariant one. Hence, $\beta_{\min}\big(\varphi^2\big) \geq 2\beta_{\min}(\varphi)$, and Theorem 1.1 readily follows from Theorem 3.1 applied to a sequence of period doubling iterations; see Sect. 3.2. The key ingredient in the proof of Theorem 3.1 is the equivariant pair-of-pants product, introduced in [39], having a very strong non-vanishing property also proved therein (see Proposition 2.3).

Finally, a few words are due on the requirement in Theorem 1.1 and Corollary 1.2 that $\beta(\psi)$ is bounded from above. First of all, note that while it would be sufficient to only have an upper bound on $\beta_{\min}(\psi)$ where $\psi = \varphi^{2^k}$

or, as in [40, Thm. A], on $\beta(\psi)$ where $\psi = \varphi^p$, all relevant results proved to date are more robust and give an upper bound on $\beta(\psi)$ for all ψ. [This is the curse (and the blessing) of symplectic topological methods in dynamics: they are very robust and general, but not particularly discriminating; they often tell the same thing about all maps. There are, however, exceptions.]

The simplest manifold for which such an a priori bound is established is $\mathbb{CP}^n$ for any coefficient field (suppressed in the notation), and the result essentially goes back to [12]. The argument is roughly as follows. (We use here the notation and conventions from Sect. 2.1.) First, recall that

$$\beta(\psi) \leq \gamma(\psi). \tag{1.2}$$

Here, $\gamma(\psi)$ is the γ-norm of ψ defined, using cohomology, as

$$\gamma(\psi) = -\big(\operatorname{c}_{\mathbb{1}}(\psi) + \operatorname{c}_{\mathbb{1}}(\psi^{-1})\big),$$

where $\operatorname{c}_\alpha(\psi)$ is the spectral invariant associated with a quantum cohomology class $\alpha \in \operatorname{HQ}(M)$ and $\mathbb{1}$ is the unit in the ordinary cohomology $\operatorname{H}(M)$ of M. (We suppress the grading in the cohomology notation when it is irrelevant.) The upper bound (1.2) holds for any closed monotone symplectic manifold and its proof is similar to the proof in [44] of the upper bound for β by the Hofer norm, but with continuation maps replaced by the multiplications by the image of $\mathbb{1}$ in $\operatorname{HF}(\psi)$ and $\operatorname{HF}\left(\psi^{-1}\right)$. (We refer the reader to [31] for some further results along these lines.) Applying the Poincaré duality in Floer cohomology (see [12]), it is not hard to show that $\operatorname{c}_{\mathbb{1}}(\psi^{-1}) = -\operatorname{c}_{\varpi}(\psi)$ when $N \geq n+1$, where ϖ is the generator of $\operatorname{H}^{2n}(M)$ and N is the minimal Chern number of M^{2n}. In particular, this is true for $M = \mathbb{CP}^n$, since then $N = n+1$. By construction, for any two classes α and ζ in $\operatorname{HQ}(M)$, the spectral invariants satisfy the Lusternik–Schnirelmann inequality $\operatorname{c}_{\alpha*\zeta}(\psi) \geq \operatorname{c}_\alpha(\psi)$. Thus, from the identity $\varpi * \zeta = \mathrm{q}\mathbb{1}$ where ζ is the generator of $\operatorname{HQ}^2(\mathbb{CP}^n)$, we conclude that $\operatorname{c}_{\mathbb{1}}(\psi) \leq \operatorname{c}_\varpi(\psi) \leq \operatorname{c}_{\mathbb{1}}(\psi) + \pi$. These inequalities, combined with (1.2), show that

$$\beta(\psi) \leq \gamma(\psi) \leq \pi$$

for any Hamiltonian diffeomorphism ψ of $\mathbb{CP}^n$.

A similar upper bound on β holds for all closed monotone manifolds M, such that $\operatorname{HQ}^{\text{even}}(M;\mathbb{F})$ for some field $\mathbb{F}$ is semi-simple, i.e., splits as an algebra into a direct sum of fields. This is [40, Thm. B] and, interestingly, this result bypasses the upper bound (1.2) in its original form. In fact, $\operatorname{HQ}(S^2 \times S^2;\mathbb{Q})$ is semi-simple, but γ is not bounded from above for $S^2 \times S^2$; see [40, Rmk. 7] and also [35, Thm. 6.2.6]. We are not aware of any algebraic criteria for an a priori bound on the γ-norm. Nor do we know how large the class of monotone symplectic manifolds with semi-simple $\operatorname{HQ}^{\text{even}}(M;\mathbb{F})$ is. In addition to $\mathbb{CP}^n$ (with any $\mathbb{F}$), the complex Grassmannians, $S^2 \times S^2$, and the one point blow-up of $\mathbb{CP}^2$ with standard monotone symplectic structures are in this class when $\operatorname{char}\mathbb{F} = 0$ (see [12] and references therein); but $S^2 \times S^2$ is not for $\mathbb{F} = \mathbb{F}_2$.

2. Preliminaries

2.1. Conventions and notation

For the reader's convenience, we set here our conventions and notation and briefly recall some basic definitions. The reader may want to consult this section only as needed.

Throughout this paper, the underlying symplectic manifold (M,ω) is assumed to be closed and strictly monotone, i.e., $[\omega]|_{\pi_2(M)} = \lambda c_1(TM)|_{\pi_2(M)} \neq 0$ for some $\lambda > 0$. The *minimal Chern number* of M is the positive generator N of the subgroup $\langle c_1(TM), \pi_2(M)\rangle \subset \mathbb{Z}$ and the *rationality constant* is the positive generator $\lambda_0 = 2N\lambda$ of the group $\langle \omega, \pi_2(M)\rangle \subset \mathbb{R}$.

A *Hamiltonian diffeomorphism* $\varphi = \varphi_H = \varphi_H^1$ is the time-one map of the time-dependent flow $\varphi^t = \varphi_H^t$ of a 1-periodic in time Hamiltonian $H\colon S^1 \times M \to \mathbb{R}$, where $S^1 = \mathbb{R}/\mathbb{Z}$. The Hamiltonian vector field X_H of H is defined by $i_{X_H}\omega = -dH$. In what follows, it will be convenient to view Hamiltonian diffeomorphisms together with the path φ_H^t, $t \in [0, 1]$, up to homotopy with fixed end points, i.e., as elements of the universal covering of the group of Hamiltonian diffeomorphisms.

Let $x\colon S^1 \to M$ be a contractible loop. A *capping* of x is an equivalence class of maps $A\colon D^2 \to M$, such that $A|_{S^1} = x$. Two cappings of x are equivalent if the integral of ω [or of $c_1(TM)$, since M is strictly monotone] over the sphere obtained by clutching the cappings is equal to zero. A capped closed curve $\bar{x}$ is, by definition, a closed curve x equipped with an equivalence class of cappings, and the presence of capping is indicated by a bar.

The action of a Hamiltonian H on a capped closed curve $\bar{x} = (x, A)$ is

$$\mathcal{A}(\bar{x}) = -\int_A \omega + \int_{S^1} H_t(x(t))\,\mathrm{d}t.$$

The space of capped closed curves is a covering space of the space of contractible loops, and the critical points of $\mathcal{A}_H$ on this space are exactly the capped 1-periodic orbits of X_H.

The k-periodic *points* of φ are in one-to-one correspondence with the k-periodic *orbits* of H, i.e., of the time-dependent flow φ^t. Recall also that a k-periodic orbit of H is called *simple* if it is not iterated. A k-periodic orbit x of H is said to be *non-degenerate* if the linearized return map $D\varphi^k\colon T_{x(0)}M \to T_{x(0)}M$ has no eigenvalues equal to one. A Hamiltonian H is non-degenerate if all its 1-periodic orbits are non-degenerate. We denote the collection of capped k-periodic orbits of H by $\bar{\mathcal{P}}_k(\varphi)$.

Let $\bar{x}$ be a non-degenerate capped periodic orbit. The *Conley–Zehnder index* $\mu(\bar{x}) \in \mathbb{Z}$ is defined, up to a sign, as in [37,38]. In this paper, we normalize μ, so that $\mu(\bar{x}) = n$ when x is a non-degenerate maximum (with trivial capping) of an autonomous Hamiltonian with small Hessian.

Fixing an almost complex structure, which will be suppressed in the notation, we denote by $(\mathrm{CF}(\varphi), d_{\mathrm{Fl}})$ and $\mathrm{HF}(\varphi)$ the Floer complex and cohomology of φ over $\mathbb{F}_2 = \mathbb{Z}_2$; see, e.g., [34,37]. (Throughout this paper, all complexes and cohomology groups are over $\mathbb{F}_2$.) The complex $\mathrm{CF}(\varphi)$ is generated by the capped 1-periodic orbits $\bar{x}$ of H, graded by the Conley–Zehnder index, and filtered by the action. The filtration level (or the action) of a chain $\xi \in \mathrm{CF}(\varphi)$ is defined by

$$\mathcal{A}(\xi) = \min\{\mathcal{A}(\bar{x}_i)\}, \quad \text{where } \xi = \sum \bar{x}_i. \tag{2.1}$$

(Note that the filtration depends on H, not just on φ, making of the notation $\mathrm{CF}(\varphi)$ somewhat misleading.) The differential d_{Fl} is the upward Floer differential: it increases the action and also the index by one. The Floer complex $\mathrm{CF}(\varphi)$ is also a finite-dimensional free module over the Novikov ring Λ. There are several choices of Λ; see, e.g., [34]. For our purposes, it is convenient to take the field of Laurent series $\mathbb{F}_2((\mathrm{q}))$ with $|\mathrm{q}| = 2N$ as Λ. With this choice, Λ naturally acts on $\mathrm{CF}(\varphi)$ by recapping, and multiplication by q corresponds to the recapping by $A \in \pi_2(M)$ with $\langle c_1(TM), A\rangle = N$. Furthermore, $\mathrm{CF}(\varphi)$ is a finite-dimensional vector space over Λ with a preferred basis formed by 1-periodic orbits with arbitrarily fixed capping.

Notationally, it is convenient to equip $\mathrm{CF}(\varphi)$ with a non-degenerate $\mathbb{F}_2$-valued pairing $\langle\, ,\rangle$ for which $\bar{\mathcal{P}}_1(\varphi)$ is an orthogonal basis: $\langle \bar{x}, \bar{y}\rangle = \delta_{\bar{x}\bar{y}}$. Then, essentially by definition

$$d_{\mathrm{Fl}}\bar{x} = \sum \langle d_{\mathrm{Fl}}\bar{x}, \bar{y}\rangle\, \bar{y}.$$

There is a canonical, grading-preserving isomorphism $\mathrm{HF}(\varphi) \xrightarrow{\cong} \mathrm{HQ}(M)[-n]$ where $\mathrm{HQ}(M)$ is the quantum cohomology of M; see, e.g., [34,37] and references therein. (Depending on the context, this is the PSS-isomorphism or the continuation map or a combination of the two.) The cohomology groups $\mathrm{HQ}(M)$ and $\mathrm{HF}(\varphi)$ are also modules over a Novikov ring Λ, and $\mathrm{HQ}(M) \cong \mathrm{H}(M) \otimes \Lambda \cong \mathrm{HF}(\varphi)$ (as a module).

The Floer complex carries a pairing

$$\mathrm{CF}(\varphi) \otimes \mathrm{CF}(\varphi) \to \mathrm{CF}\left(\varphi^2\right)[n]$$

descending, on the level of cohomology, to the so-called *pair-of-pants product*

$$\mathrm{HF}(\varphi) \otimes \mathrm{HF}(\varphi) \to \mathrm{HF}\left(\varphi^2\right)[n],$$

which we denote by $*$. Thus, with our conventions, $|\alpha * \beta| = |\alpha| + |\beta| + n$. In quantum cohomology, this product corresponds to the *quantum product*, also denoted by $*$, which makes it into a graded-commutative algebra over Λ with unit $\mathbb{1}$. This product is a deformation (in q) of the cup product: $\alpha * \beta = \alpha \cup \beta + O(\mathrm{q})$.

2.2. Equivariant Floer cohomology and the pair-of-pants product

2.2.1. Equivariant Floer cohomology: a brief introduction. The *equivariant Floer cohomology* $\mathrm{HF}_{\mathrm{eq}}\left(\varphi^2\right)$, introduced in [39], is the homology of a certain complex $\left(\mathrm{CF}_{\mathrm{eq}}\left(\varphi^2\right), d_{\mathrm{eq}}\right)$ called the *equivariant Floer complex*. As a graded $\mathbb{F}_2$-vector space or as a Λ-module

$$\mathrm{CF}_{\mathrm{eq}}\left(\varphi^2\right) = \mathrm{CF}\left(\varphi^2\right)[\mathrm{h}],$$

where $|\mathrm{h}| = 1$, and the differential d_{eq} has the form

$$d_{\mathrm{eq}} = d_{\mathrm{Fl}} + \mathrm{h}\, d_1 + \mathrm{h}^2\, d_2 + \cdots = d_{\mathrm{Fl}} + O(\mathrm{h}).$$

This differential is $\Lambda[h]$-linear and non-strictly action-increasing. It is roughly speaking defined as follows, mimicking Borel's construction of the $\mathbb{Z}_2$-equivariant Morse cohomology.

Fix a family $\tilde{J}$ of 2-periodic in t almost complex structures on M parametrized by the unit infinite-dimensional sphere $S^\infty \subset \mathbb{R}^\infty$. Here, $\mathbb{R}^\infty$ is the direct sum of infinitely many copies of $\mathbb{R}$, i.e., its elements $\xi = (\xi_0, \xi_1, \ldots)$ have only finitely many non-zero components, and $S^\infty = \{\|\xi\| = 1\}$ with $\|\xi\|^2 = \sum_k |\xi_k|^2$. The almost complex structure $\tilde{J}$ is required to satisfy the symmetry condition $\tilde{J}_{-\xi} = \tilde{J}'_\xi$, where $\tilde{J}'_\xi$ is obtained from $\tilde{J}_\xi$ by the time-shift $t \mapsto t+1$. Consider the self-indexing quadratic form $f(\xi) = \sum_k k|\xi_k|^2$ on S^∞ and an antipodally symmetric metric, such that the natural equatorial embedding $S^\infty \to S^\infty$ given by $(\xi_0, \xi_1, \ldots) \mapsto (0, \xi_0, \ldots)$ is an isometry. (Note also that the pull back of f by this embedding is $f+1$.) The almost complex structure $\tilde{J}$ must furthermore be constant in ξ near the critical points of f, invariant under the equatorial embedding, and satisfy a certain regularity requirement. Denote by $w_k^\pm$ the critical points of f of index k.

Next, consider the hybrid Morse–Floer complex of $\mathcal{A} + f$ with respect to $\tilde{J}$ and the metric on S^∞. This complex has pairs $(\bar{x}, w_k^\pm)$ with $\bar{x} \in \bar{\mathcal{P}}_2(\varphi)$ as generators and carries a natural $\mathbb{Z}_2$-action, free on the generators, sending $(\bar{x}, w_k^\pm)$ to $(\bar{x}', w_k^{-\pm})$, where $\bar{x}'$ is the time-shift of $\bar{x}$. It is easy to see that the homology of this hybrid complex is equal to $\mathrm{HF}(\varphi^2)$. By definition, $\mathrm{CF}_{\mathrm{eq}}\left(\varphi^2\right)$ is the $\mathbb{Z}_2$-invariant part of this hybrid complex, where we write $\bar{x}\,\mathrm{h}^k$ for $(\bar{x}, w_k^+) + (\bar{x}', w_k^-)$. The fact that the differential is h-linear follows from the requirement that f (up to a constant) and the auxiliary data are invariant under the equatorial embedding. Thus, in self-explanatory notation

$$d_k\bar{x} = \sum \left\langle d_k\bar{x},\, \mathrm{h}^k\bar{y}\right\rangle \bar{y}, \quad \text{where } \mu(\bar{y}) = \mu(\bar{x}) + 1 - k$$

and $\left\langle d_k\bar{x},\, \mathrm{h}^k\bar{y}\right\rangle$ counts mod 2 the total number of continuation Floer trajectories from $\bar{x}$ to $\bar{y}$ along gradient lines of f connecting w_0^+ to w_k^+ and from $\bar{x}$ to $\bar{y}'$ along gradient lines of f connecting w_0^+ to w_k^-. Clearly, the complex (and hence its cohomology) is filtered by the action $\mathcal{A}$ in addition to the filtration by $\mathcal{A}+f$. On the level of (co)chains, the filtration is defined similarly to (2.1), but with the powers of h ignored

$$\mathcal{A}(\xi) = \min\{\mathcal{A}(\bar{x}_i)\}, \quad \text{where } \xi = \sum \mathrm{h}^{m_i}\bar{x}_i.$$

The equivariant complex and the cohomology have natural continuation properties; see [39].

Example 2.1. Assume that φ is 2-perfect and φ^2 admits a regular 1-periodic almost complex structure J, i.e., for every pair $\bar{x}$ and $\bar{y}$ of 2-periodic orbits, the space of Floer trajectories connecting $\bar{x}$ to $\bar{y}$ has dimension $\mu(\bar{y}) - \mu(\bar{x})$. In particular, this space is empty when $\mu(\bar{y}) \leq \mu(\bar{x})$, except when $\bar{y} = \bar{x}$ and the space comprises one constant trajectory. Set $\tilde{J} = J$ to be a constant (i.e., independent of ξ) almost complex structure. Then, $\tilde{J}$ is also regular and $d_j = 0$ for $j \geq 1$, since continuation trajectories for a constant homotopy are just Floer trajectories. Thus, in this case, $\mathrm{HF}_{\mathrm{eq}}\left(\varphi^2\right) = \mathrm{HF}(\varphi)[\mathrm{h}]$ for any interval of action. These conditions are met, for instance, when $\varphi = \varphi_H$ is generated by a C^2-small autonomous Hamiltonian H. As a consequence, for any φ, the global cohomology $\mathrm{HF}_{\mathrm{eq}}\left(\varphi^2\right)$ is not a particularly interesting object: it is simply isomorphic to $\mathrm{HQ}(M)[\mathrm{h}]$ via the equivariant continuation (or the PSS map); see [46,47] for further details.

Remark 2.2. (Polynomials vs. formal power series) One difference between our definition of $\mathrm{CF}_{\mathrm{eq}}\left(\varphi^2\right)$ and the one in [39] is that there $\mathrm{CF}_{\mathrm{eq}}\left(\varphi^2\right) = \mathrm{CF}\left(\varphi^2\right)[[\mathrm{h}]]$; for in that setting, the expansion $d_{\mathrm{eq}}\bar{x} = \sum_k \mathrm{h}^k d_k \bar{x}$ may have infinitely many non-vanishing terms. However, as already pointed out in [39, Sect. 7], when M is strictly monotone, this expansion is necessarily finite. Indeed, otherwise, it would involve capped orbits $\bar{y} \in \bar{\mathcal{P}}_2(\varphi)$ with arbitrarily small index $\mu(\bar{y})$. However, due to monotonicity and since $\mathcal{P}_2(\varphi)$ is finite, such orbits would eventually have action strictly smaller than that of $\bar{x}$, which is impossible. This difference is essential for our proof as at some point in the argument we evaluate the elements of $\mathrm{CF}_{\mathrm{eq}}\left(\varphi^2\right)$ at $\mathrm{h} = 1$.

2.2.2. Equivariant pair-of-pants product. For our purposes, the most important feature of the equivariant Floer complex is that it is the target space of the *equivariant pair-of-pants product*, also defined in [39]. On the level of complexes this product is a chain map

$$\wp\colon\ \mathrm{CC}\left(\mathbb{Z}_2; \mathrm{CF}(\varphi) \otimes \mathrm{CF}(\varphi)\right) \to \mathrm{CF}_{\mathrm{eq}}\left(\varphi^2\right).$$

The domain of $\wp$ is the group cochain complex

$$\mathrm{CC}\left(\mathbb{Z}_2; \mathrm{CF}(\varphi) \otimes \mathrm{CF}(\varphi)\right) := \mathrm{CF}(\varphi) \otimes \mathrm{CF}(\varphi)[\mathrm{h}]$$

with the differential

$$d_{\mathbb{Z}_2} = d_{\mathrm{Fl}} + \mathrm{h}(id + \tau).$$

Here, τ is the involution $\tau(\bar{x} \otimes \bar{y}) = \bar{y} \otimes \bar{x}$ and the first term is induced by the Floer differential on $\mathrm{CF}(\varphi) \otimes \mathrm{CF}(\varphi)$. Note also that in these formulas and throughout the paper, all tensor products are over $\mathbb{F}_2$ unless specified otherwise. Furthermore, we distinguish between $\mathbb{F}_2$ and $\mathbb{Z}_2$: the former is a field and the latter is a group.

The equivariant pair-of-pants product is bilinear over $\Lambda[\mathrm{h}]$ and respects the action filtration. In particular, it can also be defined for a fixed action interval $[a, b]$ in the domain and $[2a, 2b]$ in the target, but here we will not need the filtered version of this construction. The map $\wp$ is a perturbation of the ordinary pair-of-pants product

$$\wp(\bar{x} \otimes \bar{y}) = \bar{x} * \bar{y} + O(\mathrm{h}), \tag{2.2}$$

and the $O(\mathrm{h})$ part is again polynomial in h involving only finitely many terms (depending on $\bar{x}$ and $\bar{y}$).

The cohomology of the domain of $\wp$ is the group cohomology $\mathrm{H}\left(\mathbb{Z}_2; \mathrm{CF}(\varphi) \otimes \mathrm{CF}(\varphi)\right)$ of $\mathbb{Z}_2$ with coefficients in $\mathrm{CF}(\varphi) \otimes \mathrm{CF}(\varphi)$. Thus, on the level of cohomology, the equivariant pair-of-pants product turns into a homomorphism

$$\mathrm{H}\left(\mathbb{Z}_2; \mathrm{CF}(\varphi) \otimes \mathrm{CF}(\varphi)\right) \cong \mathrm{H}\left(\mathbb{Z}_2; \mathrm{HF}(\varphi) \otimes \mathrm{HF}(\varphi)\right) \to \mathrm{HF}_{\mathrm{eq}}\left(\varphi^2\right). \tag{2.3}$$

(The first isomorphism is a consequence of the fact that $\mathrm{CF}(\varphi) \otimes \mathrm{CF}(\varphi)$ and $\mathrm{HF}(\varphi) \otimes \mathrm{HF}(\varphi)$ are equivariantly quasi-isomorphic.) The map (2.3) obviously kills the h-torsion in the domain; it is a deformation in h of the standard pair-of-pants product due to (2.2) and is closely related to a quantum deformation of the Steenrod squares; see [39, 46, 47] and also [10] for a short introduction. The map (2.3) is a monomorphism modulo h-torsion; [40]. For symplectically aspherical manifolds, but not in the strictly monotone case, (2.3) is also onto and hence an isomorphism modulo h-torsion i.e., the kernel and the cokernel are torsion modules; see [39].

On the level of complexes $\wp$ has the following extremely important feature:

Proposition 2.3. (Seidel's non-vanishing theorem [39, Prop. 6.7]) *For every $\bar{x} \in \bar{\mathcal{P}}_1(\varphi)$, we have*

$$\wp(\bar{x} \otimes \bar{x}) = \mathrm{h}^m \bar{x}^2 + \cdots , \tag{2.4}$$

where $\bar{x}^2 \in \bar{\mathcal{P}}_2(\varphi)$ is the second iterate of $\bar{x}$ and $m = 2\mu(\bar{x}) - \mu\big(\bar{x}^2\big) + n$ and the dots stand for a sum of capped orbits with action strictly greater than $2\mathcal{A}(\bar{x})$.

This non-vanishing property points to a stark difference between the equivariant and non-equivariant pair-of-pants products: $\bar{x} * \bar{x} = \bar{x}^2 + \cdots$ only when $\mu\big(\bar{x}^2\big) = 2\mu(\bar{x}) + n$, i.e., $m = 0$ in (2.4); cf. [9].

Remark 2.4. A generalization of the equivariant pair-of-pants product to the p-th iterates φ^p, where p is a prime, replacing $\mathbb{Z}_2$ by $\mathbb{Z}_p$ and $\mathbb{F}_2$ by $\mathbb{F}_p$ is constructed in [43]. This construction and the analogue of Seidel's non-vanishing theorem for the pth iterate play a crucial role in the original proof of Shelukhin's theorem in [40]; cf. Remark 5.5.

3. Floer graphs

3.1. Main result

The key to the statement of our main result is the following admittedly naive and obvious construction which has been used, at least on an informal level, for quite some time.

Let φ be a non-degenerate Hamiltonian diffeomorphism of a closed monotone symplectic manifold M. Consider the directed graph $\Gamma(\varphi)$ whose vertices are capped fixed points of φ, and two vertices $\bar{x}$ and $\bar{y}$ are connected by an arrow (from $\bar{x}$ to $\bar{y}$) if and only if $\mu(\bar{y}) = \mu(\bar{x}) + 1$ and there is an odd number of Floer trajectories from $\bar{x}$ to $\bar{y}$, i.e., $\langle d_{\mathrm{Fl}}\bar{x}, \bar{y}\rangle = 1$. The length of an arrow is the difference of actions of $\bar{y}$ and $\bar{x}$. We call $\Gamma(\varphi)$ the *Floer graph* of φ.

When M is strictly monotone as is always assumed in this paper, the group $\mathbb{Z}$ acts freely on $\Gamma(\varphi)$ by simultaneous recapping, preserving the arrow length. Sometimes, it is convenient to consider the *reduced Floer graph* $\tilde{\Gamma}(\varphi) := \Gamma(\varphi)/\mathbb{Z}$. The length of an arrow in $\tilde{\Gamma}(\varphi)$ is still well defined. Note that, unless M is symplectically aspherical, both $\Gamma(\varphi)$ and $\tilde{\Gamma}(\varphi)$ are infinite, but the latter has finitely many arrows. In particular, if $d_{\mathrm{Fl}} \neq 0$, there exists a shortest arrow. Such an arrow might not be unique, although it is unique for a generic φ, but obviously all shortest arrows have the same length.

The *equivariant Floer graph* $\Gamma_{\mathrm{eq}}(\varphi^2)$ of φ^2 is defined in a similar fashion. (We are assuming that φ^2 is non-degenerate, and hence, φ is also non-degenerate.) Its vertices are capped two-periodic orbits of φ. The vertices $\bar{x}$ and $\bar{y}$ are connected by an arrow if and only if $\bar{y}$ enters $d_{\mathrm{eq}}(\bar{x})$ with non-zero coefficient. In other words, now, we do not require the index difference to be 1, and $\bar{x}$ and $\bar{y}$ are connected by an arrow if and only if $\bar{x}$ and $\mathrm{h}^m\bar{y}$, where $m = \mu(\bar{x}) - \mu(\bar{y}) + 1$, are connected by an odd number of equivariant Floer trajectories. The length of an arrow is again the difference of actions. As in the non-equivariant case, the *reduced equivariant Floer graph* $\tilde{\Gamma}_{\mathrm{eq}}(\varphi^2) := \tilde{\Gamma}_{\mathrm{eq}}(\varphi^2)/\mathbb{Z}$ has only finitely many arrows, and hence, the shortest arrows exist.

We note that $\Gamma(\varphi^2)$ and $\Gamma_{\mathrm{eq}}(\varphi^2)$ (and their reduced counterparts) have the same vertices. Furthermore, since $d_{\mathrm{eq}} = d_{\mathrm{Fl}} + O(\mathrm{h})$, every arrow in $\Gamma(\varphi^2)$ is also an arrow in $\Gamma_{\mathrm{eq}}(\varphi^2)$, i.e., the equivariant Floer graph is obtained from its non-equivariant counterpart by adding arrows. Note that in the process, the shortest arrow length can only get shorter or remain the same. Also, observe that there is a natural one-to-one map from the vertices of $\tilde{\Gamma}(\varphi)$ to the vertices of $\tilde{\Gamma}(\varphi^2)$ sending $\bar{x}$ to $\bar{x}^2$; likewise for unreduced graphs. However, even when φ is 2-perfect, this map is not onto unless M is symplectically aspherical.

The main new result of the paper is the following theorem which relates the Floer graphs for φ and its second iterate φ^2.

Theorem 3.1. *Assume that φ is 2-perfect and φ^2 is non-degenerate. Then, $\bar{x}$ and $\bar{y}$ are connected by one of the shortest arrows in $\Gamma(\varphi)$ if and only if $\bar{x}^2$ and $\bar{y}^2$ are connected by one of the shortest arrows in $\Gamma_{\mathrm{eq}}(\varphi^2)$.*

This theorem is proved in Sect. 5.1 after we recall in Sect. 4 a few relevant facts about barcodes.

Remark 3.2. (The role of an almost complex structure) The Floer graph of φ depends on the choice of an almost complex structure J, and hence should rather be denoted by $\Gamma(\varphi, J)$. Likewise, the equivariant Floer graph depends on the parametrized almost complex structure. However, in both cases, the collection of shortest arrows is independent of this choice. This fact implicitly follows from Theorem 3.1 or can be proved directly by a continuation argument.

Note also that Floer graphs are stable under small perturbations of φ and J. To be more precise, $\Gamma(\varphi, J) = \Gamma(\tilde{\varphi}, \tilde{J})$ whenever $\tilde{\varphi}$ is sufficiently close to φ and $\tilde{J}$ is close to J. The same is true in the equivariant setting.

3.2. Implications and the proof of Theorem 1.1

Theorem 3.1 shows that when φ is perfect, the shortest arrow (or, to be more precise, every shortest arrow) persists from φ to φ^2, although in the process, it might move to the equivariant domain. This happens exactly when the difference of indices changes: $\mu(\bar{y}) - \mu(\bar{x}) = 1$ but $\mu(\bar{y}^2) - \mu(\bar{x}^2) \neq 1$. Moreover, in this case, we necessarily have $\mu(\bar{y}^2) - \mu(\bar{x}^2) < 1$. On the other hand, if the difference of indices remains equal to one, the orbits continue to be connected by one of the shortest non-equivariant arrows.

Denote by $\beta_{\min}(\varphi) = \mathcal{A}(\bar{y}) - \mathcal{A}(\bar{x})$ the length of a shortest arrow. As follows from Proposition 4.3, $\beta_{\min}(\varphi)$ is exactly equal to the shortest bar in the barcode of φ. Since every non-equivariant arrow for φ^2 is also an equivariant arrow, the shortest equivariant arrow length $\beta^{\mathrm{eq}}_{\min}(\varphi^2)$ for φ^2 does not exceed $\beta_{\min}(\varphi^2)$, that is

$$\beta^{\mathrm{eq}}_{\min}(\varphi^2) \leq \beta_{\min}(\varphi^2).$$

In the setting of Theorem 3.1

$$\beta^{\mathrm{eq}}_{\min}(\varphi^2) = \mathcal{A}(\bar{y}^2) - \mathcal{A}(\bar{x}^2) = 2\beta_{\min}(\varphi).$$

We conclude that

$$2\beta_{\min}\big(\varphi^{2^k}\big) \leq \beta_{\min}\big(\varphi^{2^{k+1}}\big)$$

as long as the iterates of φ remain perfect and non-degenerate, and hence

$$2^k\beta_{\min}(\varphi) \leq \beta_{\min}\big(\varphi^{2^k}\big).$$

In particular, when φ is perfect, the longest finite bar $\beta(\varphi)$ (and even the shortest bar) in the barcode cannot be bounded from above for the iterates of φ. This proves Theorem 1.1.

Remark 3.3. An interesting question that arises from Theorem 3.1 is if a shortest arrow could persist in the non-equivariant domain for all iterates φ^{2^k}, assuming that φ is perfect. As discussed above, this would be the case if and only if $\mu\big(\bar{y}^{2^k}\big) - \mu\big(\bar{x}^{2^k}\big) = 1$ for all $k \in \mathbb{N}$. Using a slightly simplified version of the index divisibility theorem from [24], one can show that this is impossible when φ is replaced by a suitable iterate φ^m. (This is non-obvious.) Passing to an iterate is apparently essential because there exist pairs of strongly non-degenerate elements A and B in $\widetilde{\mathrm{Sp}}(2n)$, such that $\mu\big(A^{2^k}\big) - \mu\big(B^{2^k}\big) = 1$ for all $k = 0, 1, 2, \ldots$.

4. A few words about the shortest bar

In this section we recall a few facts, well known to experts, about persistent homology in the context of Hamiltonian Floer theory. All results discussed here are contained in, e.g., [45], although in some instances implicitly and usually in a much more general setting. A reader sufficiently familiar with the material can easily skip this section. There are, however, two points the reader might want to keep in mind. Namely, our emphasis here is on the shortest bar rather than the longest finite bar (aka the boundary depth) which is more frequently used in applications to dynamics. Second, our sign conventions are different from those in [45] due to the fact that we are working with Floer cohomology.

Consider the Floer complex $\mathcal{C} := \mathrm{CF}(\varphi)$ of a non-degenerate Hamiltonian diffeomorphism φ of a strictly monotone symplectic manifold, equipped with the standard action filtration. Clearly, $\mathcal{C}$ is a finite-dimensional vector space over Λ and the collection of 1-periodic orbits of φ with fixed capping forms a basis of $\mathcal{C}$.

A finite set of vectors $\xi_i \in \mathcal{C}$ is said to be *orthogonal* if, for any collection of coefficients $\lambda_i \in \Lambda$, we have

$$\mathcal{A}\left(\sum \lambda_i \xi_i\right) = \min \mathcal{A}(\lambda_i \xi_i).$$

(Recall that with our conventions

$$\mathcal{A}(\xi) := \min \mathcal{A}(\bar{x}_i) \quad \text{when } \xi = \sum \bar{x}_i;$$

see (2.1).) It is not hard to show that an orthogonal set is necessarily linearly independent over Λ.

Example 4.1. Assume that all capped 1-periodic orbits of φ have distinct actions. Write $\xi_i = \bar{x}_i + \cdots$, where the dots stand for the orbits with action strictly greater than $\bar{x}_i$. Then, it is easy to see that the set ξ_i is orthogonal if and only if the capped orbits $\bar{x}_i$ are distinct.

Definition 4.2. A basis $\mathcal{B} = \{\alpha_i, \eta_j, \gamma_j\}$ of $\mathcal{C}$ over Λ is said to be a *singular decomposition* if

- $d_{\mathrm{Fl}}\alpha_i = 0$,
- $d_{\mathrm{Fl}}\eta_j = \gamma_j$,
- $\mathcal{B}$ is orthogonal.

It is shown in [45, Sections 2 and 3] that $\mathcal{C}$ admits a singular decomposition. For the sake of brevity, we omit the proof of this fact. In what follows we will order the pairs (η_j, γ_j), so that:

$$\mathcal{A}(\gamma_1) - \mathcal{A}(\eta_1) \leq \mathcal{A}(\gamma_2) - \mathcal{A}(\eta_2) \leq \cdots . \tag{4.1}$$

This increasing sequence is usually referred to as the *barcode* of φ (or to be more precise the collection of finite bars). The maximal entry in the sequence is called the *boundary depth* $\beta(\varphi)$, [44]. The barcode is independent of the choice of a singular decomposition (see, e.g., [45]), but here we do not use this fact. Instead, we need the following characterization of the shortest bar $\beta_{\min} = \beta_{\min}(\varphi)$:

Proposition 4.3. [45] *Set*

$$\beta_{\min} := \mathcal{A}(\gamma_1) - \mathcal{A}(\eta_1).$$

Then

$$\beta_{\min} = \inf \{\mathcal{A}(\bar{y}) - \mathcal{A}(\bar{x}) \mid \langle d_{\mathrm{Fl}}\bar{x}, \bar{y}\rangle = 1\} \tag{4.2}$$

$$= \inf \{\mathcal{A}(d_{\mathrm{Fl}}\xi) - \mathcal{A}(\xi) \mid \xi \in \mathcal{C}, \xi \neq 0\}. \tag{4.3}$$

Here, in the first equality, the infimum is taken over all capped 1-periodic orbits $\bar{x}$ and $\bar{y}$, such that $\bar{y}$ enters $d_{\mathrm{Fl}}\bar{x}$ with non-zero coefficient and, in the second, over all non-zero $\xi \in \mathcal{C}$. In particular, $\beta_{\min}(\varphi)$ is the shortest arrow length in $\Gamma(\varphi)$.

Note that the infimums in (4.2) and (4.3) are actually attained and thus can be replaced by minima, and that the proposition can be thought of as an analogue for $\mathcal{C}$ of the Courant–Fischer minimax theorem giving a variational interpretation of the eigenvalues of a quadratic form. For the sake of completeness, we include a proof of Proposition 4.3.

Proof. Let us denote the right-hand sides in (4.2) and (4.2) by $\beta'_{\min}$ and, respectively, $\beta''_{\min}$. We claim that $\beta'_{\min} = \beta''_{\min}$. Indeed, setting $\xi = \bar{x}$, in (4.3), it is easy to see that $\beta''_{\min} \leq \beta'_{\min}$. On the other hand, writing $\xi = \bar{x}_1 + \bar{x}_2 + \dots$ in the order of increasing action and $d_{\mathrm{Fl}}\xi = \sum d_{\mathrm{Fl}}\bar{x}_i = \bar{y} + \dots$, we observe that $\langle \bar{y}, d_{\mathrm{Fl}}\bar{x}_i\rangle = 1$ for some i. Then

$$\begin{aligned}\mathcal{A}(d_{\mathrm{Fl}}\xi) - \mathcal{A}(\xi) &= \mathcal{A}(\bar{y}) - \mathcal{A}(\bar{x}_1)\\ &\geq \mathcal{A}(\bar{y}) - \mathcal{A}(\bar{x}_i)\\ &\geq \beta'_{\min},\end{aligned}$$

and thus, $\beta''_{\min} \geq \beta'_{\min}$.

Next, clearly, $\beta_{\min} \geq \beta''_{\min}$. Therefore, it remains to show that $\beta_{\min} \leq \beta''_{\min}$. To this end, let us decompose ξ in the basis $\mathcal{B}$ over Λ

$$\xi = \sum \lambda_j \eta_j + \sum \lambda'_j \gamma_j + \sum \lambda''_i \alpha_i.$$

Then

$$d_{\mathrm{Fl}}\xi = \sum \lambda_j \gamma_j.$$

By orthogonality

$$\mathcal{A}(d_{\mathrm{Fl}}\xi) = \min \mathcal{A}(\lambda_j \gamma_j) = \mathcal{A}(\lambda_k \gamma_k)$$

for some k, and again by orthogonality

$$\mathcal{A}(\xi) \leq \min \mathcal{A}(\lambda_j \eta_j) \leq \mathcal{A}(\lambda_k \eta_k).$$

Therefore

$$\begin{aligned}\mathcal{A}(d_{\mathrm{Fl}}\xi) - \mathcal{A}(\xi) &\geq \mathcal{A}(\lambda_k \gamma_k) - \mathcal{A}(\lambda_k \eta_k)\\ &= \mathcal{A}(\gamma_k) - \mathcal{A}(\eta_k)\\ &\geq \mathcal{A}(\gamma_1) - \mathcal{A}(\eta_1) = \beta_{\min}.\end{aligned}$$

As a consequence, $\beta_{\min} \leq \beta''_{\min}$, which finishes the proof of the proposition. □

Remark 4.4. In conclusion, we point out that all results in this section are purely algebraic and extend in a straightforward way to any ungraded finite-dimensional complex over Λ with an "action filtration" having expected properties; see [45].

5. Proof of Theorem 3.1 and further remarks

5.1. Proof of Theorem 3.1

We begin by proving the theorem under the additional background assumption that *all actions and action differences for* φ and φ^2 *are distinct modulo* *the rationality constant* λ_0. Then, in the last step of the proof, we will show how to remove this extra assumption. Note that, in particular, this assumption guarantees that the shortest arrow is unique for $\Gamma(\varphi)$ and $\Gamma_{\text{eq}}(\varphi^2)$.

Remark 5.1. It is worth pointing out that while this background assumption is satisfied C^∞-generically, it is not quite innocuous in the context of pseudo-rotations or perfect Hamiltonian diffeomorphisms. Indeed, in this case, one can expect certain "resonance relations" between actions or actions and mean indices to hold; see [18,26].

The proof is carried out in three steps.

Step 1: The shortest arrow for φ. In this step we simply apply the machinery from Sect. 4 to $\mathrm{CF}(\varphi)$. Let $\mathcal{B} = \{\alpha_i, \eta_j, \gamma_j\}$ be a singular decomposition for $\mathrm{CF}(\varphi)$ over Λ; see Definition 4.2. Due to the background assumption, the inequalities in (4.1) are strict

$$\mathcal{A}(\gamma_1) - \mathcal{A}(\eta_1) < \mathcal{A}(\gamma_2) - \mathcal{A}(\eta_2) < \cdots . \tag{5.1}$$

Let us write

$$\gamma_1 = \bar{y}_* + \cdots \quad \text{and} \quad \eta_1 = \bar{x}_* + \cdots ,$$

where dots stand for higher action terms, and $\bar{x}_*$ and $\bar{y}_*$ are unique by the background assumption. Then, by definition

$$\mathcal{A}(\gamma_1) = \mathcal{A}(\bar{y}_*) \quad \text{and} \quad \mathcal{A}(\eta_1) = \mathcal{A}(\bar{x}_*),$$

and hence

$$\beta_{\min} := \mathcal{A}(\gamma_1) - \mathcal{A}(\eta_1) = \mathcal{A}(\bar{y}_*) - \mathcal{A}(\bar{x}_*).$$

We claim that

$$\langle d_{\text{Fl}}\bar{x}_*, \bar{y}_*\rangle = 1. \tag{5.2}$$

Indeed, $\langle d_{\text{Fl}}\bar{x}, \bar{y}_*\rangle = 1$ for some $\bar{x}$ entering η_1. Then

$$\beta_{\min} = \mathcal{A}(\bar{y}_*) - \mathcal{A}(\bar{x}_*) \geq \mathcal{A}(\bar{y}_*) - \mathcal{A}(\bar{x}) \geq \beta_{\min}.$$

It follows that the first inequality is in fact an equality and $\bar{x} = \bar{x}_*$ due to the background assumption.

Therefore, by Proposition 4.3 and (5.2), $\bar{x}_*$ and $\bar{y}_*$ are connected by the shortest arrow in $\Gamma(\varphi)$.

Step 2: The shortest arrow for φ^2. In the previous step, we have shown that $\bar{x}_*$ and $\bar{y}_*$ are connected by the shortest arrow in $\mathrm{CF}(\varphi)$. Our goal now is to prove the following key fact.

Lemma 5.2. *The iterated orbits $\bar{x}_*^2$ and $\bar{y}_*^2$ are connected by the shortest arrow in $\Gamma_{\text{eq}}(\varphi^2)$.*

Since under the background assumption, the shortest arrows in $\tilde{\Gamma}(\varphi)$ and $\Gamma_{\text{eq}}(\varphi^2)$ are unique, this will establish the theorem.

Proof of Lemma 5.2. In the notation from Sect. 2.2, set

$$\begin{aligned}
\hat{\alpha}_i &= \wp(\alpha_i \otimes \alpha_i),\\
\hat{\eta}_j &= \mathrm{h}\wp(\eta_j \otimes \eta_j) + \wp(\eta_j \otimes \gamma_j),\\
\hat{\gamma}_j &= \wp(\gamma_j \otimes \gamma_j).
\end{aligned}$$

Then, by Seidel's non-vanishing theorem (Proposition 2.3)

$$\hat{\eta}_1 = \mathrm{h}^m \bar{x}_*^2 + \cdots \quad \text{and} \quad \hat{\gamma}_1 = \mathrm{h}^{m'} \bar{y}_*^2 + \cdots$$

for some $m \geq 0$ and $m' \geq 0$, where the dots again stand for higher action terms.

Since $\wp$ is a chain map, i.e., $\wp \circ d_{\mathbb{Z}_2} = d_{\mathrm{eq}} \circ \wp$, we have

$$d_{\mathrm{eq}} \hat{\alpha}_i = 0$$

and

$$\begin{aligned} d_{\mathrm{eq}} \hat{\eta}_j &= \mathrm{h}\wp(\gamma_j \otimes \eta_j) + \mathrm{h}\wp(\eta_j \otimes \gamma_j) \\ &\quad + \wp(\mathrm{h}\eta_j \otimes \gamma_j + \mathrm{h}\gamma_j \otimes \eta_j) \\ &\quad + \wp(\gamma_j \otimes \gamma_j) \\ &= \hat{\gamma}_j. \end{aligned}$$

This indicates that the collection $\hat{\mathcal{B}} := \{\hat{\alpha}_i, \hat{\eta}_j, \hat{\gamma}_j\}$ can be thought of as a singular decomposition of $\mathrm{CF}_{\mathrm{eq}}\left(\varphi^2\right)$ with the minimal bar given by

$$\mathcal{A}(\hat{\gamma}_1) - \mathcal{A}(\hat{\eta}_1) = \mathcal{A}\big(\bar{y}_*^2\big) - \mathcal{A}\big(\bar{x}_*^2\big),$$

and, arguing similarly to Step 1, we should be able to show that $\bar{x}_*^2$ and $\bar{y}_*^2$ are connected by the shortest arrow. A minor technical difficulty that arises at this stage is that $\mathrm{CF}_{\mathrm{eq}}\left(\varphi^2\right)$ does not fit in with the algebraic framework of Sect. 4 or [45]. Namely, $\mathrm{CF}_{\mathrm{eq}}\left(\varphi^2\right)$ is not finite-dimensional over Λ; it is finite-dimensional over $\Lambda[\mathrm{h}]$, but the latter is not a field. We circumvent this difficulty by a trick which essentially amounts to setting $\mathrm{h} = 1$. (This is the point where our choice of working with polynomials in h rather than formal power series as in [39] is essential; cf. Remark 2.2.)

Consider the ungraded complex $\tilde{\mathcal{C}}$ defined as follows: $\tilde{\mathcal{C}} := \mathrm{CF}\left(\varphi^2\right) \subset \mathrm{CF}_{\mathrm{eq}}\left(\varphi^2\right)$ as a vector space over Λ with the differential $\tilde{d}\alpha := d_{\mathrm{eq}}\alpha|_{\mathrm{h}=1}$ for $\alpha \in \tilde{\mathcal{C}}$. Since d_{eq} is h-linear, we have $\tilde{d}^2 = 0$. More formally, $\tilde{\mathcal{C}}$ is the quotient complex in the short exact sequence of ungraded complexes

$$0 \longrightarrow \mathrm{CF}_{\mathrm{eq}}\left(\varphi^2\right) \xrightarrow{1+\mathrm{h}} \mathrm{CF}_{\mathrm{eq}}\left(\varphi^2\right) \xrightarrow{\pi} \tilde{\mathcal{C}} \longrightarrow 0$$

over Λ, where π is the $\mathrm{h} = 1$ evaluation map.

Remark 5.3. This exact sequence, for any action interval, gives rise to the exact triangle in Floer cohomology relating $\mathrm{H}(\tilde{\mathcal{C}})$ and $\mathrm{HF}_{\mathrm{eq}}\left(\varphi^2\right)$ via multiplication by $1+\mathrm{h}$. As any map of the form $id + O(\mathrm{h})$, this multiplication map in Floer cohomology is one-to-one, and thus

$$\mathrm{H}(\tilde{\mathcal{C}}) \cong \mathrm{HF}_{\mathrm{eq}}\left(\varphi^2\right)/(1+\mathrm{h})\,\mathrm{HF}_{\mathrm{eq}}\left(\varphi^2\right),$$

and hence, $\dim_{\mathbb{F}_2} \mathrm{H}(\tilde{\mathcal{C}}) = \mathrm{rk}_{\mathbb{F}_2[\mathrm{h}]} \mathrm{HF}_{\mathrm{eq}}\left(\varphi^2\right)$, for any action interval. For global cohomology, $\mathrm{H}(\tilde{\mathcal{C}}) \cong \mathrm{HF}\left(\varphi^2\right)$ as ungraded Λ-modules by the continuation argument and Example 2.1.

Since, by construction, $\tilde{\mathcal{C}}$ is a finite-dimensional vector space over Λ, now the machinery from [45] applies literally; see Remark 4.4. In self-explanatory notation

$$\langle d_{\mathrm{eq}}\bar{z}, \mathrm{h}^m \bar{z}'\rangle \neq 0 \quad \text{where } m = \mu(\bar{z}) - \mu(\bar{z}') + 1 \Longleftrightarrow \big\langle \tilde{d}\bar{z}, \bar{z}'\big\rangle \neq 0$$

for $\bar{z}$ and $\bar{z}'$ in $\bar{\mathcal{P}}_2(\varphi)$. Furthermore, we can also form the Floer graph for $\tilde{\mathcal{C}}$ and this graph is identical to the equivariant Floer graph $\Gamma_{\mathrm{eq}}\big(\varphi^2\big)$.

Claim 5.4. *The subset $\tilde{\mathcal{B}} := \pi(\hat{\mathcal{B}})$ in $\tilde{\mathcal{C}}$ formed by $\tilde{\alpha}_i := \pi(\hat{\alpha}_i)$ and $\tilde{\eta}_j := \pi(\hat{\eta}_j)$ and $\tilde{\gamma}_j := \pi(\hat{\gamma}_j)$ is a singular decomposition for $\tilde{\mathcal{C}}$.*

Putting aside the proof of the claim, let us first show how Lemma 5.2 follows from it. Observe that

$$\mathcal{A}(\tilde{\gamma}_j) - \mathcal{A}(\tilde{\eta}_j) = 2\big(\mathcal{A}(\gamma_j) - \mathcal{A}(\eta_j)\big). \tag{5.3}$$

Indeed, set

$$\begin{aligned} \eta_j &= \bar{x}_j + \cdots, \\ \gamma_j &= \bar{y}_j + \cdots, \end{aligned}$$

where as usual the dots stand for strictly higher action terms. (Thus, $\bar{x}_* = \bar{x}_1$ and $\bar{y}_* = \bar{y}_1$.) By Seidel's non-vanishing theorem (Proposition 2.3), we have

$$\begin{aligned} \hat{\eta}_j &= \mathrm{h}^{m_j}\bar{x}_j^2 + \cdots, \\ \hat{\gamma}_j &= \mathrm{h}^{m'_j}\bar{y}_j^2 + \cdots \end{aligned}$$

for some $m_j \geq 0$ and $m'_j \geq 0$, and hence

$$\begin{aligned} \tilde{\eta}_j &= \bar{x}_j^2 + \cdots, \\ \tilde{\gamma}_j &= \bar{y}_j^2 + \cdots . \end{aligned}$$

Therefore

$$\mathcal{A}(\tilde{\gamma}_j) - \mathcal{A}(\tilde{\eta}_j) = \mathcal{A}\big(\bar{y}_j^2\big) - \mathcal{A}\big(\bar{x}_j^2\big) = 2\big(\mathcal{A}(\bar{y}_j) - \mathcal{A}(\bar{x}_j)\big) = 2\big(\mathcal{A}(\gamma_j) - \mathcal{A}(\eta_j)\big),$$

which proves (5.3).

In particular, similarly to (5.1), we have

$$\mathcal{A}(\tilde{\gamma}_1) - \mathcal{A}(\tilde{\eta}_1) < \mathcal{A}(\tilde{\gamma}_2) - \mathcal{A}(\tilde{\eta}_2) < \cdots .$$

Therefore

$$\beta_{\min}(\tilde{\mathcal{C}}) := \mathcal{A}(\tilde{\gamma}_1) - \mathcal{A}(\tilde{\eta}_1) = \mathcal{A}\big(\bar{y}_*^2\big) - \mathcal{A}\big(\bar{x}_*^2\big)$$

is the shortest bar for $\tilde{\mathcal{C}}$. As in Step 1, we infer that

$$\big\langle \tilde{d}\bar{x}_*^2, \bar{y}_*^2\big\rangle = 1.$$

Hence, there is an arrow connecting these two orbits in the Floer graph for $\tilde{\mathcal{C}}$ and this is the shortest arrow. The Floer graph for $\tilde{\mathcal{C}}$ is defined similarly and in fact identical to the equivariant Floer graph $\Gamma_{\mathrm{eq}}\big(\varphi^2\big)$. Therefore, this arrow is also the shortest arrow in $\Gamma_{\mathrm{eq}}\big(\varphi^2\big)$, completing the proof of Lemma 5.2 modulo Claim 5.4.

Proof of Claim 5.4. Since π is a homomorphism of complexes, we have $\tilde{d}\tilde{\alpha}_i = 0$ and $\tilde{d}\tilde{\eta}_j = \tilde{\gamma}_j$. Therefore, we only need to show that $\tilde{\mathcal{B}}$ is an orthogonal basis. For this, we do not need to distinguish between different types of elements of $\mathcal{B}$. Write $\mathcal{B} = \{\xi_i\}$, where $\xi_i = \bar{z}_i + \cdots$ with the dots denoting the entries of strictly higher action. Then, by the definition of $\hat{\mathcal{B}}$ and Seidel's non-vanishing theorem, $\tilde{\mathcal{B}} = \{\tilde{\xi}_i\}$ comprises the elements

$$\tilde{\xi}_i := \pi(\hat{\xi}_i) = \bar{z}_i^2 + \cdots .$$

Now, as in Example 4.1, the orthogonality for $\mathcal{B}$ is equivalent to that the orbits $\bar{z}_i$ are distinct. Similarly, the orthogonality for $\tilde{\mathcal{B}}$ is equivalent to that the orbits $\bar{z}_i^2$ are again distinct. It follows that $\tilde{\mathcal{B}}$ is orthogonal if (in fact, iff) $\mathcal{B}$ is orthogonal which is a part of its definition. As a consequence, $\tilde{\mathcal{B}}$ is linearly independent over Λ.

Finally, since $\tilde{\mathcal{C}} = \mathrm{CF}\left(\varphi^2\right)$ as Λ-modules and φ is 2-perfect, we have

$$\dim_\Lambda \tilde{\mathcal{C}} = \dim_\Lambda \mathrm{CF}\left(\varphi^2\right) = \dim_\Lambda \mathrm{CF}(\varphi) = |\mathcal{B}| = |\tilde{\mathcal{B}}|,$$

and $\tilde{\mathcal{B}}$ is a basis. □

This concludes the proof of Lemma 5.2. □

Step 3: Removing the background assumption. Recall that the Floer graphs $\Gamma(\varphi)$ and $\Gamma_{\mathrm{eq}}\left(\varphi^2\right)$ are stable under small perturbations of φ. With this in mind, we can replace φ by a C^∞-small perturbation $\tilde{\varphi}$ meeting the background assumption, since the latter is a C^∞-generic condition. More precisely, one can change the action of a single orbit by a small amount (positive or negative) using a localized C^∞-small perturbation $\tilde{\varphi}$. Hence, given any arrow in the Floer graphs $\tilde{\Gamma}(\varphi)$ and $\tilde{\Gamma}_{\mathrm{eq}}\left(\varphi^2\right)$, pick some small $\epsilon > 0$. Then, one can apply local perturbations at the two ends to shorten its length by 2ϵ while not changing the lengths of the remaining arrows more than ϵ. It follows that every shortest arrow in the Floer graphs $\tilde{\Gamma}(\varphi)$ and $\tilde{\Gamma}_{\mathrm{eq}}\left(\varphi^2\right)$ can be perturbed into the unique shortest arrow. Now, Theorem 3.1 for φ follows from that theorem for $\tilde{\varphi}$.

Remark 5.5. (The $\mathbb{Z}_p$-equivariant analogue) This argument extends with only very minor changes to the pth iterates φ^p, where p is a prime, proving the analogue of Theorem 3.1 for $\mathbb{Z}_p$-equivariant cohomology of φ^p over $\mathbb{F}_p$ and relying on the results from [43]; cf. Remark 2.4. As a consequence, as in the proof of Theorem 1.1, if φ is strongly non-degenerate, β is *a priori* bounded from above and $|\mathcal{P}(\varphi)|$ is greater than the sum of Betti numbers of M over $\mathbb{Q}$, then there exists a simple p-periodic orbit for every sufficiently large prime p as is shown in [40].

5.2. Degenerate case

Perhaps, the simplest way to extend our arguments and, in particular, Theorem 1.1 and Corollary 1.2 to include some degenerate Hamiltonian diffeomorphisms as in [40] is by bypassing Theorem 3.1 and using a somewhat less precise argument. Below, we outline the key steps of this generalization, some of which again overlap with [40]. The account is deliberately brief. The main new point here is the construction of the (equivariant) Floer graph in the degenerate case.

Assume that φ is 2-perfect and that the second iteration is admissible: -1 is not an eigenvalue of $D\varphi_x$ for any $x \in \mathcal{P}_1(\varphi)$. (The latter requirement is satisfied once φ is replaced by its sufficiently high iterate φ^{2^k}.) Then, as shown in [19], for every $\bar{x} \in \bar{\mathcal{P}}_1(\varphi)$, we have a canonical isomorphism in local Floer cohomology

$$\mathrm{HF}(\bar{x}) \xrightarrow{\cong} \mathrm{HF}\left(\bar{x}^2\right) \tag{5.4}$$

up to a shift of grading. By the Smith inequality in local Floer cohomology, which can be proved by exactly the same argument as in [39] (see also [8,40]), we have $\mathrm{HF}_{\mathrm{eq}}\left(\bar{x}^2\right) \cong \mathrm{HF}\left(\bar{x}^2\right)[\mathrm{h}]$, where, strictly speaking, on the left, we have the graded module associated with the h-adic filtration of $\mathrm{HF}_{\mathrm{eq}}\left(\bar{x}^2\right)$. (We expect that in this situation $d_{\mathrm{eq}} = d_{\mathrm{Fl}}$, and hence $\mathrm{HF}_{\mathrm{eq}}\left(\bar{x}^2\right) \cong \mathrm{HF}\left(\bar{x}^2\right)[\mathrm{h}]$ literally, without passing to graded modules, but we have not been able to prove this.)

For every $\bar{x} \in \bar{\mathcal{P}}_1(\varphi)$, fix a basis $\xi_{i,\bar{x}}$ in $\mathrm{HF}(\bar{x})$, so that this system of bases is recapping-invariant. Applying (5.4) to this system, we obtain bases $\xi'_{i,\bar{x}}$ in $\mathrm{HF}\left(\bar{x}^2\right)$ with $\bar{x} \in \bar{\mathcal{P}}_1(\varphi)$, and this system extends to a recapping-invariant system over the entire $\bar{\mathcal{P}}_2(\varphi)$.

We also have a recapping-invariant system of bases in $\mathrm{HF}_{\mathrm{eq}}\left(\bar{x}^2\right)$ arising from $\wp(\xi_{i,\bar{x}} \otimes \xi_{i,\bar{x}}) \in \mathrm{HF}_{\mathrm{eq}}\left(\bar{x}^2\right)$. To be more precise, it is convenient to replace the equivariant cohomology (local or global) by the homology of the ungraded complex $\tilde{\mathcal{C}}$ obtained by setting $\mathrm{h} = 1$ as in the proof of Theorem 3.1. For the sake of brevity, we keep the notation $\mathrm{HF}_{\mathrm{eq}}$ for this cohomology suppressing the projection π in the notation. Set $\xi^{\mathrm{eq}}_{\bar{x},i} := \wp(\xi_{i,\bar{x}} \otimes \xi_{i,\bar{x}})$. We claim that this is a basis in $\mathrm{HF}_{\mathrm{eq}}\left(\bar{x}^2\right)$ which is now just a vector space over $\mathbb{F}_2$. Then, extending, we get a recapping invariant family of bases over $\bar{\mathcal{P}}_2(\varphi)$.

To show that $\{\xi^{\mathrm{eq}}_{\bar{x},i}\}$ is indeed a basis, we first recall that, without changing $D\varphi_x$ and the local cohomology, φ can be deformed near x to the direct product of degenerate and totally non-degenerate maps; see [19, Sect. 4.5]. This essentially reduces the question to the case, which for the sake of brevity, we will focus on, where x is totally degenerate, i.e., all eigenvalues of $D\varphi_x$ are equal to 1 and in particular φ can be made C^1-close to the identity. Furthermore, recall that $\mathrm{HF}(\bar{x}) \cong \mathrm{HF}(\varphi_f) \cong \mathrm{HM}(f)$ by [17, Sect. 3.3 and 6], where HM stands for the local Morse cohomology, f is the generating function of φ and φ_f is the germ of the Hamiltonian diffeomorphism generated by f. These isomorphisms come from continuation maps and there are similar isomorphisms (equivariant and non-equivariant) for $\bar{x}^2$ and $\varphi_{2f} = \varphi_f^2$, where we can replace the generating function for φ^2 by $2f$; see [19, Sect. 4.3]. Now, as in Example 2.1 and Remark 5.3, we arrive at the continuation map identifications

$$\mathrm{HF}_{\mathrm{eq}}\left(\bar{x}^2\right) \cong \mathrm{HF}\left(\bar{x}^2\right) \cong \mathrm{HF}(\bar{x}) \cong \mathrm{H}(Y_f), \tag{5.5}$$

where Y_f is a certain topological space (the Conley index) associated with the critical point x of f. Furthermore, the map $\alpha \mapsto \wp(\alpha \otimes \alpha)$ turns into the Steenrod square Sq on $\mathrm{H}(Y_f)$; see [47]. Thus, with these identifications in mind, $\xi_{\bar{x},i} = \xi'_{\bar{x},i}$ and

$$\xi^{\rm eq}_{\bar{x},i} = \mathrm{Sq}(\xi_{\bar{x},i}) = \xi_{\bar{x},i} + \cdots, \tag{5.6}$$

where the dots stand for the terms of *higher degree* in $\mathrm{H}(Y_f)$. It follows that the vectors $\xi^{\rm eq}_{\bar{x},i}$ are linearly independent and, since $\dim_{\mathbb{F}_2} \mathrm{HF}_{\rm eq}\big(\bar{x}^2\big) = \dim_{\mathbb{F}_2} \mathrm{HF}(\bar{x})$ by (5.5), this system is a basis.

The action filtration spectral sequence in Floer cohomology has $E_1 = \bigoplus_{\bar{x}} \mathrm{HF}(\bar{x})$ and converges to $\mathrm{HF}(\varphi)$. With bases fixed, we can canonically collapse this spectral sequence into one complex with the same features as the ordinary Floer complex including the action filtration and cohomology equal to $\mathrm{HF}(\varphi)$; cf. [25, Sect. 2.1.3 and 2.5]. These data are sufficient to define the Floer graph $\Gamma(\varphi)$ of φ with vertices $\xi_{\bar{x},i}$. (Note that the orbits with $\mathrm{HF}(\bar{x}) = 0$ do not contribute to $\Gamma(\varphi)$ and the graph depends on the choice of the bases $\{\xi_{\bar{x},i}\}$.) It is also worth keeping in mind that even in the non-degenerate case, this graph and the complex might differ from the Floer graph as defined in Sect. 3 and from the Floer complex. However, they have the same formal properties as $\mathrm{CF}(\varphi)$ and the original graph, and the resulting homology is isomorphic to the Floer cohomology $\mathrm{HF}(\varphi)$; cf. [25].

A similar construction applies to φ^2 in the ordinary and equivariant settings and $\xi'_{\bar{x},i} \leftrightarrow \xi^{\rm eq}_{\bar{x},i}$ gives rise to an action-preserving one-to-one correspondence between the vertices of $\Gamma\big(\varphi^2\big)$ and $\Gamma_{\rm eq}\big(\varphi^2\big)$. The condition that the sum (1.1) with $\mathbb{F} = \mathbb{F}_2$ is strictly greater than the sum of Betti numbers guarantees that the graph $\Gamma(\varphi)$, and hence $\Gamma\big(\varphi^2\big)$ and $\Gamma_{\rm eq}\big(\varphi^2\big)$, have at least one arrow.

Denote by $\beta_{\min}$ the length of the shortest arrows in a Floer graph. Our goal is to show that φ cannot be 2^k-perfect, where k is sufficiently large, assuming an a priori upper bound on $\beta_{\min}\big(\varphi^{2^k}\big)$ as in Theorem 1.1. (Note that in contrast with the non-degenerate case, the Floer graphs are now sensitive to small perturbations of φ and we usually cannot make the shortest arrow unique without changing the graph unless $\dim_{\mathbb{F}_2} \mathrm{HF}(x) = 1$ for all $x \in \mathcal{P}_1(\varphi)$.)

The equivariant pair-of-pants product $\wp$ extends to the complexes we have constructed, and Seidel's non-vanishing theorem takes the form

$$\wp(\xi_{\bar{x},i} \otimes \xi_{\bar{x},i}) = \xi^{\rm eq}_{\bar{x},i} + \cdots, \tag{5.7}$$

where now the dots stand for terms with action greater than or equal to the action of $\xi^{\rm eq}_{\bar{x},i}$, but with the provision that the first term enters the whole sum with non-zero coefficient. (This is a consequence of (5.6) and Seidel's non-vanishing theorem applied to the non-degenerate part in the splitting of φ at x.)

Pick one of the shortest arrows, say v, in $\Gamma_{\rm eq}\big(\varphi^2\big)$. After recapping, we can ensure that the beginning of v has the form $\xi^{\rm eq}_{\bar{x},i}$. Using (5.7) and the facts that $\wp$ is a chain map and v is a shortest arrow, it is not hard to see that $\xi_{\bar{x},i}$ is the beginning of an arrow in $\Gamma(\varphi)$ whose length is at most $\beta^{\rm eq}_{\min}\big(\varphi^2\big)/2$. Hence

$$2\beta_{\min}(\varphi) \le \beta^{\rm eq}_{\min}\big(\varphi^2\big). \tag{5.8}$$

(This proves a somewhat weaker version of Theorem 3.1: every shortest equivariant arrow comes from an arrow for φ.)

On the other hand

$$\beta^{\mathrm{eq}}_{\min}\big(\varphi^2\big) \leq \beta_{\min}\big(\varphi^2\big). \tag{5.9}$$

Indeed, $\dim_{\mathbb{F}_2} \mathrm{HF}^I\left(\varphi^2\right) \geq \mathrm{rk}_{\mathbb{F}_2[\mathrm{h}]} \mathrm{HF}^I_{\mathrm{eq}}\left(\varphi^2\right)$ for any action interval I, as is easy to see from the h-adic filtration spectral sequence. Applying this to an interval tightly enclosing one of the shortest arrows in $\Gamma\big(\varphi^2\big)$, we obtain (5.9). In fact, we expect that, as in the non-degenerate case, $\Gamma_{\mathrm{eq}}\big(\varphi^2\big)$ incorporates all arrows of $\Gamma\big(\varphi^2\big)$ (and, perhaps, more). This is a stronger statement than (5.9), but (5.9) is sufficient for our purposes.

Combining (5.8) and (5.9), we see that $2\beta_{\min}(\varphi) \leq \beta_{\min}\big(\varphi^2\big)$. (This inequality can also be extracted from some of the results in [40].) As a consequence, $\beta_{\min}\big(\varphi^{2^k}\big) \geq 2^k\beta_{\min}(\varphi)$ as long as φ is 2^k-perfect. When $\beta_{\min}\big(\varphi^{2^k}\big)$ is bounded from above, this is impossible for large k.

We note in conclusion that in the non-degenerate case, this proof reduces to an argument which does not rely on persistence homology and is ultimately simpler and more direct, although arguably less structured, than our proof of Theorem 1.1 via Theorem 3.1.

Acknowledgements

The authors are grateful to Egor Shelukhin for useful discussions.

References

[1] Atallah, M.S., Shelukhin, E.: Hamiltonian no-torsion. Preprint. arXiv:2008.11758

[2] Banik, M.: On the dynamics characterization of complex projective spaces. J. Fixed Point Theory Appl. **22**(1) (2020). https://doi.org/10.1007/s11784-020-0760-5

[3] Batoreo, M.: On periodic points of symplectomorphisms on surfaces. Pac. J. Math. **294**, 19–40 (2018)

[4] Bramham, B.: Periodic approximations of irrational pseudo-rotations using pseudoholomorphic curves. Ann. Math. (2) **181**, 1033–1086 (2015)

[5] Bramham, B.: Pseudo-rotations with sufficiently Liouvillean rotation number are C^0-rigid. Invent. Math. **199**, 561–580 (2015)

[6] Bramham, B., Hofer, H.: First steps towards a symplectic dynamics. Surv. Differ. Geom. **17**, 127–178 (2012)

[7] Çineli, E.: Conley conjecture and local Floer homology. Arch. Math. (Basel) **111**, 647–656 (2018)

[8] Çineli, E., Ginzburg, V.L.: On the iterated Hamiltonian Floer homology. Commun. Contemp. Math. https://doi.org/10.1142/S0219199720500261

[9] Çineli, E., Ginzburg, V.L., Gürel, B.Z.: Pseudo-rotations and holomorphic curves. Sel. Math. **26**, 78 (2020). Preprint arXiv:1905.07567

[10] Çineli, E., Ginzburg, V.L., Gürel, B.Z.: From pseudo-rotations to holomorphic curves via quantum Steenrod squares. Int. Math. Res. Not. (2020). https://doi.org/10.1093/imrn/rnaa173

[11] Collier, B., Kerman, E., Reiniger, B., Turmunkh, B., Zimmer, A.: A symplectic proof of a theorem of Franks. Compos. Math. **148**, 1969–1984 (2012)

[12] Entov, M., Polterovich, L.: Calabi quasimorphism and quantum homology. Int. Math. Res. Not. **2003**(30), 1635–1676 (2003). https://doi.org/10.1155/S1073792803210011

[13] Fayad, B., Katok, A.: Constructions in elliptic dynamics. Ergod. Theory Dyn. Syst. **24**, 1477–1520 (2004)

[14] Franks, J.: Geodesics on S^2 and periodic points of annulus homeomorphisms. Invent. Math. **108**, 403–418 (1992)

[15] Franks, J.: Area preserving homeomorphisms of open surfaces of genus zero. N. Y. J. Math. **2**, 1–19 (1996)

[16] Franks, J., Handel, M.: Periodic points of Hamiltonian surface diffeomorphisms. Geom. Topol. **7**, 713–756 (2003)

[17] Ginzburg, V.L.: The Conley conjecture. Ann. Math. (2) **172**, 1127–1180 (2010)

[18] Ginzburg, V.L., Gürel, B.Z.: On the generic existence of periodic orbits in Hamiltonian dynamics. J. Mod. Dyn. **4**, 595–610 (2009)

[19] Ginzburg, V.L., Gürel, B.Z.: Local Floer homology and the action gap. J. Symplectic Geom. **8**, 323–357 (2010)

[20] Ginzburg, V.L., Gürel, B.Z.: Hyperbolic fixed points and periodic orbits of Hamiltonian diffeomorphisms. Duke Math. J. **163**, 565–590 (2014)

[21] Ginzburg, V.L., Gürel, B.Z.: The Conley conjecture and beyond. Arnold Math. J. **1**, 299–337 (2015)

[22] Ginzburg, V.L., Gürel, B.Z.: Conley conjecture revisited. Int. Math. Res. Not. (2017). https://doi.org/10.1093/imrn/rnx137

[23] Ginzburg, V.L., Gürel, B.Z.: Hamiltonian pseudo-rotations of projective spaces. Invent. Math. **214**, 1081–1130 (2018)

[24] Ginzburg, V.L., Gürel, B.Z.: Pseudo-rotations vs. rotations. Preprint arXiv:1812.05782

[25] Ginzburg, V.L., Gürel, B.Z.: Lusternik–Schnirelmann theory and closed Reeb orbits. Math. Z. (2019). https://doi.org/10.1007/s00209-019-02361-2

[26] Ginzburg, V.L., Kerman, E.: Homological resonances for Hamiltonian diffeomorphisms and Reeb flows. Int. Math. Res. Not. **2010**(1), 53–68 (2010). https://doi.org/10.1093/imrn/rnp120

[27] Gürel, B.Z.: On non-contractible periodic orbits of Hamiltonian diffeomorphisms. Bull. Lond. Math. Soc. **45**, 1227–1234 (2013)

[28] Gürel, B.Z.: Periodic orbits of Hamiltonian systems linear and hyperbolic at infinity. Pac. J. Math. **271**, 159–182 (2014)

[29] Hofer, H., Zehnder, E.: Symplectic Invariants and Hamiltonian Dynamics. Birkäuser, Basel (1994)

[30] Kerman, E.: On primes and period growth for Hamiltonian diffeomorphisms. J. Mod. Dyn. **6**, 41–58 (2012)

[31] Kislev, A., Shelukhin, E.: Bounds on spectral norms and applications. Preprint. arXiv:1810.09865

[32] Le Calvez, P.: Periodic orbits of Hamiltonian homeomorphisms of surfaces. Duke Math. J. **133**, 125–184 (2006)

[33] Le Roux, F., Seyfaddini, S.: The Anosov-Katok method and pseudo-rotations in symplectic dynamics. Preprint. arXiv:2010.06237

[34] McDuff, D., Salamon, D.: J-Holomorphic Curves and Symplectic Topology. Colloquium Publications, vol. 52. AMS, Providence (2004)

[35] Polterovich, L., Rosen, D.: Function Theory on Symplectic Manifolds. CRM Monograph Series, vol. 34. American Mathematical Society, Providence (2014)

[36] Polterovich, L., Shelukhin, E.: Autonomous Hamiltonian flows, Hofer's geometry and persistence modules. Sel. Math. (N.S.) **22**, 227–296 (2016)

[37] Salamon, D.A.: Lectures on Floer homology. In: Eliashberg, Y., Traynor, L. (eds.) Symplectic Geometry and Topology. IAS/Park City Mathematics Series, vol. 7, pp. 143–229. American Mathematical Society, Providence (1999)

[38] Salamon, D., Zehnder, E.: Morse theory for periodic solutions of Hamiltonian systems and the Maslov index. Commun. Pure Appl. Math. **45**, 1303–1360 (1992)

[39] Seidel, P.: The equivariant pair-of-pants product in fixed point Floer cohomology. Geom. Funct. Anal. **25**, 942–1007 (2015)

[40] Shelukhin, E.: On the Hofer–Zehnder conjecture. Preprint. arXiv:1905.04769

[41] Shelukhin, E.: Pseudorotations and Steenrod squares. J. Mod. Dyn. (to appear). Preprint. arXiv:1905.05108

[42] Shelukhin, E.: Pseudo-rotations and Steenrod squares revisited. Preprint. arXiv:1909.12315

[43] Shelukhin, E., Zhao, J.: The $\mathbb{Z}/(p)$-equivariant product-isomorphism in fixed point Floer cohomology. Preprint. arXiv:1905.03666

[44] Usher, M.: Boundary depth in Floer theory and its applications to Hamiltonian dynamics and coisotropic submanifolds. Isr. J. Math. **184**, 1–57 (2011)

[45] Usher, M., Zhang, J.: Persistent homology and Floer–Novikov theory. Geom. Topol. **20**, 3333–3430 (2016)

[46] Wilkins, N.: Quantum Steenrod squares and the equivariant pair-of-pants in symplectic cohomology. Preprint. arXiv:1810.02738

[47] Wilkins, N.: A construction of the quantum Steenrod squares and their algebraic relations. Geom. Topol. **24**, 885–970 (2020)

Erman Çineli and Viktor L. Ginzburg
Department of Mathematics
UC Santa Cruz
Santa Cruz CA
95064
USA
e-mail: scineli@ucsc.edu;
ginzburg@ucsc.edu

Başak Z. Gürel
Department of Mathematics
University of Central Florida
Orlando FL
32816
USA
e-mail: basak.gurel@ucf.edu

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:49
https://doi.org/10.1007/s11784-022-00942-z
Published online June 10, 2022

Journal of Fixed Point Theory and Applications

Higher symplectic capacities and the stabilized embedding problem for integral elllipsoids

Daniel Cristofaro-Gardiner, Richard Hind and Kyler Siegel

Abstract. The third named author has been developing a theory of "higher" symplectic capacities. These capacities are invariant under taking products, and so are well suited for studying the stabilized embedding problem. The aim of this note is to apply this theory, assuming its expected properties, to solve the stabilized embedding problem for integral ellipsoids, when the eccentricity of the domain has the opposite parity of the eccentricity of the target and the target is not a ball. For the other parity, the embedding we construct is definitely not always optimal; also, in the ball case, our methods recover previous results of McDuff, and of the second named author and Kerman. There is a similar story, with no condition on the eccentricity of the target, when the target is a polydisc: a special case of this implies a conjecture of the first named author, Frenkel, and Schlenk concerning the rescaled polydisc limit function. Some related aspects of the stabilized embedding problem and some open questions are also discussed.

Mathematics Subject Classification. 53D.

Contents

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

1. Introduction

1.1. The main results

Let X_1 and X_2 be four-dimensional symplectic manifolds. There has recently been considerable interest in understanding the **stabilized symplectic embedding problem**, namely the question of whether or not there exists a symplectic embedding

$$X_1 \times \mathbb{C}^N \overset{s}{\hookrightarrow} X_2 \times \mathbb{C}^N. \tag{1}$$

Indeed, certain techniques which are available for studying four-dimensional embedding problems do not have a clear analogue in higher dimensions, and so it is interesting to understand how different the stabilized problem is from the four-dimensional one. For more about the problem, we refer the reader to [7,8,14,15,21], the references therein, and the discussion below.

The embedding problem (1) is already quite subtle when X_1 and X_2 are simple shapes, like **ellipsoids**

$$E(a,b) := \left\{ \frac{\pi|z_1|^2}{a} + \frac{\pi|z_2|^2}{b} \leq 1 \right\} \subset \mathbb{C}^2,$$

balls $B(c) := E(c,c)$, **polydiscs**

$$P(a,b) := \left\{ \frac{\pi|z_1|^2}{a} \leq 1, \frac{\pi|z_2|^2}{b} \leq 1 \right\} \subset \mathbb{C}^2,$$

and **cubes** $C(c) := P(c,c)$. (Here, $\mathbb{C}^N$ is equipped with its standard symplectic form.) For example, what is known about the stabilized ellipsoid-into-ball problem has a curious mix of rigidity and flexibility: much about this question remains unknown. In contrast, the stabilized polydisc-into-ball problem is completely solved [29, Thm. 1.3.5] (for another approach, see [13]) and the answer is described by a very simple function, namely a piecewise linear function with two pieces.

The starting point for our investigations here is the stabilized ellipsoid-into-ellipsoid problem. This is a special case of Problem 44 in the influential problem list [22, Ch. 14] by McDuff and Salamon, which asks for a solution to the symplectic embedding problem for $2n$-dimensional symplectic ellipsoids: we can view stabilized ellipsoids as $2n$-dimensional ellipsoids with most arguments infinite. Consider the function $c^N_{b,ell}(a)$, defined to be the infimum, over λ, such that there exists an embedding

$$E(1,a) \times \mathbb{C}^N \overset{s}{\hookrightarrow} \lambda \cdot E(1,b) \times \mathbb{C}^N, \tag{2}$$

where we write $\lambda \cdot E(a,b)$ for $E(\lambda a, \lambda b)$. This function for $a, b \geq 1$ completely determines the stabilized ellipsoid-into-ellipsoid problem, and we would ideally like to compute it.

At present, this looks out of reach. As mentioned above, even the case $b = 1$ seems quite subtle; in fact, it is the focus of a conjecture by McDuff [21]. And, when $b > 1$, almost nothing is currently known. However, it turns out that when a and b are integers, there is a lot more traction.

Theorem 1.1. *Assume that $b > 1$ is an integer, and let $a \geq b+1$ be any integer with parity the opposite of b. Then, for $N \geq 1$*

$$c^N_{b,ell}(a) = \frac{2a}{a+b-1}.$$

We discuss the hypothesis $a \geq b+1$ here in Sect. 1.2.2, where we show that it is essentially necessary.

A key aspect of our proof of the above theorem, which is one of the motivations for writing this note, involves the obstructions required to prove it. Symplectic embedding problems are profitably studied by **symplectic capacities**; see, e.g., [3]. The third named author has recently defined a new sequence of symplectic capacities $\mathfrak{g}_k$ which play a starring role here. These capacities $\mathfrak{g}_k$ are invariant under taking products with $\mathbb{C}$ and so give obstructions to the stabilized problem. As we will see in the proof of Theorem 1.1, the $\mathfrak{g}_k$ are very well adapted to proving Theorem 1.1, and the obstructive side of the proof follows quite quickly once we can marshal them to our benefit. The constructive side of the proof comes from a variant of the stabilized folding construction pioneered by the second named author.

Disclaimer 1.2. Our high-level discussion of symplectic capacities in Sect. 2 follows [28], which in turn assumes the existence of rational symplectic field theory with its expected functoriality properties as outlined in [9]. Apart from simple special cases, such a formalism is known to require a virtual perturbation framework such as the theory of polyfolds; for the current status of this and related projects, we refer the reader to, e.g., [1,10,16,18,19,26] and the references therein.

The proofs of our main results on embedding obstructions in Sect. 3 take the properties of the capacities $\mathfrak{g}_k$ summarized in Theorem 2.1 as a black box, together with some computations from [28] which we recall in Sect. 3.1.2. Our proof of Theorem 1.1 furthermore requires the formula for $\mathfrak{g}_k(E(1,a))$ which will appear in the forthcoming work [25]. The latter reference also constructs an ersatz version of these capacities in the special case of ellipsoids without appealing to virtual perturbations; these give equivalent obstructions for stabilized embeddings between four-dimensional ellipsoids, and the method also readily adapts to the case of ellipsoid domain and polydisk target. Our proof of Proposition 1.7 further depends on the formalism of [29], which is based on [28] and the forthcoming [30].

In dimension four, when b is integral, there is an equivalence of embeddings

$$E(1,a) \overset{s}{\hookrightarrow} \lambda P(1,b), \quad E(1,a) \overset{s}{\hookrightarrow} \lambda E(1,2b), \tag{3}$$

that is, one of these embeddings exists if and only if the other does, see for example [6, Rmk. 1.2.1]. Therefore, it is natural to compare Theorem 1.1 with the stabilized ellipsoid-into-polydisc problem. Here, we get a somewhat parallel, but in fact stronger result. Define $c^N_{b,poly}(a)$ to be the infimum, over λ, such that an embedding

$$E(1,a) \times \mathbb{C}^N \overset{s}{\hookrightarrow} \lambda \cdot P(1,b) \times \mathbb{C}^N \tag{4}$$

exists.

Theorem 1.3. *Let $a \geq 2b-1$ be any odd integer. Then, for $N \geq 1$*

$$c^N_{b,poly}(a) = \frac{2a}{a+2b-1}.$$

We remark that, in contrast to Theorem 1.1, there is no requirement here that b is an integer. As with the previous theorem, the hypothesis $a \geq 2b-1$ is discussed in Sect. 1.2.2, where it is shown to be necessary.

1.2. Applications and remarks

1.2.1. Steps and the rescaled embedding function. One of our motivations for studying Theorem 1.3 is that it readily implies a conjecture of the first author, Frenkel, and Schlenk about the stabilized ellipsoid-into-polydisc function, namely Conjecture 1.4 in [6], which we now explain.

First, we explain the motivation behind that conjecture. As alluded to above, at present, fully computing the function $c^N_{b,poly}(a)$ for $N \geq 1$ seems quite difficult. However, there is a related function, called the **rescaled limit function** $\hat{c}^N_{b,poly}$, see (5), that looks more tractable and in particular could be computed given a resolution of the aforementioned Conjecture 1.4.

To elaborate, the function $c^0_{b,poly}(a)$ for $b \in \mathbb{Z}_{\geq 2}$ was previously computed by the first author, Frenkel and Schlenk in [6]. It was shown that the function $c^0_{b,poly}(a)$ is given by the volume constraint $\sqrt{\frac{a}{2b}}$, except on finitely many intervals. On all but one of these intervals, the function $c^0_{b,poly}(a)$ is given by a "linear step": it is piecewise linear, with a single non-smooth point, called its corner, where its graph changes from lying on a line through the origin to being horizontal. On the remaining interval, it is also piecewise linear with a single non-smooth point, but the linear piece does not lie on a line through the origin—it has an intercept, and so we call it the "affine step". For more detail, see [6].

Conjecture 1.4 asserts that the linear steps from above are "stable". Of course, for any a, we have $c^N_{b,poly}(a) \leq c^0_{b,poly}(a)$, by taking the product with the identity mapping. The conjecture, then, is that for a in the domain of the linear steps, we have $c^N_{b,poly}(a) = c^0_{b,poly}(a)$. To state that conjecture precisely, we define, for $k \in \{0, 1, 2, \ldots, \lfloor\sqrt{2b}\rfloor\}$, the numbers

$$u_b(k) = \frac{(2b+k)^2}{2b}, \quad v_b(k) = 2b\left(\frac{2b+2k+1}{2b+k}\right)^2.$$

We always have $u_b(k) < v_b(k)$ except if $k^2 = 2b$; for $u_b(k) < a < v_b(k)$, the graph of $c^N_{b,poly}(a)$ is precisely the linear steps mentioned above.

Corollary 1.4. (Conj. 1.4 of [6]) *Assume that b is an integer and*

$$u_b(k) \leq a \leq v_b(k).$$

Then

$$c^0_{b,poly}(a) = c^N_{b,poly}(a) = c^0_{2b,ell}(a) = c^N_{2b,ell}(a).$$

The final two equalities here, concerning the ellipsoid-into-ellipsoid function, were not actually part of Conjecture 1.4; however, they fall out immediately from our proof.

We now state the relevance of this to the rescaled limit function. The background is that [6] defined[1] the rescaled functions

$$\hat{c}^N_{b,poly}(a) := 2bc^N_{b,poly}(a+2b) - 2b, \quad a \geq 0, \tag{5}$$

to capture the qualitative behavior of the obstructive part of the embedding function $c^0_{b,poly}$ that goes beyond Gromov's non-squeezing theorem. It was shown in [6, Eq. 1.3] that the functions $\hat{c}^0_{b,poly}(a)$ converge, as $b \to \infty$, uniformly on bounded sets to a pleasing answer, namely the "infinite regular staircase" described by the function $c_\infty(a) : [0,\infty) \to \mathbb{R}$ whose graph consists of infinitely many linear steps of width 2; see [6, Fig. 1.7] and Fig. 1 below. For more about the motivation for studying the rescaled function, we refer the reader to the discussion in [6, Sec. 1.2].

Corollary 1.5. *The rescaled limit function is stable. That is, for any $N \in \mathbb{Z}_{\geq 0}$ and integral b, we have*

$$\lim_{b\to\infty} \hat{c}^N_{b,poly}(a) = c_\infty(a), \quad a \in [0,\infty)$$

uniformly on bounded sets.

We will explain the proofs of these corollaries in Sect. 3.2.

[1] Actually, only the $N = 0$ case of these functions was defined, but the definition extends verbatim to general N, and that will be our working definition here.

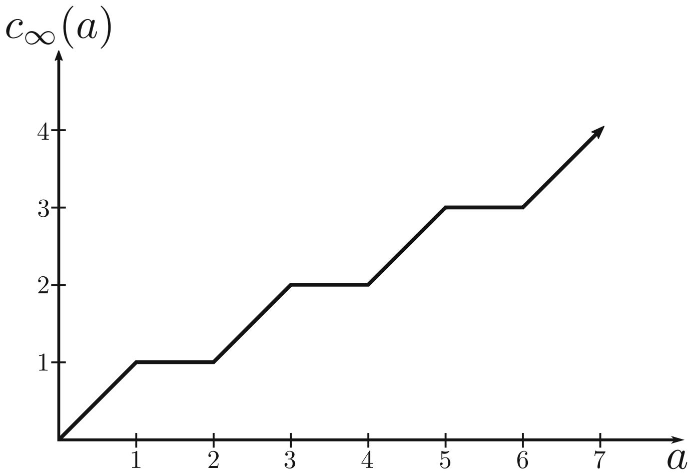


FIGURE 1. The rescaled limit function. Each step has width two, and consists of a line of slope one and a horizontal line

1.2.2. The first step. We next remark that, in the context of Theorem 1.1, the lower bound on a is essentially necessary. Indeed, if $a \leq b$, then inclusion gives an embedding which Gromov's non-squeezing theorem shows is optimal. That is, $c^N_{b,ell}(a) = 1$ for all $N \geq 0$. There is a similar story for Theorem 1.3 for $a \leq 2b-1$, but it requires a more interesting embedding. With a little more work, we can extend the range of a to work out at least part[2] of the "first step" of the embedding functions considered in this note.

Proposition 1.6. *Let $b \in \mathbb{R}_{\geq 1}$. Then:*

- *The function $c^N_{b,ell}$ starts as follows:*
  - *We have $c^N_{b,ell}(a) = 1, \quad 1 \leq a \leq b$.*
  - *We have $c^N_{b,ell}(a) = \frac{a}{b}, \quad b \leq a \leq \lfloor b \rfloor + 1$.*
- *The function $c^N_{b,poly}$ starts as follows. Let a_0 be the smallest odd integer that is no less than $2b-1$.*
  - *We have $c^N_{b,poly}(a) = 1, \quad 1 \leq a \leq \frac{a_0-1}{2} + b$.*
  - *We have $c^N_{b,poly}(a) = \frac{2}{a_0+2b-1}a, \quad \frac{a_0-1}{2} + b \leq a \leq a_0$.*

Note that there is no restriction above that a or b be integral, in contrast to the theorems in the previous section.

1.2.3. The case $b = 1$. In view of Theorem 1.1, it is natural to ask about the case $b = 1$. This was previously studied by McDuff [21], who proved an analogous result for any integer congruent to two, modulo three; we can recover this result with our methods, as well; see Example 1 in Sect. 3.1. Comparing our result to McDuff's, it is interesting to note the switch from three periodicity to two periodicity as b increases from one. There is a substantial

[2]In fact, Proposition 1.6 likely describes the entirety of the first step, although we do not address this here.

mystery about the structure as b ranges from 1 to 2, see Sect. 4.3, which we plan to investigate in follow-up work.

1.2.4. The other parity. In view of the above results, it is natural to ask: what happens for a an integer of a parity not covered by our theorems. We certainly do not have a satisfactory answer to this at present. However, using the more general calculus of [29], together with the aid of the computer, we can show for example:

Proposition 1.7. *For $6 \leq a \leq 100$ an even integer, the conclusion of Theorem 1.1 holds for $b = 2$, that is for $N \geq 1$, we have*

$$c^N_{2,ell}(a) = \frac{2a}{a+1}.$$

Similarly, for $6 \leq a \leq 100$ an even integer, the conclusion of Theorem 1.3 holds for $b = 1$, that is for $N \geq 1$, we have

$$c^N_{1,poly}(a) = \frac{2a}{a+1}.$$

Remark 1.8. The assumption $a \geq 6$ in Proposition 1.7 is necessary. Indeed, for a, less than the squared silver ratio $\sigma^2 \approx 5.83$, $c^0_{1,poly}(a)$ is an infinite staircase [11]. In particular, we have $c^N_{1,poly}(a) \leq c^0_{1,poly}(a)$, and $c^0_{1,poly}(a)$ is strictly less than $\frac{2a}{a+1}$ for $a = 2, 4$. The same applies for $c^N_{2,ell}$, since we have $c^0_{2,ell} = c^0_{1,poly}$.

For more examples, suppose that $a = 2b + 2k + 2$ is an even integer. Referring to Sect. 1.2.1, we see that $v_b(k) \leq a \leq u_b(k+1)$ which for $k \geq 2$ implies that $c^0_{b,poly}(a) = \sqrt{\frac{a}{2b}}$, that is, there is a volume filling embedding from $E(1, a)$ into a scaling of $P(1, b)$ (the point $a = 2b + 4$ lies in the affine step). By (3), this is equivalent to the existence of a volume filling embedding from $E(1, a)$ into a scaling of $E(1, 2b)$. Now, volume filling embeddings in dimension 4 improve on the folding construction giving Theorem 1.1 when $a < b + 1 + 2\sqrt{b}$. Hence, the conclusion of Theorem 1.1 is false when a and b are even and $b + 4 < a < b + 1 + 2\sqrt{b}$.

Structure of the note

In Sect. 2, we review the construction of the higher symplectic capacities of the third named author; our discussion here includes some informal elements to help convey the intuition. Then, in Sect. 3, we give the proofs of our results. The final Sect. 4 discusses some natural follow-up questions to this work.

2. New capacities

We first briefly review the capacities $\mathfrak{g}_k$ defined for $k \in \mathbb{Z}_{\geq 1}$ in [28]. These are part of a more general family of capacities $\mathfrak{g}_\mathfrak{b}$ indexed by elements in the symmetric tensor algebra $\overline{S}\mathbb{Q}[t] = \bigoplus_{k=1}^\infty (\otimes^k \mathbb{Q}[t])/\Sigma_k$. We give here only an impressionistic sketch, omitting some of the more technical details. In addition to the computations described in Sect. 3.1.2, the key structural properties we will need are summarized in the following:

Theorem 2.1. [28] *For any Liouville domain X and $k \in \mathbb{Z}_{\geq 1}$, we have $\mathfrak{g}_k(X) \in \mathbb{R}_{>0}$ with the following properties:*

(1) symplectomorphism invariance*: if X' is another Liouville domain which is symplectomorphic to X, we have $\mathfrak{g}_k(X) = \mathfrak{g}_k(X')$*
(2) scaling*: if X' is the Liouville domain obtained by scaling the Liouville form of X by a constant $c \in \mathbb{R}_{>0}$, we have $\mathfrak{g}_k(X') = c\mathfrak{g}_k(X)$*
(3) monotonicity*: if X' is another Liouville domain and there exists a symplectic embedding $X \overset{s}{\hookrightarrow} X'$, then we have $\mathfrak{g}_k(X) \leq \mathfrak{g}(X')$*
(4) stabilization*: we have $\mathfrak{g}_k(X \times B^2(S)) = \mathfrak{g}_k(X)$, provided that $S > \mathfrak{g}_k(X)$.*[3]

Note that (3) actually implies (1).

2.1. The first approximation

Suppose that X is a Liouville domain. We work with almost complex structures J on the symplectic completion $\widehat{X}$ which are admissible in the sense of symplectic field theory (SFT). Fix a point $p \in X$ along with a local J-holomorphic divisor D passing through p. To first approximation, $\mathfrak{g}_k(X)$ is simply the minimal energy of a punctured J-holomorphic sphere $u : \Sigma \to \widehat{X}$ with some number $l \geq 1$ of positive ends asymptotic to Reeb orbits in ∂X, such that u passes through p and is tangent to D to order $k-1$. We denote this tangency constraint by $\lll \mathcal{T}^{k-1}p \ggg$ (see [24] and the references therein for more details).

To see why this should be monotone with respect to symplectic embeddings, the basic point is that given such a curve u in $\widehat{X}$ and a symplectic embedding $X' \overset{s}{\hookrightarrow} X$, we can neck-stretch along $\partial X'$. This forces u to break into a pseudoholomorphic building consisting of

- a curve u_{top} (possibly disconnected) in the completed symplectic cobordism $\widehat{X \setminus X'}$ with the same positive asymptotics as u
- a curve u_{bot} in $\widehat{X'}$ which inherits the tangency constraint $\lll \mathcal{T}^{k-1}p \ggg$.

Since u_{bot} is a candidate minimizer for $\mathfrak{g}_k(X')$ and it has energy at most that of u, this shows that $\mathfrak{g}_k(X') \leq \mathfrak{g}_k(X)$.

2.2. Behavior under stabilization

One role of the local tangency constraint in the definition of $\mathfrak{g}_k$ is to cut down the dimension of families of curves, thereby giving access to curves of higher Fredholm index. There are certainly other natural geometric constraints which lower the index, the most obvious being to impose k distinct point constraints. In fact, doing so leads to the "rational symplectic field theory capacities" (RSFT) first considered in [17].

However, point constraints behave in a rather complicated way under dimensional stabilization. The RSFT capacities are therefore perhaps not well suited for stabilized problems (although they may have other applications yet

[3]Strictly speaking, $X \times B^2(S)$ is not a Liouville domain, since it has corners, although these can be removed by an arbitrarily small smoothing. See [28, §5.4] for a more precise formulation. Property (1) is of course automatic given property (3).

to be discovered). For example, note that each point constraint is codimension 2 when $\dim X = 4$, but is generally codimension $2n-2$ when $\dim X = 2n$. This means that the same curve with the same point constraints has negative total index after stabilizing by $\mathbb{C}^N$ with N large enough.

By contrast, local tangency constraints behave quite well with respect to stabilization. This is closely related to the observation of Hind and Kerman from [14] that punctured rational curves with exactly one negative end have stable Fredholm index. The stabilization property in Theorem 2.1 is also closely related to the stabilization theorems appearing in the works [7,8,21].

2.3. The naive chain complex

Unfortunately, the definition given in Sect. 2.1 is not particularly robust, since it might depend on the choice of almost complex structure J. Indeed, if we try to deform J to some other almost complex structure J', somewhere along the way the curve u might degenerate into a pseudoholomorphic building and then disappear. Therefore, to get something which is truly a symplectomorphism invariant, we have to be a bit more "homological". This is where the chain complexes coming from Floer theory or symplectic field theory become essential.

The idea is to associate with X a filtered chain complex $C(X)$, where

- as a vector space, $C(X)$ is the (graded) polynomial algebra on the (not necessarily primitive) Reeb orbits of ∂X
- the differential is defined by counting rigid-up-to-translation connected rational curves in $\mathbb{R} \times \partial X$ with several positive ends and one negative end
- the filtration is by the symplectic action functional, or equivalently by the periods of Reeb orbits.

Similarly, given an exact[4] symplectic cobordism W with positive end $\partial^+ W = \partial X$ and negative end $\partial^- W = \partial X'$, we define a chain map from $C(X)$ to $C(X')$ by counting rigid possibly disconnected rational curves in W, such that each component has several positive ends and one negative end. By Stokes' theorem, both the differential and the cobordism map are action-nondecreasing and hence preserve the filtrations.

However, the above prescription does not work on face value due to transversality issues. Namely, to show that the differential squares to zero and that the cobordism map is a chain map, the typical strategy is to analyze analogous moduli spaces of dimension one and show that (after compactifying) their boundaries give precisely the desired relations. However, it is well known that the relevant SFT moduli spaces are rarely transversely cut out for any choice of generic J. Multiply covered curves tend to appear with higher-than-expected dimension, and this spoils our strategy.

2.4. Input from symplectic field theory

One way is get around this issue is to count curves in a "virtual" sense, by introducing suitable abstract perturbations which allow more room to achieve

[4]There is also a nice story extending the theory to non-exact symplectic cobordisms, but we will ignore this for simplicity.

transversality. This is the basic strategy being pursued to define SFT in full generality by various groups, with much recent progress but consensus not yet achieved (see, e.g., [1,10,16,18,19,26] and the references therein).

In the setting of SFT, the desired invariant $C(X)$ can be written as $\bar{C}H_{\text{lin}}(X)$. Here, $CH_{\text{lin}}(X)$ is the linearized contact homology of X, which is roughly the chain complex generated by Reeb orbits of ∂X with differential counting cylinders in the symplectization $\mathbb{R} \times \partial X$.[5] Linearized contact homology only involves curves with one positive end, but by incorporating curves with several positive ends, we get an $\mathcal{L}_\infty$ structure, consisting of l-to-1 operations for all $l \geq 1$ satisfying various compatibility conditions. We can conveniently package this $\mathcal{L}_\infty$ structure into one large chain complex $\bar{C}H_{\text{lin}}(X)$, the *bar complex*.

2.5. From spectral invariants to capacities

Getting back to the high-level viewpoint, we have a filtered chain complex $C(X)$ for each Liouville domain X, and filtration-preserving chain maps $\Xi : C(X) \to C(X')$ for any (exact) symplectic embedding $X' \overset{s}{\hookrightarrow} X$. Now, for any class α in the homology of $C(X)$, define $c_\alpha(X)$ to be the minimal action of any closed element of $C(X)$ which represents α. By a simple diagram chase, we have $c_{[\Xi](\alpha)}(X') \leq c_\alpha(X)$, where $[\Xi]$ denotes the homology-level map induced by Ξ.

At first glance, this construction appears to give a new family of symplectic capacities indexed by homology classes of $C(X)$. However, there is still one issue, which is that we need a canonical way to reference these homology classes. Indeed, in principle, the homology level map $[\Xi]$ might be quite nontrivial, so how do we know when two numbers $c_\alpha(X)$ and $c_\beta(X')$ can be compared to each other?

This is where the tangency constraints come in. The claim is that by counting possibly disconnected curves in $\widehat{X}$ with each component u_i satisfying a $\lessdot\mathcal{T}^{k_i-1}p\gtrdot$ constraint for some $k_i \in \mathbb{Z}_{>0}$, we get a chain map

$$\epsilon_X\lessdot\mathcal{T}^\bullet\gtrdot : C(X) \to \overline{S}\mathbb{Q}[t].$$

For example, a term $t^3 \odot t^2 \odot t^5$ in $\overline{S}\mathbb{Q}[t]$ corresponds to counting curves with three components which satisfy constraints $\lessdot\mathcal{T}^3p\gtrdot$, $\lessdot\mathcal{T}^2p\gtrdot$, and $\lessdot\mathcal{T}^5p\gtrdot$, respectively. Moreover, these maps are natural in the sense that the composition $\epsilon_{X'}\lessdot\mathcal{T}^\bullet\gtrdot \circ \Xi$ agrees with $\epsilon_X\lessdot\mathcal{T}^\bullet\gtrdot$ up to filtered chain homotopy.

Now, for any $\mathfrak{b} \in \overline{S}\mathbb{Q}[t]$, we define the capacity $\mathfrak{g}_\mathfrak{b}(X) \in \mathbb{R}_{>0}$ by

$$\mathfrak{g}_\mathfrak{b}(X) := \inf\{c_\alpha(X) : \; [\epsilon_X\lessdot\mathcal{T}^\bullet\gtrdot](\alpha) = \mathfrak{b}\}.$$

This defines a symplectomorphism invariant which scales like symplectic area, and for any symplectic embedding $X' \overset{s}{\hookrightarrow} X$, we have $\mathfrak{g}_\mathfrak{b}(X') \leq \mathfrak{g}_\mathfrak{b}(X)$. In the case that X is Liouville deformation equivalent to a ball, one can show that $\epsilon_X\lessdot\mathcal{T}^\bullet\gtrdot$ is actually a chain homotopy equivalence, so every spectral invariant of $C(X)$ corresponds to some choice of $\mathfrak{b}$.

[5]More precisely, we only allow "good" Reeb orbits, and we count cylinders which are additionally "anchored" in X.

Finally, to define the simplified capacities $\mathfrak{g}_k$, let $\pi_1 : \overline{S}\mathbb{Q}[t] \to \mathbb{Q}[t]$ denote the projection to tensors of length 1 (e.g., $t^2 + t^3 \odot t^2 \odot t^5$ maps to t^2). We define

$$\mathfrak{g}_k(X) := \inf_{\mathfrak{b}\colon \pi_1(\mathfrak{b})=t^{k-1}} \mathfrak{g}_{\mathfrak{b}}(X).$$

In essence, this means we look for the collection of Reeb orbits in ∂X of minimal action which is closed with respect to the differential of $C(X)$, and which bounds a connected rational curve in $\widehat{X}$ satisfying a $\lessdot\mathcal{T}^{k-1}p\gtrdot$ constraint (but disregarding any disconnected curves bounded by the same collection).

2.6. The case of ellipsoids

To get some intuition for $\mathfrak{g}_{\mathfrak{b}}(X)$, we note that when X is an irrational ellipsoid $E(a_1, \ldots, a_n)$, the differential on $C(X)$ vanishes for degree parity reasons. This means that $C(X)$ already agrees with its homology, and the map

$$\epsilon_X\lessdot\mathcal{T}^\bullet\gtrdot : C(X) \to \overline{S}\mathbb{Q}[t]$$

is in fact an isomorphism. Then, $\mathfrak{g}_{\mathfrak{b}}(X)$ is simply the action of the unique element $\left(\epsilon_X\lessdot\mathcal{T}^\bullet\gtrdot\right)^{-1}(\mathfrak{b}) \in C(X)$ which corresponds to $\mathfrak{b}$. However, recall that the map $\epsilon_X\lessdot\mathcal{T}^\bullet\gtrdot$ is defined by counting curves in $E(a_1, \ldots, a_n)$ satisfying local tangency constraints, so it could be quite nontrivial even in the case $n = 2$. Indeed, in the very special case of the nearly round ball $E(1, 1 + \epsilon)$, a closely related problem is to count rational curves in $\mathbb{CP}^2$ satisfying local tangency constraints, which was recently solved in [24]. For other ellipsoids, including those in higher dimensions, and for more general Liouville domains, computing $\mathfrak{g}_{\mathfrak{b}}$ seems to involve some very interesting and challenging enumerative problems.

We discuss the computation of the capacities $\mathfrak{g}_k$ for four-dimensional ellipsoids in Sect. 3.1.2, based on the forthcoming work [25]. As for the larger family of capacities $\mathfrak{g}_{\mathfrak{b}}$, a general recursive algorithm for their computation is given in [29], and this will be utilized in the proof of Proposition 1.7.

3. Optimal embeddings

3.1. The main theorems

We now prove our main results. To prove Theorem 1.1, we need a new construction and new obstructions. These two parts of our argument are logically independent of each other and can be done in either order. To prove Theorem 1.3, we can use an existing construction and so we just need the obstructions.

3.1.1. The construction. We begin with the construction.

Proposition 3.1. *For all $a > 1$ and $S > 0$, let $\frac{a}{a+1} \leq \mu \leq \frac{a}{2}$ and $\lambda = 1 - \frac{\mu}{a}$. There exists a symplectic embedding of $E(a, 1, S)$ into an arbitrary neighborhood of*

$$\{(z_1, z_2) \mid \pi|z_1|^2 \leq \lambda + \mu, \pi|z_2|^2 \leq f(\pi|z_1|^2)\} \times \mathbb{C},$$

where

$$f(t) = \begin{cases} 2\lambda - t/2 & \text{when} \quad 0 \le t \le 2\mu\frac{2\lambda-1}{\lambda+\mu-1}; \\ 1 - \frac{(1-\lambda)(t-2\lambda+1)}{1-\lambda+\mu} & \text{when} \quad 2\mu\frac{2\lambda-1}{\lambda+\mu-1} \le t \le \lambda+\mu. \end{cases}$$

Remark 3.2. Using the work of Pelayo-Vũ Ngọc [27, Theorem 4.4], we can extend to $S = \infty$ and embed the interior of the ellipsoid into the domain itself, rather than into a neighborhood.

We defer the proof for a moment, first stating some key corollaries we will need.

Corollary 3.3. *For any $N \ge 1$ and $a \ge 1$, $1 \le b \le 2$, there exists a symplectic embedding*

$$\mathrm{int}E(a,1) \times \mathbb{C}^N \overset{s}{\hookrightarrow} \frac{a(b+2)}{(a+1)b} \cdot \left(E(b,1) \times \mathbb{C}^N\right).$$

Here, "int" denotes the interior.

Proof of Corollary 3.3. It clearly suffices to prove this when $N = 1$. In Proposition 3.1, set $\mu = \frac{a}{a+1}$ so $\lambda = 1 - \frac{\mu}{a} = \mu$. In this case, $f(t) = 2\lambda - t/2$ for all $0 \le t \le 2\lambda = \lambda + \mu$ and we see that the domain $\{(z_1, z_2) \mid \pi|z_1|^2 \le \lambda+\mu, \pi|z_2|^2 \le f(\pi|z_1|^2)\}$ is simply $P(2\lambda, 2\lambda) \cap E(4\lambda, 2\lambda)$. This sits inside $E(cb, c)$ when $c \ge \frac{a(b+2)}{(a+1)b}$.

This deals with the case when $a > 1$. When $a = 1$, we still have an embedding into an arbitrarily small neighborhood, and so can still apply [27] for the precise result. □

Corollary 3.4. *Let $b \in \mathbb{R}_{\ge 2}$. Then, for any $N \ge 1$ and $a \ge b - 1$, there exists a symplectic embedding*

$$\mathrm{int}E(a,1) \times \mathbb{C}^N \overset{s}{\hookrightarrow} \frac{2a}{a+b-1} \cdot \left(E(b,1) \times \mathbb{C}^N\right).$$

Proof of Corollary 3.4. Note that when $a > 1$, we have $\frac{1-\lambda}{1-\lambda+\mu} < \frac{1}{2}$, and so the graph of $f(t)$ is convex. Hence, $f(t)$ is bounded above by the linear function between $(0, 2\lambda)$ and $(\lambda+\mu, \lambda)$ and our domain is a subset of $P(\lambda+\mu, 2\lambda) \cap E(2(\lambda+\mu), 2\lambda)$.

In the context of Proposition 3.1, set $\mu = \frac{a(b-1)}{a+b-1}$. We note that $\frac{a}{a+1} \le \mu \le \frac{a}{2}$ exactly when $2 \le b \le a+1$. Then, $\lambda = \frac{a}{a+b-1}$ and we find a symplectic embedding

$$E(a,1) \times \mathbb{C} \overset{s}{\hookrightarrow} \left(P\left(\frac{ab}{a+b-1}, \frac{2a}{a+b-1}\right) \cap E\left(\frac{2ab}{a+b-1}, \frac{2a}{a+b-1}\right)\right) \times \mathbb{C}$$
$$\subset \frac{2a}{a+b-1} E(b,1) \times \mathbb{C}.$$

□

We now give the promised proof of the proposition.

Proof of Proposition 3.1. Before the proof, we fix some notation.

Write $A \subset_\varepsilon B$ to mean that the set A lies in an ε neighborhood of B, or $z \in_\varepsilon B$ to mean that a point z lies ε close to B.

Let $\pi : \mathbb{C}^3 \to \mathbb{C}$ be the projection onto the z_1 plane.

In the z_1 plane, we fix sets $W_0 = [0,1] \times [0,\mu]$ and $W_i = [2i, 2i+1] \times [0,\lambda]$ for $i \geq 1$.

Finally, $D(a)$ denotes the round closed disk in the plane centered at the origin of area a, and A_i are the subsets of the z_3 plane given by $A_1 = D(S+\varepsilon)$ and $A_i = D(i(S+\varepsilon)) \backslash D((i-1)(S+\varepsilon))$ for $i \geq 2$. □

Proof. The condition $\mu \geq \frac{a}{a+1}$ is equivalent to $\mu \geq 1 - \frac{\mu}{a} = \lambda$, and the condition $\mu \leq \frac{a}{2}$ is equivalent to $2\lambda \geq 1$. Both of these inequalities will be used in our construction.

We apply a slightly generalized version of Lemma 2.2 from [12]. This says that, given ε, there exists a large K and a symplectomorphism ϕ from $E(a,1,S)$ to a set F_K with the following properties. For $z \in \mathbb{C}$, we write $F_z = \pi^{-1}(z) \cap F_K$.

(1) $\pi(F_K) \subset_\varepsilon \bigcup_{i=1}^K ([2i-1, 2i] \times \{0\}) \bigcup_{i=0}^K W_i$;
(2) if $z = (u,v) \in_\varepsilon W_0$ then $F_z \subset_\varepsilon D(1 - \frac{u\mu}{a}) \times A_1$;
(3) if $z \in_\varepsilon [2i-1, 2i] \times \{0\}$ and i is odd, then $F_z \subset_\varepsilon D(\lambda) \times A_i$;
(4) if $z \in_\varepsilon [2i-1, 2i] \times \{0\}$ and i is even, then $F_z \subset_\varepsilon (D(2\lambda) \backslash D(\lambda)) \times A_i$;
(5) if $z = (2i+u, v) \in_\varepsilon W_i$ and i is odd, then $F_z \subset_\varepsilon D((1+u)\lambda) \times (A_i \cup A_{i+1})$;
(6) if $z = (2i+u, v) \in_\varepsilon W_i$ and $i \geq 2$ is even, then $F_z \subset_\varepsilon D((2-u)\lambda) \times (A_i \cup A_{i+1})$.

Apart from slight changes of notation, the modification from Lemma 2.2 consists in increasing the area of W_0 (the original lemma fixed $\mu = \lambda = \frac{x}{x+1}$) and a refined description of the fibers over W_0. The estimate in item (2) follows easily, because $\pi^{-1}(W_0)$ is the set $\{\pi|z_1|^2 \leq \mu\} \subset E(a,1,S)$ and restricted to this set ϕ takes the form $\phi(z_1,z_2,z_3) = (\psi(z_1), z_2, z_3)$ where we may assume for all $0 \leq u \leq 1$ that ψ maps points with $\pi|z_1|^2 \leq \mu u$ (outside of which the fiber lies in $\pi|z_2|^2 < 1 - \frac{u\mu}{x}$) to an ε neighborhood of the set $[0,u] \times [0,\mu]$. Then, if $\psi(z_1) = (u,v)$, we have $\pi|z_1|^2 \geq \mu u - \varepsilon$ and so $\pi|z_2|^2 \leq 1 - \frac{u\mu}{a} + \varepsilon$.

The next step is to follow Step 3 of the proof from [12, page880] and apply a symplectic immersion $\tau : \pi(F_K) \to \mathbb{C}$. This can be arranged to restrict to an embedding on each of the W_i and each of the intervals $[2i-1, 2i] \times \{0\}$, so that the W_i with i odd map into a neighborhood of $[-1,0] \times [0,\lambda]$, the W_i with i even map into $[0,1] \times [0,\mu]$, and the ε neighborhoods of the intervals $[2i-1, 2i] \times \{0\}$ map close to the origin, remaining disjoint from the image of the W_i. The condition on W_i with i even is possible, since $\lambda \leq \mu$.

Let ι_{23} be the identity map on the (z_2, z_3)-plane. Then, we note that $(\tau \times \iota_{23}) : F_K \to \mathbb{C}^3$ is an embedding. Indeed, the fibers of π over W_i and W_j intersect only if $|i-j| \leq 1$ [since otherwise by items (5) and (6) their z_3 coordinates lie in different A_k], and in particular are disjoint if i and j have the same parity. Also, the fibers over neighborhoods of different intervals $[2i-1, 2i] \times \{0\}$ are disjoint by items (3) and (4).

We refine the immersion τ slightly to also satisfy the following.

- if $z = (2i + u, v) \in W_i$ and i is odd, then $\tau(z) \in_\varepsilon [-1 + u, 0] \times [0, \lambda]$;
- if $z = (u, v) \in W_0$, then $\tau(z) \in_\varepsilon [0, u] \times [0, \mu]$
- if $z = (2i + u, v) \in W_i$ and $i \geq 2$ is even, then $\tau(z) \in_\varepsilon [0, \frac{u\lambda}{\mu}] \times [0, \mu]$.

The following describes the fibers of the image of $\tau \times \iota_{23}$.

Lemma 3.5. *Let (z_1, z_2, z_3) lie in the image of $\tau \times \iota_{23}$ and $z_1 = (u, v)$.*
If $-1 \leq u \leq 0$, then $F_z \subset_\varepsilon D((2 + u)\lambda) \times \mathbb{C}$;
if $0 \leq u \leq \frac{2\lambda-1}{\lambda+\mu-1}$, then $F_z \subset_\varepsilon D(2\lambda - u\mu) \times \mathbb{C}$;
if $\frac{2\lambda-1}{\lambda+\mu-1} \leq u \leq 1$, then $F_z \subset_\varepsilon D(1 - \frac{u\mu}{a}) \times \mathbb{C}$.

Proof. The description of the fibers when $u \leq 0$ follows directly from item (5) in the description of F_K and the properties of τ. Also, if $\frac{\lambda}{\mu} \leq u \leq 1$, then by our description of τ restricted to the W_i, we see that (u, v) is the image of a point in W_0, and so, the property follows from item (2). (Note that $\frac{\lambda}{\mu} \geq \frac{2\lambda-1}{\lambda+\mu-1}$ because $\lambda < 1$ and $\mu \geq \lambda$.)

If $0 < u \leq \frac{\lambda}{\mu}$, then either $(u, v) = \tau(u', v')$ where $(u', v') \in W_0$ and $u' \geq u$, or $(u, v) = \tau(2i + u', v')$ where $(2i + u', v') \in W_i$ for $i \geq 2$ even and $u' \geq \frac{u\mu}{\lambda}$. In the first case, by item (2), the z_2 coordinate of the fiber lies in $D(1 - \frac{u\mu}{a})$ and in the second case, by (6), the z_2 coordinate of the fiber lies in $D(2\lambda - u\mu)$. Thus, the lemma follows from the fact that $2\lambda - u\mu \geq 1 - \frac{u\mu}{a}$ exactly when $u \leq \frac{2\lambda-1}{\lambda+\mu-1}$ (using the assumption that $2\lambda \geq 1$). $\square$

Finally, we apply the map $\sigma \times \iota_{23}$, where σ is an embedding of a neighborhood of $([-1, 0] \times [0, \lambda]) \cup ([0, 1] \times [0, \mu])$ in the z_1 plane to a neighborhood of the disk $D(\lambda + \mu)$. We can choose σ to satisfy the following.

- if $u \in [-\frac{\mu}{\lambda}t, t]$ and $0 \leq t \leq \frac{2\lambda-1}{\lambda+\mu-1}$, then $\sigma(u, v) \in_\varepsilon D(2t\mu)$ for all v;
- if $u \in [-\frac{2\lambda-1+(1-\lambda)t}{\lambda}, t]$ and $\frac{2\lambda-1}{\lambda+\mu-1} \leq t \leq 1$, then $\sigma(u, v) \in_\varepsilon D((2\lambda - 1) + (1 - \lambda + \mu)t)$ for all v.

Such a map σ exists, because the intersection of $([-1, 0] \times [0, \lambda]) \cup ([0, 1] \times [0, \mu])$, the image of τ, with $\{u \in [-\frac{\mu}{\lambda}t, t]\}$ has area $2\mu t$ and the intersection of the image of τ with $\{u \in [-\frac{2\lambda-1+(1-\lambda)t}{\lambda}, t]\}$ has area $(2\lambda-1)+(1-\lambda+\mu)t$. When $t = \frac{2\lambda-1}{\lambda+\mu-1}$, we have that $\frac{\mu}{\lambda}t = \frac{2\lambda-1+(1-\lambda)t}{\lambda}$, and so, we are imposing a condition on the image of all (u, v).

Claim. The image of $\sigma \times \iota_{23}$ lies in an ε neighborhood of $\{(z_1, z_2) \mid \pi|z_1|^2 \leq \lambda + \mu, \pi|z_2|^2 \leq f(\pi|z_1|^2)\} \times \mathbb{C}$, concluding the proof.

Proof of the claim. We check the fibers of π over points $w \in D(\lambda + \mu)$. First, if w is in the image of a point in one of the segments $[2i - 1, 2i] \times \{0\}$ then w is close to 0 and the z_2 coordinate of the fiber lies in $D(2\lambda)$.

Next, suppose that $\pi|w|^2 = s+\varepsilon$ where $s \leq 2\mu\frac{2\lambda-1}{\lambda+\mu-1}$. Then, $w = \sigma(u, v)$ where either $u > \frac{s}{2\mu}$ or $u < -\frac{s}{2\lambda}$ (since by our conditions on σ points with $u \in [-\frac{s}{2\lambda}, \frac{s}{2\mu}]$ are mapped into $D(s)$). By Lemma 3.5, in the first case, the z_2 coordinate of the fiber lies ε close to $D(2\lambda - \frac{s}{2})$, and in the second case, the z_2 coordinate of the fiber also lies in an ε neighborhood of $D((2 - \frac{s}{2\lambda})\lambda)$. Hence, $\pi|z_2|^2 \leq 2\lambda - \pi|z_1|^2/2$.

Finally, suppose that $\pi|w|^2 = s+\varepsilon$ where $2\mu\frac{2\lambda-1}{\lambda+\mu-1} \le s \le \lambda+\mu$. Then, we see that $w = \sigma(u,v)$ where either $u > \frac{s-(2\lambda-1)}{1-\lambda+\mu}$ or $u < -\frac{(2\lambda-1)\mu+(1-\lambda)s}{\lambda(1-\lambda+\mu)}$. This again follows from our conditions on σ. Indeed, if

$$u \in \left[-\frac{(2\lambda-1)\mu+(1-\lambda)s}{\lambda(1-\lambda+\mu)}, \frac{s-(2\lambda-1)}{1-\lambda+\mu}\right],$$

then, rewriting, $u \in [-\frac{2\lambda-1+(1-\lambda)t}{\lambda}, t]$ with $t = \frac{s-(2\lambda-1)}{1-\lambda+\mu}$. The bounds on s imply that $\frac{2\lambda-1}{\lambda+\mu-1} \le t \le 1$ and so by the second bullet point in our description of σ points with u in this range are mapped into $D((2\lambda-1)+(1-\lambda+\mu)t) = D(s)$.

Concluding by Lemma 3.5, if $u > \frac{s-(2\lambda-1)}{1-\lambda+\mu}$, then the z_2 coordinate of the fiber lies ε close to $D(1-\frac{s-(2\lambda-1)}{1-\lambda+\mu}\frac{\mu}{a}) = D(1-\frac{(1-\lambda)(s-2\lambda+1)}{1-\lambda+\mu})$, recalling that $\lambda = 1-\frac{\mu}{a}$. If $u < -\frac{s}{2\lambda}$, then the z_2 coordinate of the fiber lies ε close to $D(2\lambda - \frac{(2\lambda-1)\mu+(1-\lambda)s}{1-\lambda+\mu})$ which we check is also $D(1-\frac{(1-\lambda)(s-2\lambda+1)}{1-\lambda+\mu})$. Hence, $\pi|z_2|^2 \le 1-\frac{(1-\lambda)(\pi|z_1|^2-2\lambda+1)}{1-\lambda+\mu}+\varepsilon$. □

With the claim proven, we have completed the proof of the proposition. □

3.1.2. Some obstructions. We now turn our attention to the obstructive side. Notably, this will be quite short, because we can cite work on these higher capacities that has previously been done or is forthcoming. Namely, here, we only recall the following computations for the capacities of ellipsoids and polydisks from [28, §6.3]:

$$\mathfrak{g}_k(P(1,a)) = \min(k, a+\lceil \tfrac{k-1}{2}\rceil) \qquad \text{for } a \ge 1,\ k \ge 1 \text{ odd} \tag{6}$$
$$\mathfrak{g}_k(E(1,a)) = k \qquad \text{for } a \ge 1,\ 1 \le k \le a. \tag{7}$$

It seems plausible that the computation for $P(1,a)$ is also valid for k even. This would follow if we knew that the capacities $\mathfrak{g}_k$ are nondecreasing with k, although this is not yet clear.

We will also need the following more general expected formula for ellipsoids, which will be proved in [25]. For $1 \le a \le 3/2$, we have

$$\mathfrak{g}_k(E(1,a)) = \begin{cases} 1+ia & \text{for } k = 1+3i \text{ with } i \ge 0 \\ a+ia & \text{for } k = 2+3i \text{ with } i \ge 0 \\ 2+ia & \text{for } k = 3+3i \text{ with } i \ge 0. \end{cases} \tag{8}$$

For $a > 3/2$, we have

$$\mathfrak{g}_k(E(1,a)) = \begin{cases} k & \text{for } 1 \le k \le \lfloor a \rfloor \\ a+i & \text{for } k = \lceil a\rceil + 2i \text{ with } i \ge 0 \\ \lceil a\rceil + i & \text{for } k = \lceil a\rceil + 2i + 1 \text{ with } i \ge 0. \end{cases} \tag{9}$$

3.1.3. The proofs. We now give the promised proofs.

Proof of Theorem 1.1. Let a, b, and N be as in the statement of the theorem. Then, by Corollary 3.4, we have

$$c^N_{b,ell}(a) \leq \frac{2a}{a+b-1}.$$

To prove the opposite inequality, we use the higher capacities $\mathfrak{g}_k$. That is, take $k = a$. Then, by (7) and (9), we have

$$\mathfrak{g}_k(E(1,a)) = a, \quad \mathfrak{g}_k(E(1,b)) = \frac{a+b-1}{2}.$$

Hence, by the scaling, monotonicity, and stabilization properties of the $\mathfrak{g}_k$ in Theorem 2.1, we have

$$c^N_{b,ell}(a) \geq \frac{2a}{a+b-1},$$

hence the theorem. □

Remark 3.6. Note that in the above proof, we only need the inequality $\mathfrak{g}_a(E(1,b)) \leq \frac{a+b-1}{2}$, and in the case that b is even (and hence, $a \geq b+1$ is odd), this can be deduced directly from (6). Indeed, by (3), there is an embedding $E(1,b) \overset{s}{\hookrightarrow} P(1,b/2)$, whence we have

$$\mathfrak{g}_a(E(1,b)) \leq \mathfrak{g}_a(P(1,b/2)) = b/2 + \lceil (a-1)/2 \rceil = \frac{a+b-1}{2}.$$

Proof of Theorem 1.3. The proof is similar to the previous one. Let a, b, and N be as in the statement of the theorem.

The bound

$$c^N_{b,poly}(a) \leq \frac{2a}{a+2b-1}$$

follows from the existence of a variant of the embedding from above, which was previously shown to exist in [6, Lem.1.3].

To show that no better embedding exists, we use the above capacities. Namely, let $k = a$. Then, by (6) and (7), we have

$$\mathfrak{g}_k(E(1,a)) = a, \qquad \mathfrak{g}_k(P(1,b)) = b + \frac{a-1}{2}.$$

The theorem now follows by the same argument as above. □

Example 1. It is interesting to compare the above methods with the case $b = 1$. For this, we recall for the convenience of the reader an argument from [28, §1.4]. There, a variant of the embedding used in the previous theorems, constructed in [12], gives

$$c^N_{1,ell}(a) \leq \frac{3a}{a+1}.$$

On the other hand, if a is an integer congruent to two, modulo three, then taking $k = a$ as above yields

$$\mathfrak{g}_k(E(1,a) \times \mathbb{C}^n) = a, \qquad \mathfrak{g}_k(E(1,1) \times \mathbb{C}^n) = \frac{1+a}{3}.$$

Hence, combining these inequalities, we get that for a congruent to two modulo three

$$c^N_{1,ell}(a) = \frac{3a}{a+1}.$$

This recovers the result of McDuff [21, Thm. 1.1].

3.2. The rescaled embedding function

We now provide the proofs of the promised corollaries regarding the conjecture of the second named author, Frenkel, and Schlenk.

Proof of Corollary 1.4. We will first prove the statement about $c^N_{b,poly}$, after which the result about $c^N_{b,ell}$ will follow easily.

The function $c^N_{b,poly}(a)$ is nonincreasing in N. We want to show that it is in fact constant in N for a in the intervals given by the theorem. The computation of $c^0_{b,poly}(a)$ from [6], together with Theorem 1.3 from above, shows that it does not depend on N for the exterior (middle) corner of each linear step.

Now, note that if an embedding

$$E(1,a) \times \mathbb{C}^n \overset{s}{\hookrightarrow} \lambda P(1,b) \times \mathbb{C}^n$$

exists, then for any $a' > a$, by scaling, there is an embedding

$$E(1,a') \times \mathbb{C}^n \overset{s}{\hookrightarrow} \frac{a'}{a}\lambda P(1,b) \times \mathbb{C}^n.$$

Thus, $c^N_{b,poly}(a') \leq \frac{a'}{a} c^N_{b,poly}(a)$. Therefore, given $y_0 = c^N(a)$, the graph of $c^N(a')$ for $a' > a$ cannot lie above the line through (a, y_0) and the origin. For future reference, we call this the **subscaling property**. We can now prove the corollary.

Consider any linear step for $c^0_{b,poly}(a)$. Recall that this consists of a linear part, then an exterior corner, and then a horizontal part. Consider the linear part. We want to show that this stabilizes. We know that $c^N_{b,poly}(a) \leq c^0_{b,poly}(a)$. If there were any a value for which strict inequality held, then by the linearity property above, at the exterior corner a_0 of the step, we would have $c^N_{b,poly}(a_0) < c^0_{b,poly}(a_0)$. However, above we saw in Theorem 1.3 that the exterior corner is stable. Hence, the whole linear part must stabilize. As for the horizontal part, we know that we must have $c^N_{b,poly} \leq c^0_{b,poly}$, but on the other hand the function $c^N_{b,poly}$ is nondecreasing, and so must be constant here. Thus, the whole step stabilizes, so all the linear steps do.

In view of Theorem 1.1, the exact same argument implies the result about $c^N_{2b,ell}$, since for $N = 0$, there is an equivalence of embeddings (3). □

Proof of Corollary 1.5. Corollary 1.4 shows that, after the initial part of the graph, where $c^N_{b,poly}(a) = 1$, the graph has $\lceil \sqrt{2b} \rceil + 1$ linear steps that are all stable. The length of these steps is given by the formula $\ell_b(k)$ from [6, p.6]. In particular, as explained there, the length of the k^{th} step converges to 2 as b tends to infinity. Since the steps are centered at the odd numbers, increase in number without bound as b increases, and our rescaled function is centered, so that the initial part of the graph with height one, that is, the part determined by Gromov's non-squeezing theorem, does not appear, the result follows. □

3.3. The first step

We now prove Proposition 1.6.

Proof of Proposition 1.6. The key is the following lemma.

Lemma 3.7. *Let a_0 be the smallest odd integer that is no less than $2b-1$. There is a symplectic embedding*

$$\operatorname{int}\left(E\left(1,\frac{a_0-1}{2}+b\right)\right)\overset{s}{\hookrightarrow}P(1,b). \tag{10}$$

Proof. We first explain why it suffices to prove the lemma for b rational. Given an irrational b, we can choose rational numbers b_n converging to b from below. Then, if the lemma is true for each b_n and the b_n are sufficiently close to b, composing with the inclusion $P(1,b_n)\subset P(1,b)$ gives embeddings $\operatorname{int}\left(E\left(1,\frac{a_0-1}{2}+b_n\right)\right)\overset{s}{\hookrightarrow}P(1,b)$, hence the desired embedding (10) by [4, Cor. 1.6].

We thus henceforth assume that b is rational. Then, by for example [4][Thm. 2.1], it is equivalent to find an embedding

$$\operatorname{int}(E(1,b))\cup\operatorname{int}\left(E\left(1,\frac{a_0-1}{2}\right)\right)\overset{s}{\hookrightarrow}P(1,b). \tag{11}$$

Indeed, the argument for [4, Thm. 2.1] implies that both (10) and (11) are equivalent to ball packing problems of the $P(1,b)$, where in the first case, the size of the balls is given by the weight sequence defined in [4, §2] for $(a_0-1)/2+b$, and in the second case, the size of the balls is given by the union of the weight sequence for b and for $(a_0-1)/2$. Since $(a_0-1)/2$ is an integer, the first $(a_0-1)/2$ of the weights for $(a_0-1)/2+b$ will be 1, so (10) and (11) are equivalent to the same ball packing problem.

We know that $a_0\leq 2b+1$, and hence

$$\frac{a_0-1}{2}\leq b. \tag{12}$$

We can therefore find an embedding as in (11) as follows. We think the moment image of $P(1,b)$ as a union of two triangles, joined along the diagonal that does not contain the origin. The triangle with legs on the axes contains an $E(1,b)$ factor by inclusion. As for the other triangle, it is affine equivalent to the first, via multiplication by $-I_2$, where I_2 is the two-by-two identity matrix. Hence, by the Traynor trick, see for example [31] and [2, Lem. 1.8], it also contains a copy of an $\operatorname{int}(E(1,b))$, disjoint from the interior of the first $E(1,b)$. Now, by (12), this latter $\operatorname{int}(E(1,b))$ contains a copy of $\operatorname{int}(E(1,(a_0-1)/2))$. □

We can now prove the proposition. We first prove the second bullet point. By Lemma 3.7, we know that $c^N_{b,poly}\leq 1$, for a in the given range. However, by Gromov's non-squeezing theorem, we also know that $c^N_{b,poly}\geq 1$, for a in this range. As for the rest of the second bullet point, this follows from the subscaling property of $c^N_{b,poly}$, as in the proof of Corollary 1.4 above, given the lower bound on $c^N_{b,poly}(a_0)$ coming from Theorem 1.3.

We now prove the first bullet point. The result for $1 \leq a \leq b$ follows, because inclusion gives an embedding for a in this range, which is optimal by Gromov's non-squeezing theorem. Similarly, for $b \leq a \leq \lfloor b \rfloor + 1$, scaling gives an embedding as in the subscaling property, which is optimal by the $(\lfloor b \rfloor + 1)^{st}$ Ekeland–Hofer capacity; see, e.g., [3][§2.3.1, §4.1.1] for the relevant formula. $\square$

3.4. The other parity

The proof of the remaining proposition, Proposition 1.7, requires the $\mathfrak{g}_b$ and computer assistance as well. It turns out that the simplified capacities $\mathfrak{g}_k$ do not suffice in these cases. For example, for $E(1,6) \times \mathbb{C}^N \overset{s}{\hookrightarrow} \lambda \cdot P(1,1) \times \mathbb{C}^N$, one can check that the simplified capacities give only $\lambda \geq 5/3$, whereas we have in fact $c^N_{1,poly}(6) = 12/7$ for $N \in \mathbb{Z}_{\geq 1}$.

On the other hand, we have the more general capacities $\mathfrak{g}_{\mathfrak{b}}$, which could in principle give sharp obstructions for all $a \in \mathbb{R}_{\geq 1}$ and $b \in \mathbb{Z}_{\geq 1}$ in (2) and (4). This is related to the discussion at the end of [28, §6.3], where it is observed that the simplified capacities $\mathfrak{g}_k$ do not generally give sharp obstructions for $E(1,a) \times \mathbb{C}^N \overset{s}{\hookrightarrow} \lambda \cdot E(1,1) \times \mathbb{C}^N$, but the capacities $\mathfrak{g}_{\mathfrak{b}}$ necessarily give sharp obstructions at least for $a \leq \tau^4$. Moreover, the formalism from [29] gives an explicit recursive algorithm to compute the capacities $\mathfrak{g}_{\mathfrak{b}}$ for all convex toric domains, although, unfortunately, it appears to be somewhat difficult to compute with "by hand".

Proof of Proposition 1.7. We begin with the computation of $c^N_{1,poly}(a)$ for $a = 6, 8, \dots, 100$. By [12], we have the upper bound $c^N_{1,poly}(a) \leq \frac{2a}{a+1}$, so it suffices to establish the lower bound $c^N_{1,poly}(a) \geq \frac{2a}{a+1}$. Suppose that we have a symplectic embedding $E(1,a) \times \mathbb{C}^N \overset{s}{\hookrightarrow} \lambda \cdot P(1,1) \times \mathbb{C}^N$.

Following the notation and exposition of [29], the idea is as follows. By [29, Cor. 1.2.3], there is a filtered $\mathcal{L}_\infty$ homomorphism $Q : V_{P(\lambda,\lambda)} \to V_{E(1,a)}$ which is unfiltered $\mathcal{L}_\infty$ homotopic to the identity. Here, V is an explicit DGLA with generators $\alpha_{i,j}$ for $i,j \in \mathbb{Z}_{\geq 1}$ and $\beta_{i,j}$ for $i,j \in \mathbb{Z}_{\geq 0}$ not both zero. The filtered DGLA $V_{P(\lambda,\lambda)}$ is just V as an unfiltered DGLA, and its filtration is specified by

$$\mathcal{A}_{P(\lambda,\lambda)}(\alpha_{i,j}) = \mathcal{A}_{P(\lambda,\lambda)}(\beta_{i,j}) = \lambda i + \lambda j.$$

Similarly, the filtered DGLA $V_{E(1,a)}$ is just V as an unfiltered DGLA, with filtration specified by

$$\mathcal{A}_{E(1,a)}(\alpha_{i,j}) = \mathcal{A}_{E(1,a)}(\beta_{i,j}) = \max(i, aj).$$

Recall that an $\mathcal{L}_\infty$ homomorphism $Q : V_{P(\lambda,\lambda)} \to V_{E(1,a)}$ consists of a sequence of maps $Q^l : \odot^l V_{P(\lambda,\lambda)} \to V_{E(1,a)}$ for $l = 1, 2, 3, \dots$, and these must satisfy an infinite sequence of certain quadratic relations.

Any element of the form $\beta_{i_1,j_1} \odot \cdots \odot \beta_{i_k,j_k}$ defines a cycle in the bar complex $\bar{V}_{P(\lambda,\lambda)}$. In particular, $\widehat{Q}(\beta_{i_1,j_1} \odot \cdots \odot \beta_{i_k,j_k})$ must be homologous to $\beta_{i_1,j_1} \odot \cdots \odot \beta_{i_k,j_k}$ in $\bar{V}_{E(1,a)}$. Moreover, there is a filtered $\mathcal{L}_\infty$ homomorphism $\Phi_{1,a} : V_{E(1,a)} \to V^{\mathrm{can}}_{E(1,a)}$, where $V^{\mathrm{can}}_{E(1,a)}$ denotes the homology of $V_{E(1,a)}$ (viewed as a filtered $\mathcal{L}_\infty$ algebra with trivial $\mathcal{L}_\infty$ operations), and hence, $(\widehat{\Phi}_{1,a} \circ \widehat{Q})(\beta_{i_1,j_1} \odot \cdots \odot \beta_{i_k,j_k})$ is homologous to $\widehat{\Phi}_{1,a}(\beta_{i_1,j_1} \odot \cdots \odot \beta_{i_k,j_k})$ in $\bar{V}^{\mathrm{can}}_{E(1,a)}$.

Now, suppose that we have $a = p/q$ with $p + q = 2d$ for some $p, q, d \in \mathbb{Z}_{\geq 1}$. Consider some $d_1, d_2 \in \mathbb{Z}_{\geq 0}$ satisfying $d_1 + d_2 = d$, and suppose that we have

$$\Phi^d_{1,a}(\odot^{d_1}\beta_{1,0} \odot \odot^{d_2}\beta_{0,1}) \neq 0. \tag{13}$$

Then, we claim that we have $\lambda \geq \frac{2a}{a+1}$, which gives the desired lower bound. Indeed, for a general input of the form $\beta_{i_1,j_1} \odot \cdots \odot \beta_{i_k,j_k}$, it follows by degree considerations that $\Phi^k_{1,a}(\beta_{i_1,j_1} \odot \cdots \odot \beta_{i_k,j_k})$ is either trivial, or else it is the unique element up to scaling in $V^{\mathrm{can}}_{E(1,a)}$ of its given degree. In the latter case, its action is given by the lth Ekeland–Hofer capacity of $E(1,a)$, i.e., $c^{\mathrm{EH}}_l(E(1,a))$, for $l = \sum_{m=1}^k (i_m + j_m) + k - 1$. Also, the action of the input is given by

$$\mathcal{A}_{P(\lambda,\lambda)}(\beta_{i_1,j_1} \odot \cdots \odot \beta_{i_k,j_k}) = \sum_{m=1}^{k} \mathcal{A}_{P(\lambda,\lambda)}(\beta_{i_m,j_m}) = \sum_{m=1}^{k} (\lambda i_m + \lambda j_m).$$

Specializing to the case of input $\odot^{d_1}\beta_{1,0} \odot \odot^{d_2}\beta_{0,1}$ and $l = 2d - 1$, using $a = p/q$ and $p + q = 2d$, it is straightforward to check that we have $c^{\mathrm{EH}}_l(E(1,a)) = p$. Since $\widehat{\Phi}_{1,a} \circ \widehat{Q}$ is filtration-preserving and $\Phi^d(\odot^{d_1}\beta_{1,0} \odot \odot^{d_2}\beta_{0,1})$ is a summand of the image of $[\odot^{d_1}\beta_{1,0} \odot \odot^{d_2}\beta_{0,1}]$ under $[\widehat{\Phi}_{1,a} \circ \widehat{Q}]$, we must have $\lambda(d_1 + d_2) \geq p$, and hence

$$\lambda \geq \frac{p}{d} = \frac{2p}{p+q} = \frac{2a}{a+1},$$

as claimed.

Let us now specialize to the case that a is an even integer. Then, we have $a = p/q$ for $p = 2a$ and $q = 2$, and hence, $p + q = 2d$ for $d = a + 1$. By computer calculations, (13) holds for $d_1 = 3$ and $d_2 = d - d_1 = a - 2$ for $a = 6, \ldots, 100$. Geometrically, this corresponds to a nonvanishing count of rational curves in $\mathbb{CP}^1 \times \mathbb{CP}^1 \setminus \frac{1}{\lambda} \cdot E(1,a)$ of bidegree (d_1, d_2) with one negative puncture asymptotic to the $p = 2a$ fold cover of the short simple Reeb orbit. Curiously, the analogous counts for $d_1 = 1, 2$ vanish.

The computation of $c^N_{2,ell}(a)$ for $a = 6, 8, \ldots, 100$ is similar. In this case, we suppose that we have a symplectic embedding $E(1,a) \times \mathbb{C}^N \overset{s}{\hookrightarrow} \lambda \cdot E(1,2) \times \mathbb{C}^N$, and we take our input cycle to be of the form $\odot^3\beta_{2,1} \odot \odot^{d-3}\beta_{1,0}$, for $d = a - 2$. By computer calculation, we have

$$\Phi^d_{1,a}(\odot^3\beta_{2,1} \odot \odot^{d-3}\beta_{0,1}) \neq 0 \tag{14}$$

for $a = 6, 8, \ldots, 100$. The action of the output is that of the lth Ekeland–Hofer capacity of $E(1,a)$ for $l = 5 + 2d$, and we have $c^{\mathrm{EH}}_l(E(1,a)) = 2a$. Meanwhile, the action of the input is

$$\mathcal{A}_{E(1,2)}(\odot^3\beta_{2,1} \odot \odot^{d-3}\beta_{1,0}) = 6 + (d-3) = a + 1,$$

whence the lower bound $\lambda \geq \frac{2a}{a+1}$ readily follows. □

4. Discussion

We close by discussing some natural follow-up questions to our work.

4.1. Beyond the rescaled function

One can of course ask whether the function $c^N_{b,poly}(a)$ can in any sense be computed completely. As explained in [6, Lem.1.3], and mentioned previously here, a previous folding construction of the second named author gives the bound

$$c^N_{b,poly}(a) \leq \frac{2a}{a+2b-1}.$$

This bound cannot be optimal for all a. For example, as we have seen in this paper, there are sometimes four-dimensional embeddings beating this bound, and these can be stabilized by taking the product with the identity. For a sufficiently large with respect to b, though, in particular for

$$a \geq (\sqrt{2b}+1)^2, \tag{15}$$

the above folding bound beats the four-dimensional volume obstruction, and so must give a better construction than any stabilized four-dimensional one. The main question at the moment here is as follows.

Question 4.1. *Is it the case that either $c^N_{b,poly}(a) = c^0_{b,poly}(a)$, or*

$$c^N_{b,poly}(a) = \frac{2a}{a+2b-1}?$$

If this is true, it looks hard to prove. For example, if $a < (\sqrt{2b}+1)^2$, then the volume bound is strictly below the folding bound from above. On the other hand, for $b \in \mathbb{Z}_{\geq 2}$, it is known that there are entire intervals of the subset $a < (\sqrt{2b}+1)^2$ for which the volume bound is optimal for $c^0_{b,poly}$: for example, for $b = 2$, [6, Thm.1.1] states that there is an interval on which $c^0_{b,poly}$ is given by the volume starting at $a = 7.84$, but on the other hand by (15), the folding curve is above the volume curve up until $a = 9$. Finding the holomorphic curves needed to show that this volume bound stabilizes would be a completely new phenomenon.

The same question, but concerning $c^N_{b,ell}$ is also open and just as interesting.

4.2. The opposite parity

It is also natural to ask what happens for the stabilized embedding problem for ellipsoids, when the parity of the domain and target are the same. For example, one might hope that an analogue of our Proposition 1.7 holds in the case $b > 2$. If this is true, however, it is not so clear how to prove it: our preliminary computer search to generalize the method required to prove it has not turned up promising candidates. It would be very interesting to find a candidate of curves to solve this problem, or to find another embedding.

4.3. The region from $b = 1$ to $b = 2$

For $b \geq 2$, our Corollary 3.4 produces an embedding, such that

$$c^N_{b,ell}(a) \leq \frac{2a}{a+b-1}.$$

Meanwhile, for $1 \leq b \leq 2$, Corollary 3.3 shows

$$c^N_{b,ell}(a) \leq \frac{a(b+2)}{(a+1)b}. \tag{16}$$

It is interesting to ask when this bound is sharp, for instance whether there are sequences of a where this holds. We now list some facts, suggesting that the answer may not be straightforward.

Note that when $b = 1$, the bound gives

$$c^N_{1,ell}(a) \leq \frac{3a}{a+1},$$

which as mentioned above is sharp when $a \equiv 2$ modulo 3, [21]. There is another sequence starting at $a = 2$ where (16) is an equality. By work of the first and second named authors, [7], we have $c^N_{1,ell}(a) = c^0_{1,ell}(a)$ for all $1 \leq a \leq \tau^4$. This region of the graph is an infinite staircase, that is, piecewise linear with infinitely many singular points accumulating at τ^4, see [23]. Between these singular points the graph alternates between being constant and sitting on a line through the origin. One can check the corners of the stairs, the left endpoints of the constant intervals, lie on the folding graph $\frac{3a}{a+1}$.

When $b = 2$, our bound gives

$$c^N_{2,ell}(a) \leq \frac{2a}{a+1}.$$

The graph of $c^0_{2,ell}$ also begins with an infinite staircase, see [5, 11], and again, the tips of the stairs lie on the graph $\frac{2a}{a+1}$. It seems extremely likely that at such a, we have $c^N_{2,ell}(a) = c^0_{2,ell}(a)$ for all N, so the bound (16) is again sharp.

However, when $b = 3/2$, the situation is mysterious. Now, our bound gives

$$c^N_{3/2,ell}(a) \leq \frac{7}{3}\frac{a}{a+1}.$$

Here again, work of the first named author and Kleinman shows that $c^0_{3/2,ell}(a)$ has an infinite staircase [5], but now, the tips of the stairs lie on the graph $\frac{2a}{a+1}$. Moreover, the $\mathfrak{g}_k$ show that $c^N_{3/2,ell}(a) \geq \frac{2a}{a+1}$ at integer a. It is unclear whether an improved construction can show this lower bound is indeed sharp, or whether enhanced obstructions can be used to show that even though the folding graph (16) lies strictly above the infinite staircase, it is still asymptotically sharp.

4.4. A combinatorial rule?

While the functions $c^0_{b,ell}$ and $c^0_{b,poly}$ themselves are known to be quite complicated (see for example [23,32]), they are governed by simple to state combinatorial rules. For example, McDuff shows in [20] that $c^0_{b,ell}$ is completely determined by the combinatorics of the sequence $N(a,b)$, whose kth term is the $(k+1)^{st}$ smallest entry among the nonnegative integer linear combinations of a and b. It would be extremely interesting if the functions $c^N_{b,ell}$ and $c^N_{b,poly}$ are also governed by some kind of relatively simple to state combinatorial rule. It might be easier to find such a rule than to actually compute these functions explicitly.

Acknowledgements

We thank Felix Schlenk for his encouragement, and for helping the first and third named authors better understand constructions of embeddings between stabilized ellipsoids. We would also like to thank the referee for carefully reading our paper and for many useful comments.

Our paper is dedicated to Claude Viterbo on the occasion of his 60th birthday. We are immensely grateful to Claude for his visionary leadership of our field.

This research was completed, while the first named author was on a von Neumann fellowship at the Institute for Advanced Study; he thanks the Institute for their support. The first named author is partially supported by NSF grant DMS-1711976 and the second named author by Simons Foundation Grant no. 633715.

References

[1] Bao, E., Honda, K.: Semi-global Kuranishi charts and the definition of contact homology. (2015). arXiv:1512.00580

[2] Choi, K., Cristofaro-Gardiner, D., Frenkel, D., Hutchings, M., Ramos, V.: Symplectic embeddings into four-dimensional concave toric domains. J. Topol. **7**, 1054–1076 (2014)

[3] Cieliebak, K., Hofer, H., Latschev, J., Schlenk, F.: Quantitative symplectic geometry. Dyn. Ergod. Theory Geom. Dedic. Anatole Katok **54**, 1–44 (2007)

[4] Cristofaro-Gardiner, D.: Symplectic embeddings from concave toric domains into convex ones. J. Differ. Geom. **112**, 199–232 (2019)

[5] Cristofaro-Gardiner, D., Kleinman, A.: Ehrhart functions and symplectic embeddings of ellipsoids. J. Lond. Math. Soc. **101**, 1090–1111 (2020)

[6] Cristofaro-Gardiner, D., Frenkel, D., Schlenk, F.: Symplectic embeddings of four-dimensional ellipsoids into integral polydiscs. Algebraic Geom. Topol. **17**, 1189–1260 (2017)

[7] Cristofaro-Gardiner, D., Hind, R.: Symplectic embeddings of products. Comment. Math. Helv. **93**, 1–32 (2018)

[8] Cristofaro-Gardiner, D., Hind, R., McDuff, D.: The ghost stairs stabilize to sharp symplectic embedding obstructions. J. Topol. **11**, 309–378 (2018)

[9] Eliashberg, Y., Givental, A., Hofer, H.: Introduction to symplectic field theory. Visions in mathematics: GAFA 2000 Special volume, pp. 560–673 (2000)

[10] Fish, J.W., Hofer, H.: Lectures on polyfolds and symplectic field theory. (2018). arXiv:1808.07147

[11] Frenkel, D., Müller, D.: Symplectic embeddings of four-dimensional ellipsoids into cubes. J. Symplectic Geom. **13**, 765–847 (2015)

[12] Hind, R.: Some optimal embeddings of symplectic ellipsoids. J. Topol. **8**, 871–883 (2015)

[13] Hind, R.: Stabilized symplectic embeddings. In: Angella, D., Medori, C., Tomassini, A. (eds) Complex and Symplectic Geometry. Springer INdAM Series, vol 21. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-62914-8_7

[14] Hind, R., Kerman, E.: New obstructions to symplectic embeddings. Invent. Math. **196**, 383–452 (2014)

[15] Hind, R., Kerman, E.: Correction to: New obstructions to symplectic embeddings. Invent. Math. **214**, 1023–1029 (2018)

[16] Hofer, H., Wysocki, K., Zehnder, E.: Polyfold and Fredholm Theory. Springer, Berlin (2021)

[17] Hutchings, M.: Rational SFT using only q variables.https://floerhomology.wordpress.com/2013/04/23/rational-sft-using-only-q-variables (2013)

[18] Hutchings, M., Nelson, J.: Cylindrical contact homology for dynamically convex contact forms in three dimensions. J. Symplectic Geom. **14**, 983–1012 (2016)

[19] Ishikawa, S.: Construction of general symplectic field theory. (2018). arXiv:1807.09455

[20] McDuff, D.: The Hofer conjecture on embedding symplectic ellipsoids. J. Differ. Geom. **88**, 519–532 (2011)

[21] McDuff, D.: A remark on the stabilized symplectic embedding problem for ellipsoids. Eur. J. Math. **4**, 356–371 (2018)

[22] McDuff D., Salamon D.: Introduction to symplectic topology (third edition). In: Oxford Graduate Texts in Mathematics (2017)

[23] McDuff, D., Schlenk, F.: The embedding capacity of 4-dimensional symplectic ellipsoids. Ann. Math. **175**, 1191–1282 (2012)

[24] McDuff, D., Siegel, K.: Counting curves with local tangency constraints. J. Topol. (2021) **(to appear)**

[25] McDuff, D., Siegel, K.: Symplectic capacities, unperturbed curves, and convex toric domains. arXiv:2111.00515

[26] Pardon, J.: Contact homology and virtual fundamental cycles. J. Am. Math. Soc. **32**, 825–919 (2019)

[27] Pelayo, Á., Ngọc, S.V.: Hofer's question on intermediate symplectic capacities. Proc. Lond. Math. Soc. **110**, 787–804 (2015)

[28] Siegel, K.: Higher symplectic capacities. (2019). arXiv:1902.01490

[29] Siegel, K.: Computing higher symplectic capacities I. Int. Math. Res. Not. **(to appear)**

[30] Siegel, K.: Computing higher symplectic capacities II **(in preparation)**

[31] Traynor, L.: Symplectic packing constructions. J. Differ. Geom. **42**, 411–429 (1995)

[32] Usher, M.: Infinite staircases in the symplectic embedding problem for four-dimensional ellipsoids into polydisks. Algebraic Geom. Topol. **19**, 1935–2022 (2019)

Daniel Cristofaro-Gardiner
University of Maryland
College Park
MD
USA
e-mail: dcristof@umd.edu

Richard Hind
University of Notre Dame
Notre Dame
IN
USA
e-mail: hind.1@nd.edu

Kyler Siegel
University of Southern California
Los Angeles
CA
USA
e-mail: kyler.siegel@usc.edu

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:19
https://doi.org/10.1007/s11784-022-00962-9
Published online March 30, 2022

Journal of Fixed Point Theory and Applications

Check for updates

Closed geodesics on reversible Finsler 2-spheres

Guido De Philippis, Michele Marini, Marco Mazzucchelli and Stefan Suhr

Dedicated to Claude Viterbo on the occasion of his 60th birthday.

Abstract. We extend two celebrated theorems on closed geodesics of Riemannian 2-spheres to the larger class of reversible Finsler 2-spheres: Lusternik–Schnirelmann's theorem asserting the existence of three simple closed geodesics, and Bangert–Franks–Hingston's theorem asserting the existence of infinitely many closed geodesics. To prove the first theorem, we employ the generalization of Grayson's curve shortening flow developed by Angenent–Oaks.

Mathematics Subject Classification. 53C22, 58E10, 53C44.

Keywords. Closed geodesics, reversible Finsler metrics, curve shortening flow, Lusternik–Schnirelmann theory.

Contents

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

1. Introduction

Since the seminal work of Hadamard [21], Poincaré [40], Birkhoff [9], and Morse [36], it became evident that closed Riemannian manifolds of dimension at least 2 tend to have many closed geodesics (that is, periodic orbits of the geodesic flow). This evidence was supported by celebrated theorems of Gromoll–Meyer [19] and Vigué-Poirrier–Sullivan [43], which together assert that simply connected closed Riemannian manifolds with a non-monogenic rational cohomology ring always have infinitely many closed geodesics. This statement covers a large class of simply connected closed manifolds, leaving out those with the cohomology of a compact rank-one symmetric space: S^n, $\mathbb{CP}^n, \mathbb{HP}^n$, and $\mathrm{Ca}\mathbb{P}^2$. As of 2019, it is an open conjecture whether these spaces admit infinitely many closed geodesics for any choice of the Riemannian metric. The only known case is the one of S^2, for which the proof required a combination of spectacular work by Bangert [8], Franks [16], and Hingston [22] (either Franks' or Hingston's work, together with Bangert's one, provide the full result). The starting point for this work is another celebrated result due to Lusternik–Schnirelmann [30], asserting that every Riemannian 2-sphere possesses at least three simple closed geodesics (that is, closed geodesics that are embedded circle in the Riemannian manifold). For many decades Lusternik–Schnirelmann's theorem was considered controversial due to a gap in their construction of a pseudo-gradient flow for the length function of simple closed curves, that have been subsequently addressed by many authors [6,23,25,26,28,42]. Nowadays, the gap is considered filled, for instance thanks to the work of Grayson [20] on the curve shortening flow.

The closed geodesic problem can be studied on closed Finsler manifolds as well. A Finsler metric on a manifold M is a continuous function $F : \mathrm{T}M \to [0, \infty)$, smooth outside the zero section of $\mathrm{T}M$, positively homogeneous of degree 1 (i.e., $F(x, \lambda v) = \lambda F(x, v)$ for all $(x, v) \in \mathrm{T}M$ and $\lambda \geq 0$), and such that the restriction of its square F^2 to any fiber of $\mathrm{T}M$ has positive definite Hessian everywhere outside the origin. In the literature, a more general notion of Finsler metric is sometimes employed, but the one given here is the most appropriate for the study of geodesic flows. Many results, such as Gromoll–Meyer's one, remain valid essentially with the same proof in the Finsler category (see [31] and references therein). However, a striking example due to Katok [27], and further explored by Ziller [45], shows that Lusternik–Schnirelmann's and Bangert–Franks–Hingston's theorems fail: there exists a Finsler metric on S^2 having only two closed geodesics.

A Finsler metric $F : \mathrm{T}M \to [0, \infty)$ is called reversible when $F(x, -v) = F(x, v)$ for all $(x, v) \in \mathrm{T}M$. The Katok's Finsler metric does not satisfy this property. In the current paper, we show that all the above mentioned results valid for Riemannian 2-spheres remain valid for reversible Finsler 2-spheres.

1.1. The curve shortening semi-flow

In [39], Oaks provided a generalization of Grayson's curve shortening flow [20]. As remarked by Angenent [5], such generalization allows to provide a curve shortening flow on orientable reversible Finsler surfaces: a tool to shrink embedded circles without creating self-intersections. In this section, we state a theorem that summarizes all the properties of this flow (actually, a semi-flow) that we will need for the application to the closed geodesics problem.

Let M be a closed oriented surface, equipped with a reversible Finsler metric F. We denote by $S^1 := \mathbb{R}/\mathbb{Z}$ the 1-periodic circle, and by $\mathrm{Emb}(S^1, M)$ the space of smooth embedded loops $\gamma : S^1 \hookrightarrow M$, endowed with the C^∞ topology (that is, the topology whose basis is given by the open sets $\mathcal{U} \subset \mathrm{Emb}(S^1, M)$ of the C^k topology, for all $k \in \mathbb{N}$). We consider the Finsler length functional

$$L : \mathrm{Emb}(S^1, M) \to (0, \infty), \quad L(\gamma) = \int_0^1 F(\gamma(u), \dot\gamma(u))\, \mathrm{d}u. \tag{1.1}$$

The group of diffeomorphisms $\mathrm{Diff}(S^1)$ acts freely on $\mathrm{Emb}(S^1, M)$ by reparametrizations. Notice that, since the Finsler metric F is homogeneous of degree 1 and reversible, the length functional is invariant by the $\mathrm{Diff}(S^1)$-action, i.e., $L(\gamma) = L(\gamma \circ \theta)$ for all $\gamma \in \mathrm{Emb}(S^1, M)$ and $\theta \in \mathrm{Diff}(S^1)$.

We fix an auxiliary Riemannian metric g on M, and we will simply write $\|\cdot\|$ or $\|\cdot\|_g$ for its associated norm on tangent vectors. Since (M, g) is an orientable Riemannian surface, it admits a canonical complex structure $J \in \mathrm{End}(\mathrm{T}M)$, i.e., Jv is obtained by rotating $v \in \mathrm{T}_x M$ of a positive angle $\pi/2$ measured with g. The positive normal to $\gamma \in \mathrm{Emb}(S^1, M)$ is the vector field

$$N_\gamma(u) := \frac{1}{\|\dot\gamma(u)\|} J\dot\gamma(u),$$

where $\|\cdot\|$ is the Riemannian norm associated to g. We set

$$V_\gamma(u) := \frac{\big(\frac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma(u), \dot\gamma(u)) - F_x(\gamma(u), \dot\gamma(u))\big) N_\gamma(u)}{\|\dot\gamma(u)\|}. \tag{1.2}$$

In the expression of V_γ, the terms F_x and F_v denote the partial derivatives of F with respect to some local coordinates on M (or, more precisely, local coordinates on $\mathrm{T}M$ induced by local coordinates on M). However, the covector

$$\tfrac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma(u), \dot\gamma(u)) - F_x(\gamma(u), \dot\gamma(u)) \in \mathrm{T}^*_{\gamma(u)} M \tag{1.3}$$

is independent of the choice of local coordinates, and vanishes identically if and only if γ is a closed geodesic of (M,F). Since $\dot\gamma$ is always in the kernel of this covector, we actually conclude that V_γ vanishes identically if and only if γ is a closed geodesic of (M,F). Therefore, for each $\ell \geq \mathrm{injrad}(M,F)$ and $\epsilon > 0$, the open subset

$$\mathcal{U}(\ell,\epsilon) := \Big\{\gamma \in \mathrm{Emb}(S^1,M) \ \Big|\ L(\gamma) \in (\ell-\epsilon^2, \ell+\epsilon^2),\ \max_{s\in S^1} |V_\gamma(s)| < \epsilon\Big\}, \tag{1.4}$$

is a neighborhood of the set of simple closed geodesics of (M,F) with length ℓ. We will employ the notation

$$\mathrm{Emb}(S^1,M)^{<\ell} := \{\gamma \in \mathrm{Emb}(S^1,M) \mid L(\gamma) < \ell\}$$

to denote the sublevel sets of the length functional.

Remark 1.1. In the literature, closed geodesics γ are usually required to be parametrized with constant positive speed, that is, the function $t \mapsto F(\gamma(t),\dot\gamma(t))$ is required to be constant and positive. In this paper, instead, we allow closed geodesics to be parametrized arbitrarily as immersed curves. Indeed, the equation $V_\gamma \equiv 0$ is independent of the parametrization of γ (Lemma 2.3). From Sect. 4 on, we will often consider closed geodesics that are non-trivial critical points of the energy function, and thus parametrized with constant speed. □

We consider the evolution equation

$$\partial_t \gamma_t = V_{\gamma_t} N_{\gamma_t}$$

with prescribed initial condition $\gamma_0 \in \mathrm{Emb}(S^1,M)$, where $\gamma_t \in C^\infty(S^1,M)$ for all t for which it is defined. Up to slowing down the time evolution when γ_t becomes short, the solutions of this equation give rise to a curve shortening semi-flow, whose properties are summarized in the next statement.

Theorem 1.2. *Let (M,F) be a closed, orientable, reversible Finsler manifold, and $\rho_0 > 0$. There exists a continuous map*

$$\psi : [0,\infty) \times \mathrm{Emb}(S^1,M) \to \mathrm{Emb}(S^1,M), \quad \psi(t,\gamma) = \psi_t(\gamma),$$

with the following properties:

(i) *It is a semi-flow, i.e., $\psi_0 = \mathrm{id}$ and $\psi_{t_2} \circ \psi_{t_1} = \psi_{t_1+t_2}$ for all $t_1, t_2 \geq 0$.*
(ii) *It is equivariant with respect to the action of circle diffeomorphisms, i.e., $\psi_t(\gamma \circ \theta) = \psi_t(\gamma) \circ \theta$ for all $\gamma \in \mathrm{Emb}(S^1,M)$ and $\theta \in \mathrm{Diff}(S^1)$.*
(iii) *The length function is not increasing along the trajectories of ψ_t. More precisely, $\frac{\mathrm{d}}{\mathrm{d}t} L(\psi_t(\gamma)) \leq 0$ with equality if and only if γ is a closed geodesic or $L(\gamma) \leq \rho_0$.*
(iv) *For each $\ell > 2\rho_0$ and $\epsilon > 0$, there exists $\delta > 0$ and a continuous function*

$$\tau : \mathrm{Emb}(S^1,M)^{<\ell+\delta} \to (0,\infty)$$

such that

$$\psi_t(\gamma) \in \mathcal{U}(\ell,\epsilon) \cup \mathrm{Emb}(S^1,M)^{<\ell-\delta}, \quad \forall \gamma \in \mathrm{Emb}(S^1,M)^{<\ell+\delta},\ t \geq \tau(\gamma).$$

(v) *If there are no simple closed geodesics with length in* $[\ell_1, \ell_2] \subset (2\rho_0, \infty)$, *then there exists a continuous function*

$$\tau : \mathrm{Emb}(S^1, M)^{<\ell_2} \to (0, \infty)$$

such that

$$\psi_t(\gamma) \subset \mathrm{Emb}(S^1, M)^{<\ell_1}, \quad \forall \gamma \in \mathrm{Emb}(S^1, M)^{<\ell_2},\ t \geq \tau(\gamma).$$

Most of the points in this theorem follow from Oaks [39], except point (iv), which is crucial for the applications.

1.2. Closed geodesics on Finsler 2-spheres

We already anticipated that the semi-flow of Theorem 1.2 allows to extend the celebrated Lusternik–Schnirelmann's theorem [30] to the reversible Finsler setting. Actually, it will also allow to extend the characterization of simple Zoll geodesic flows on the 2-sphere, originally claimed in the Riemannian case by Lusternik [29] and rigorously proved in [37]. We recall that a Finsler manifold is called Zoll when all its unit-speed geodesics are closed with the same minimal period, and simple Zoll if, in addition, all the geodesics are simple closed. We denote by $\sigma_{\mathrm{s}}(S^2, F)$ the simple length spectrum of a Finsler 2-sphere, which is the set of lengths of its simple closed geodesics.

Theorem 1.3. *Every reversible Finsler* 2*-sphere* (S^2, F) *has at least three geometrically distinct simple closed geodesics. More precisely*:

(i) *If* $\sigma_{\mathrm{s}}(S^2, F)$ *is a singleton, then* (S^2, F) *is simple Zoll.*
(ii) *If* $\sigma_{\mathrm{s}}(S^2, F)$ *contains exactly two elements, then there exists* $\ell \in \sigma_{\mathrm{s}}(S^2, F)$ *such that every point of* S^2 *lies on a simple closed geodesic of length* ℓ.
(iii) *Assume that, for any compact interval* $[\ell_1, \ell_2] \subset (0, \infty)$, (S^2, F) *has only finitely many simple closed geodesics with length in* $[\ell_1, \ell_2]$. *Then,* (S^2, F) *has three simple closed geodesics* $\gamma_1, \gamma_2, \gamma_3$ *with lengths* $L(\gamma_1) < L(\gamma_2) < L(\gamma_3)$ *and such that, for each* $i = 1, 2, 3$, γ_i *has non-trivial local homology in degree* i *with* $\mathbb{Z}_2$ *coefficients.*

For the definition of the local homology of a closed geodesic, we refer the reader to Sect. 4.3. Point (iii) in Theorem 1.3 may look technical, but it is a crucial ingredient for the proof of Theorem 1.5 here below. Even though it is claimed in [22], it does not have a proper proof in the published literature.

Remark 1.4. As it was pointed out in [17] and [15, Remark 5.3], a Zoll reversible Finsler 2-sphere is actually simple Zoll provided it has a simple closed geodesic. This, together with Theorem 1.3, implies that any Zoll reversible Finsler 2-sphere is automatically simple Zoll. □

Finally, we can state the last result, that generalizes Bangert–Franks–Hingston's theorem.

Theorem 1.5. *Every reversible Finsler* 2*-sphere* (S^2, F) *has infinitely many geometrically distinct closed geodesics.*

The main ideas for this theorem remain the same as in the Riemannian case, but nevertheless we provide a full and rather self-contained account, which insures that certain arguments of the original proof that looked Riemannian can indeed be carried over in the Finsler case. At the same time, our treatment fills some expository gaps present in the literature.

Finally, we would like to mention a related problem that saw major advances in recent years. Closed geodesics on Riemannian surfaces are in particular minimal hypersurfaces. In 1982, Yau [44] conjectured that every closed Riemannian 3-manifold has infinitely many smooth, closed, immersed minimal hypersurfaces. An even stronger statement was proved by Irie–Marques–Neves [24]: on any closed n-manifold, with $3 \leq n \leq 7$, equipped with a C^∞-generic Riemannian metric, the union of all smooth, closed, embedded minimal hypersurfaces is dense. We refer the reader to the survey [32] for more background and details.

1.3. Organization of the paper

In Sect. 2, we provide a construction of the curve shortening semi-flow, and prove Theorem 1.2. In Sect. 3, we prove Theorem 1.3, except the technical point (iii). In Sect. 4, we provide the background on the classical critical point theory for the Finsler energy function, and we will prove Theorem 1.3(iii) at the end of the section. Finally, in Sect. 5, we prove Theorem 1.5.

2. The curve shortening semi-flow

2.1. The evolution equation

We consider a 1-parameter family of curves $\gamma_t \in \mathrm{Emb}(S^1, M)$ evolving according to the partial differential equation

$$\partial_t \gamma_t(u) = w_t(u) n_t(u) \tag{2.1}$$

where $w_t := V_{\gamma_t}$ and $n_t := N_{\gamma_t}$. For every $\gamma_0 \in \mathrm{Emb}(S^1, M)$, we denote by $\tau_{\gamma_0} \in [0, \infty]$ the largest extended real number such that there is a well-defined solution $\gamma_t \in \mathrm{Emb}(S^1, M)$ of (2.1) for all $t \in [0, \tau_{\gamma_0})$, with $\gamma_t|_{t=0} = \gamma_0$. We set

$$\mathcal{U} := \big\{(t, \gamma_0) \bigm| \gamma_0 \in \mathrm{Emb}(S^1, M),\ t \in [0, \tau_{\gamma_0})\big\}.$$

Theorem 2.1. *There is a unique map*

$$\phi : \mathcal{U} \to \mathrm{Emb}(S^1, M), \quad \phi(t, \gamma_0) = \phi_t(\gamma_0) = \gamma_t,$$

where γ_t is the solution of (2.1) with initial condition γ_0, satisfying the following properties:

(i) *The subset $\mathcal{U} \subset [0, \infty) \times \mathrm{Emb}(S^1, M)$ is an open neighborhood of $\{0\} \times \mathrm{Emb}(S^1, M)$, and ϕ is continuous.*
(ii) *The map ϕ is equivariant under the action of $\mathrm{Diff}(S^1)$ on $\mathrm{Emb}(S^1, M)$, i.e., $\phi_t(\gamma \circ \theta) = \phi_t(\gamma) \circ \theta$ for all $\gamma \in \mathrm{Emb}(S^1, M)$ and $\theta \in \mathrm{Diff}(S^1)$.*
(iii) *For each $\gamma \in \mathrm{Emb}(S^1, M)$ we have $\frac{\mathrm{d}}{\mathrm{d}t} L(\phi_t(\gamma)) \leq 0$, with equality if and only if γ is a closed geodesic of (M, F).*

(iv) *For each* $\gamma \in \mathrm{Emb}(S^1, M)$*, if*

$$\ell_\gamma := \lim_{t\to\tau_\gamma^-} L(\phi_t(\gamma)) > 0$$

then $\tau_\gamma = \infty$*.*

The proof of this theorem will be carried out in the rest of the section: point (i) will be proved in Sect. 2.3; point (ii) is a consequence of Lemma 2.3; point (iii) will be proved in Sect. 2.2. The fact that ϕ is well defined as a map of the above form (i.e., mapping the space $\mathrm{Emb}(S^1, M)$ into itself) and point (iv) will be proved in Sect. 2.4. In analogy with the Riemannian case, we call ϕ_t the *curve shortening semi-flow* of (M, F). Notice that ϕ_t is not a flow (despite in the Riemannian literature it is often called a flow): indeed, it is only defined for $t \geq 0$, and thus satisfies $\phi_{t_1} \circ \phi_{t_2} = \phi_{t_1+t_2}$ only for $t_1, t_2 \geq 0$.

All closed geodesics of a closed Finsler surface (M, F) have length strictly larger than the injectivity radius $\mathrm{injrad}(M, F)$. It is sometimes convenient to have a well-defined curve shortening semi-flow defined for all positive times even for those trajectories that are not converging to a closed geodesic. We can achieve this by slowing down the curve shortening semi-flow lines in the sublevel set $\{L < \mathrm{injrad}(M, F)\}$, as follows. We fix a constant

$$\rho_0 > 0, \tag{2.2}$$

which will be chosen smaller than $\mathrm{injrad}(M, F)$ in the applications. We consider a monotone increasing smooth function $\chi : [0, \infty) \to [0, 1]$ such that $\mathrm{supp}(\chi) = [\rho_0, \infty)$ and $\chi(\ell) = 1$ for all $\ell \in [2\rho_0, \infty)$. We define

$$\psi : [0, \infty) \times \mathrm{Emb}(S^1, M) \to \mathrm{Emb}(S^1, M), \quad \psi(t, \gamma_0) = \psi_t(\gamma_0) = \gamma_t,$$

where γ_t is the solution of the partial differential equation

$$\partial_t\gamma_t(u) = \chi(L(\gamma_t))V_{\gamma_t}(u)N_{\gamma_t}(u) \tag{2.3}$$

The semi-flow ψ_t is the one that we employ for Theorem 1.2. Its properties, except Theorem 1.2(iv) and (v), will be direct consequences of the above Theorem 2.1 by means of the following lemma.

Lemma 2.2. *There exists a smooth function* $T : \mathrm{Emb}(S^1, M) \times [0, \infty) \to [0, \infty)$ *monotone increasing in the second variable such that* $T(\gamma, \cdot) < \tau_\gamma$ *and*

$$\psi_t(\gamma) = \phi_{T(\gamma,t)}(\gamma), \quad \forall \gamma \in \mathrm{Emb}(S^1, M),\ t \in [0, \infty).$$

Moreover,

(i) $T(\gamma, t_1 + t_2) = T(\phi_{T(\gamma,t_1)}(\gamma), t_2)$,
(ii) $T(\gamma, t) = t$ *if* $L(\phi_t(\gamma)) \geq 2\rho_0$,
(iii) $T(\gamma, t) = 0$ *if* $L(\gamma) \leq \rho_0$,
(iv) $T(\gamma \circ \theta, t) = T(\gamma, t)$ *for all* $\theta \in \mathrm{Diff}(S^1)$.

Proof. We denote $\gamma_0 := \gamma$ and $\gamma_t := \phi_t(\gamma_0)$. The smooth map $(s, t) \mapsto \gamma_{T(\gamma,t)}(s)$ is a solution of (2.3) if and only if

$$\chi(L(\gamma_{T(\gamma,t)}))V_{\gamma_{T(\gamma,t)}}N_{\gamma_{T(\gamma,t)}} = \partial_t\gamma_{T(\gamma,t)} = (\partial_t T(\gamma, t))V_{\gamma_{T(\gamma,t)}}N_{\gamma_{T(\gamma,t)}}.$$

Therefore, the desired function $t \mapsto T(\gamma, t)$ is a solution of the ordinary differential equation

$$\begin{aligned} \partial_t T(\gamma,t) &= \chi(L(\gamma_{T(\gamma,t)})), \\ T(\gamma,0) &= 0. \end{aligned} \tag{2.4}$$

This readily implies that T is smooth as a function of (γ, t), and not decreasing. Point (i) readily follows from the semi-flow property $\phi_{t_1+t_2} = \phi_{t_1} \circ \phi_{t_2}$ of the curve shortening. If $L(\gamma_{T(\gamma,t)}) \geq 2\rho_0$, then $L(\gamma_{T(\gamma,t')}) \geq 2\rho_0$ and $\chi(L(\gamma_{T(\gamma,t')})) = 1$ for all $t' \in [0,t]$, which implies point (ii). If $L(\gamma) \leq \rho_0$, then $L(\gamma_t) \leq \rho_0$ and $\chi(L(\gamma_{T(\gamma,t)})) = 0$ for all $t \in (0, \tau_\gamma)$, which implies point (iii). Finally, if we set $T_\theta(\gamma,t) := T(\gamma \circ \theta, t)$ for some $\theta \in \mathrm{Diff}(S^1)$, we readily see that T_θ is also a solution of the ordinary differential equation (2.4). Since such equation has a unique solution, we have point (iv). $\square$

The function V_γ is a generalization of the Riemannian curvature of immersed curves in oriented Riemannian surfaces. Theorem 2.1(ii) readily follows from the following statement.

Lemma 2.3. *For each $\theta \in \mathrm{Diff}(S^1)$, we have*

$$N_{\gamma\circ\theta} = \mathrm{sign}(\dot\theta) N_\gamma \circ \theta, \quad V_{\gamma\circ\theta} = \mathrm{sign}(\dot\theta) V_\gamma \circ \theta.$$

Proof. The statement concerning the normal vector N_γ is clear. Since the Finsler metric F is 1-homogeneous in the fibers $\mathrm{T}_x M$, we have $F_v(x, \lambda v) = F_v(x,v)$ for all $\lambda > 0$. Moreover, we have $F_x(x,\lambda v) = \lambda F_x(x,v)$, $F_{xv}(x,\lambda v) = F_{xv}(x,v)$, $F_{vv}(x,v) = \lambda F_{vv}(x,\lambda v)$. Therefore, if we set $r = \theta(u)$,

$$\begin{aligned} V_{\gamma\circ\theta}(u) &:= \frac{\big(\frac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma(\theta(u)), \dot\gamma(\theta(u))) - \dot\theta(u) F_x(\gamma(\theta(u)), \dot\gamma(\theta(u)))\big) N_{\gamma\circ\theta}(u)}{\|\dot\gamma(\theta(u))\|\, |\dot\theta(u)|} \\ &= \frac{\dot\theta(u)\big(\frac{\mathrm{d}}{\mathrm{d}r} F_v(\gamma(r), \dot\gamma(r)) - F_x(\gamma(r), \dot\gamma(r))\big) N_{\gamma\circ\theta}(u)}{\|\dot\gamma(r)\|\, |\dot\theta(u)|} \\ &= \mathrm{sign}(\dot\theta(u)) \frac{\big(\frac{\mathrm{d}}{\mathrm{d}r} F_v(\gamma(r), \dot\gamma(r)) - F_x(\gamma(r), \dot\gamma(r))\big) N_\gamma(r)}{\|\dot\gamma(r)\|} \\ &= \mathrm{sign}(\dot\theta(s)) V_\gamma(\theta(u)). \end{aligned}$$

$\square$

2.2. The anti-gradient of the length

The space $\mathrm{Emb}(S^1, M)$, equipped with the C^∞ topology, is a Fréchet manifold (indeed, it is an open subset of the Fréchet manifold $C^\infty(S^1, M)$). The tangent space $\mathrm{T}_\gamma \mathrm{Emb}(S^1, M)$ is precisely the space of smooth 1-periodic vector field X along γ. The length function

$$L : \mathrm{Emb}(S^1, M) \to (0,\infty), \quad L(\gamma) = \int_0^1 F(\gamma(u), \dot\gamma(u))\, \mathrm{d}u$$

is Gateaux differentiable (it is actually smooth, but we will not need it throughout this paper). Its differential can be computed as

$$\mathrm{d}L(\gamma)X = \int_0^1 \big(F_x(\gamma(u), \dot\gamma(u)) - \tfrac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma(u), \dot\gamma(u))\big)\, X(u)\, \mathrm{d}u. \tag{2.5}$$

Lemma 2.4. *For each $a \in C^\infty(S^1, \mathbb{R})$, we have* $\mathrm{d}L(\gamma)a\dot\gamma = 0$.

Proof. If we set $\gamma_\epsilon(u) := \gamma(u + \epsilon a(u))$, we have $a\dot\gamma = \partial_\epsilon \gamma_\epsilon|_{\epsilon=0}$. Since, for all $|\epsilon|$ small enough, γ_ϵ is an embedded curve obtained by reparametrization of γ, we have $L(\gamma_\epsilon) = L(\gamma)$ and $\mathrm{d}L(\gamma)a\dot\gamma = \frac{\mathrm{d}}{\mathrm{d}\epsilon}\big|_{\epsilon=0} L(\gamma_\epsilon) = 0$. □

The Riemannian metric g introduces an L^2 Riemannian metric on $\mathrm{Emb}(S^1, M)$ given by

$$\langle\langle X, Y\rangle\rangle_\gamma = \int_{S^1} g(X(u), Y(u))\|\dot\gamma(u)\|\,\mathrm{d}u, \quad \forall X, Y \in \mathrm{T}_\gamma\mathrm{Emb}(S^1, M). \quad (2.6)$$

Thanks to the factor $\|\dot\gamma(u)\|$ in the integrand, the inner product is invariant under the action of $\mathrm{Diff}(S^1)$, i.e.,

$$\langle\langle X \circ \theta, Y \circ \theta\rangle\rangle_{\gamma\circ\theta} = \langle\langle X, Y\rangle\rangle_\gamma, \quad \forall \theta \in \mathrm{Diff}(S^1). \quad (2.7)$$

We denote by ∇L the gradient of the length functional with respect to this inner product. Namely, $\nabla L(\gamma)$ is the 1-periodic vector field along γ defined by

$$\mathrm{d}L(\gamma)X = \langle\langle \nabla L(\gamma), X\rangle\rangle_\gamma.$$

Lemma 2.5. $\nabla L(\gamma) = -V_\gamma N_\gamma$.

Proof. Consider an arbitrary $X \in \mathrm{T}_\gamma\mathrm{Emb}(S^1, M)$, which we can uniquely write as $X(u) = a(u)\dot\gamma + b(u)N_\gamma$, where $b(u) = g(X(u), N_\gamma(u))$. By Lemma 2.4 and Eq. (2.5), we compute

$$\begin{aligned}
\mathrm{d}L(\gamma)X &= \mathrm{d}L(\gamma)a\dot\gamma + \mathrm{d}L(\gamma)bN_\gamma = \mathrm{d}L(\gamma)bN_\gamma \\
&= \int_0^1 \big(F_x(\gamma(u), \dot\gamma(u)) - \tfrac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma(u), \dot\gamma(u))\big)\, b(u)\, N(u)\,\mathrm{d}u \\
&= \int_0^1 g(-V_\gamma(u)N_\gamma(u), X(u))\, \|\dot\gamma(u)\|\,\mathrm{d}u.
\end{aligned}$$

□

Therefore, the curve shortening equation (2.1) can be seen as the anti-gradient flow equation of L associated to the L^2-Riemannian metric on $\mathrm{Emb}(S^1, M)$, i.e.,

$$\partial_t \gamma_t = -\nabla L(\gamma_t). \quad (2.8)$$

The invariance (2.7), together with Lemma 2.5, provides an alternative proof of Lemma 2.3. Moreover, if a solution γ_t is well defined for $t \in [a, b]$, then

$$L(\gamma_a) - L(\gamma_b) = \int_a^b \|\nabla L(\gamma_t)\|^2 \mathrm{d}t = \int_a^b \int_{S^1} V_{\gamma_t}(u)^2\, \|\dot\gamma_t(u)\|\,\mathrm{d}u\,\mathrm{d}t. \quad (2.9)$$

It is well known that the closed geodesics of (M, F) are critical points of L, that is, those γ such that $V_\gamma \equiv 0$. Therefore, $\partial_t L(\gamma_t) \le 0$ with equality if and only if γ_t is a closed geodesic of (M, F). This settles Theorem 2.1(iii).

Remark 2.6. (Alternative curve shortening) The PDE (2.1) of the curve shortening is not canonically associated to the Finsler metric F, as it also involves the auxiliary Riemannian metric g. This choice of curve shortening semi-flow turns out to be the most convenient for the later computations. Alternatively, one could also study a curve shortening semi-flow whose definition does not involve an auxiliary Riemannian metric: this is done by replacing, in (2.8), the gradient ∇ with the one induced by the following Riemannian metric on $\mathrm{Emb}(S^1, M)$

$$\langle\!\langle X, Y\rangle\!\rangle'_\gamma = \int_{S^1} F(\gamma(u), \dot\gamma(u)) \left(\tfrac{1}{2}F^2\right)_{vv} (\gamma(u), \dot\gamma(u))[X(u), Y(u)]\, \mathrm{d}u,$$
$$\forall X, Y \in \mathrm{T}_\gamma \mathrm{Emb}(S^1, M).$$

For each $v \in \mathrm{T}_q M$, we define v^F to be the positive orthogonal to v with respect to the inner product $(F^2)_{vv}(q, v)[\cdot, \cdot]$ with norm $F(q, v^F) = F(q, v)$. If we set

$$Z_\gamma(u) := \frac{\left(\frac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma(u), \dot\gamma(u)) - F_x(\gamma(u), \dot\gamma(u))\right)\dot\gamma(u)^F}{F(\gamma(u), \dot\gamma(u))},$$

the alternative curve shortening semi-flow is precisely given by

$$\partial_t \gamma_t(u) = \frac{Z_{\gamma_t}(u)}{F(\gamma_t(u), \dot\gamma_t(u))} \dot\gamma_t(u)^F.$$

□

2.3. Short-time existence

To prove Theorem 2.1(i), it is convenient to work in suitable local coordinates around a fixed curve $\gamma_0 \in \mathrm{Emb}(S^1, M)$. We denote by $\exp : \mathrm{T}M \to M$ the exponential map of (M, g). There exists $\rho > 0$ and an open set $U \subset M$ of $\gamma_0(S^1)$ such that the map

$$\xi : S^1 \times (-\rho, \rho) \to U, \quad \xi(u, r) = \exp_{\gamma_0(u)}(r\, N_{\gamma_0}(u))$$

is a diffeomorphism.

We define the smooth map

$$\Xi : C^\infty(S^1, (-\rho, \rho)) \to \mathrm{Emb}(S^1, M), \quad \Xi(z)(u) = \xi(u, z(u)).$$

Let us show that this map is open and injective. We first define the vector field N on U by

$$N(\xi(u, r)) = \frac{\mathrm{d}}{\mathrm{d}r}\xi(u, r) = \mathrm{d}\exp_{\gamma_0(u)}(r\, N_{\gamma_0}(u)) N_{\gamma_0}(u),$$

and notice that $\|N(q)\| = 1$ for all $q \in U$. Thus, we have $\mathrm{d}\Xi(z)w = W$, where

$$W(u) = w(u) N(\Xi(z)(u)),$$

and this latter vector field along $\Xi(z)$ is non-zero provided the function w is non-zero. Hence Ξ is an immersion. Clearly, Ξ is injective, for ξ is a diffeomorphism. Finally, the equality

$$\mathrm{dist}(\Xi(z)(u), \gamma_0(u)) = |z(u)|, \quad \forall z \in C^\infty(S^1, (-\rho, \rho)),\ u \in S^1$$

implies that Ξ is an open map onto its image.

Since $\mathrm{Diff}(S^1)$ acts freely on $\mathrm{Emb}(S^1, M)$, the map

$$\Psi : C^\infty(S^1, (-\rho, \rho)) \times \mathrm{Diff}(S^1) \to \mathrm{Emb}(S^1, M),$$
$$\Psi(z, \theta)(u) = \Xi(z)(\theta(u)) = \exp_{\gamma_0(\theta(u))}\big(z(\theta(u))N_{\gamma_0}(\theta(u))\big)$$

is open and injective onto a neighborhood of γ_0. The differential of Ψ is given by

$$\mathrm{d}\Psi(z, \theta)(v, \tau) = V,$$

where

$$V(u) = v(\theta(u))N(\Psi(z, \theta)(u)) + \tau(\theta(u))\Xi(z)^\cdot(\theta(u)).$$

Here, we have denoted $\Xi(z)^\cdot(u) := \frac{\partial}{\partial u}\Xi(z)(u)$

The map Ψ pulls-back the L^2 inner product (2.6) to

$$\begin{aligned}
&\langle\!\langle\!\langle (v, \tau), (w, \sigma)\rangle\!\rangle\!\rangle_{(z,\theta)} \\
&\quad := \langle\!\langle \mathrm{d}\Psi(z, \theta)(v, \tau), \mathrm{d}\Psi(z, \theta)(w, \sigma)\rangle\!\rangle_{\Psi(z,\theta)} \\
&\quad = \int_{S^1} \Big(v(u)\, w(u) + v(u)\, \sigma(u)\, a_z(u) + w(u)\, \tau(u)\, a_z(u) \\
&\qquad + \tau(u)\, \sigma(u)\, b_z(u)^2\Big)\, b_z(u)\, \mathrm{d}u,
\end{aligned}$$

where

$$\begin{aligned}
a_z(u) &:= g(N(\Xi(z)(u)), \Xi(z)^\cdot(u)), \\
b_z(u) &:= \|\Xi(z)^\cdot(u)\|.
\end{aligned}$$

Notice that this inner product is actually independent of $\theta \in \mathrm{Diff}(S^1)$, and therefore, we will simply write

$$\langle\!\langle\!\langle (v, \tau), (w, \sigma)\rangle\!\rangle\!\rangle_{(z,\theta)} = \langle\!\langle\!\langle (v, \tau), (w, \sigma)\rangle\!\rangle\!\rangle_z. \tag{2.10}$$

To write expressions in local coordinates, let us pull-back the Finsler metric F by ξ. We obtain the Finsler metric $G := \xi^* F$ on $S^1 \times (-\epsilon, \epsilon)$ given by

$$G(q, v) = F(\xi(q), \mathrm{d}\xi(q)v), \quad \forall q \in S^1 \times (-\rho, \rho),\ v \in \mathbb{R}^2.$$

The composition of the length functional L with Ψ reads

$$\begin{aligned}
L \circ \Psi(z, \theta) = L \circ \Xi(z) &= \int_{S^1} F\big(\tfrac{\mathrm{d}}{\mathrm{d}u}\Xi(z)(u)\big)\, \mathrm{d}u \\
&= \int_{S^1} G(\underbrace{(u, z(u))}_{q}, \underbrace{(1, \dot z(u))}_{v})\, \mathrm{d}u.
\end{aligned}$$

Let us compute the derivative

$$\begin{aligned}
\mathrm{d}(L \circ \Xi)(z)w &= \int_{S^1} \big(G_{q_2}\, w + \partial_{v_2} G\, \dot w\big)\, \mathrm{d}u \\
&= \int_{S^1} \big(G_{q_2} - G_{q_1 v_2} - G_{q_2 v_2}\, \dot z - G_{v_2 v_2}\, \ddot z\big) w\, \mathrm{d}u. \qquad (2.11)
\end{aligned}$$

We denote by $(v_z, \tau_z) := \nabla(L \circ \Psi)(z)$ the gradient of $L \circ \Psi$ with respect to the inner product (2.10), i.e.,

$$\langle\!\langle\!\langle (v_z, \tau_z), (w, \sigma) \rangle\!\rangle\!\rangle_z = \mathrm{d}(L \circ \Psi)(z, \theta)(w, \sigma).$$

Since $L \circ \Psi(z, \theta)$ is independent of $\theta \in \mathrm{Diff}(S^1)$, we have

$$0 = \langle\!\langle\!\langle (v_z, \tau_z), (0, \sigma) \rangle\!\rangle\!\rangle_z = \int_{S^1} \Big(v_z(u)\, a_z(u)\, b_z(u) + \tau_z(u)\, b_z(u)^3 \Big)\, \sigma(u)\, \mathrm{d}u,$$

which implies that

$$\tau_z(u) = -v_z(u)\, \frac{a_z(u)}{b_z(u)^2}, \quad \forall u \in S^1.$$

On the other hand, we have

$$\begin{aligned} \langle\!\langle\!\langle (v_z, \tau_z), (w, 0) \rangle\!\rangle\!\rangle_z &= \int_{S^1} \Big(v_z(u) + \tau_z(u)\, a_z(u) \Big)\, b_z(u)\, w(u)\, \mathrm{d}u \\ &= \int_{S^1} \Big(1 - \tfrac{a_z(u)^2}{b_z(u)^2} \Big) v_z(u)\, b_z(u)\, w(u)\, \mathrm{d}u \\ &= \mathrm{d}(L \circ \Xi)(z) w. \end{aligned} \tag{2.12}$$

Notice that the quotient $a_z(u)/b_z(u)$ is well defined. Indeed, the curve $s \mapsto \Xi(z)(u)$ is transverse to the vector field N, and therefore

$$\tfrac{a_z(u)^2}{b_z(u)^2} = g\big(N(\Xi(z)(u)), \tfrac{\Xi(z)^\cdot(u)}{|\Xi(z)^\cdot(u)|_{\Xi(z)(u)}}\big)^2 < 1.$$

Equations (2.11) and (2.12) imply that

$$v_z = \Big(1 - \tfrac{a_z^2}{b_z^2}\Big)^{-1} b_z(u)^{-1} \big(G_{q_2} - G_{q_1 v_2} - G_{q_2 v_2}\, \dot z - G_{v_2 v_2}\, \ddot z \big).$$

The integral curves of the anti-gradient $-\nabla(L \circ \Xi)$ are solutions

$$(z, \theta) : [0, T) \times S^1 \to (-\rho, \rho) \times S^1$$

of the partial differential equation

$$\partial_t (z, \theta) = (-v_z, v_z a/b^2). \tag{2.13}$$

In particular, z is a solution of the partial differential equation

$$\partial_t z = \tfrac{b_z}{b_z^2 - a_z^2} \Big(G_{v_2 v_2}\, \partial_u^2 z + G_{q_2 v_2}\, \partial_u z + G_{q_1 v_2} - G_{q_2} \Big). \tag{2.14}$$

Since G is a Finsler metric, the second derivative $G_{vv}(q, v)$ is positive semi-definite and its kernel is generated by v. Therefore, $G_{v_2 v_2}((u, z(u)), (1, \dot z(u))) \neq 0$, and (2.14) is a parabolic partial differential equation. The local theory for this class of equations (see, e.g., [35]) provides the following statement.

Proposition 2.7. *For each $z_0 \in C^\infty(S^1, (-\rho, \rho))$, there exists $\epsilon > 0$ and a unique smooth solution $z : [0, \epsilon) \times S^1 \to (-\rho, \rho)$ of (2.14) such that $z(0, \cdot) = z_0$. Moreover, z depends continuously on the initial condition z_0 in the C^∞ topology.* $\square$

Assume that $z : [0, \epsilon) \times S^1 \to (-\rho, \rho)$ is the smooth solution given by Proposition 2.7. Up to reducing $\epsilon > 0$, we can easily find a smooth $\theta : [0, \epsilon) \times S^1 \to S^1$ such that (z, θ) is a solution of the curve shortening equation (2.13) with $\theta(0, \cdot) = \mathrm{id}$. Indeed, for each $s \in S^1$, such a θ is the unique smooth solution of the ordinary differential equation

$$\partial_t \theta(t, s) = -\tau_z(\theta(t, s)).$$

The smooth map

$$\gamma : [0, \epsilon) \times S^1 \to M, \quad \gamma(t, s) = \Psi(z(t, \cdot), \theta(t, \cdot))(s) = \xi(\theta(t, s), z(t, \theta(t, s)))$$

is thus the unique smooth solution of the curve shortening equation (2.1) with $\gamma(0, \cdot) = \Xi(z) = \gamma_0$. Summing up, we have proved the following statement, which implies Theorem 2.1(i).

Theorem 2.8. (Local existence and uniqueness) *For each* $\gamma_0 \in \mathrm{Emb}(S^1, M)$, *there exists* $\epsilon > 0$ *and a unique smooth solution* $\gamma : [0, \epsilon) \times S^1 \to M$ *of the curve shortening equation* (2.1) *such that* $\gamma(0, \cdot) = \gamma_0$. *Moreover,* γ *depends continuously on the initial condition* γ_0 *in the* C^∞ *topology.* □

2.4. Long-time existence

We denote by SM the unit tangent bundle of M with respect to the auxiliary Riemannian metric g, i.e.,

$$SM = \big\{(x, v) \in \mathrm{T}M \ \big|\ \|v\| = 1\big\}. \tag{2.15}$$

To prove that ϕ is well defined as a map onto $\mathrm{Emb}(S^1, M)$, and that there is long-time existence of solutions of the curve shortening equation (Theorem 2.1(iv)), it suffices to show that the factor w_t in the right-hand side of (2.1) can be expressed by means of a suitable smooth function

$$V : \mathbb{R} \times SM \to \mathbb{R}, \quad V(\kappa_t(u), \gamma_t(u), \tau_t(u)),$$

and invoke the general results of Angenent [3] and Oaks [39]. Here, κ_t denotes the Riemannian curvature of γ_t measured with respect to the auxiliary Riemannian metric g, and $\tau_t(u) := \dot\gamma_t(u)/\|\dot\gamma_t(u)\|$ its unit tangent vector. By expanding the definition of w_t, we have

$$\begin{aligned} w_t &= \frac{\big(\frac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma_t, \dot\gamma_t) - F_x(\gamma_t, \dot\gamma_t)\big) n_t}{\|\dot\gamma_t\|} \\ &= F_{vv}(\gamma_t, \tau_t)[\ddot\gamma_t/\|\dot\gamma_t\|^2, n_t] + F_{xv}(\gamma_t, \tau_t)[\tau_t, n_t] - F_x(\gamma_t, \tau_t) n_t. \end{aligned}$$

Since $F_{vv}(x, v)v = 0$, the first summand in the last line can be rewritten as

$$\begin{aligned} F_{vv}(\gamma_t, \tau_t)[\ddot\gamma_t/\|\dot\gamma_t\|^2, n_t] &= F_{vv}(\gamma_t, \tau_t)[n_t, n_t]\, g(\ddot\gamma_t/\|\dot\gamma_t\|^2, n_t) \\ &= F_{vv}(\gamma_t, \tau_t)[n_t, n_t]\kappa_t - F_{vv}(\gamma_t, \tau_t)[n_t, n_t] g(\Gamma_{\gamma_t}[\tau_t, \tau_t], n_t). \end{aligned}$$

Here,

$$\Gamma_x[v, w] = \Gamma^k_{ij}(x) v_i w_j \partial_{x_k},$$

where Γ_{ij}^k are the Christoffel symbols of the metric g with respect to the local coordinates employed in the above expression. Inserting this into the expression of w_t, we obtain

$$\begin{aligned} w_t &= F_{vv}(\gamma_t, \tau_t)[n_t, n_t]\kappa_t - F_{vv}(\gamma_t, \tau_t)[n_t, n_t] g(\Gamma_{\gamma_t}[\tau_t, \tau_t], n_t) \\ &\quad + F_{xv}(\gamma_t, \tau_t)[\tau_t, n_t] - F_x(\gamma_t, \tau_t) n_t. \end{aligned}$$

Notice that the first summand $F_{vv}(\gamma_t, \tau_t)[n_t, n_t]\kappa_t$ is well defined independently of the local coordinates, as F_{vv} is simply the fiberwise Hessian of F. Therefore, since w_t is also well defined, the remaining summands

$$-F_{vv}(\gamma_t, \tau_t)[J\tau_t, J\tau_t] g(\Gamma_{\gamma_t}[\tau_t, \tau_t], J\tau_t) + F_{xv}(\gamma_t, \tau_t)[\tau_t, J\tau_t] - F_x(\gamma_t, \tau_t)$$

are well defined independently of the local coordinates as well. The expression above shows that w_t is of the form $w_t = V(\kappa_t, \gamma_t, \tau_t)$, where $V : \mathbb{R} \times SM \to \mathbb{R}$ is the smooth function

$$\begin{aligned} V(\kappa, x, v) &= F_{vv}(x, v)[Jv, Jv]\kappa - F_{vv}(x, v)[Jv, Jv] g(\Gamma_x[v, v], Jv) \\ &\quad + F_{xv}(x, v)[v, Jv] - F_x(x, v) Jv \\ &=: A(x, v)\kappa + B(x, v). \end{aligned} \tag{2.16}$$

The reversibility of F readily imply that $V(\kappa, x, v) = -V(-\kappa, x, -v)$. The function V thus satisfies in particular the assumptions required in [3, 39]. By [4, Theorem 1.3], the map ϕ takes values inside $\mathrm{Emb}(S^1, M)$. Finally, Theorem 2.1(iv) follows from [39, Corollary 6.2].

2.5. L^∞ bounds on V_γ

For any $\gamma_0 \in \mathrm{Emb}(S^1, M)$, we will write the corresponding solution of (2.1) by

$$\gamma_t = \phi_t(\gamma_0)$$

and its length by

$$\ell_t := L(\gamma_t).$$

We denote by ∇_t, ∇_u, and ∇_s the covariant derivatives associated with the Levi-Civita connection of g. It is convenient to introduce the vector field

$$D_s = \|\dot\gamma_t(u)\|^{-1} \partial_u$$

on $\mathbb{R} \times S^1$, which acts on smooth real-valued functions $f : \mathbb{R} \times S^1 \to \mathbb{R}$, $f(t, u) = f_t(u)$ by

$$D_s f_t(u) = \frac{\partial_u f_t(u)}{\|\dot\gamma_t(u)\|}.$$

We recall the classical Frenet formulas from plane Riemannian geometry:

$$\nabla_u \tau_t = \kappa_t \|\dot\gamma_t\| n_t, \quad \nabla_u n_t = -\kappa_t \dot\gamma_t.$$

By means of the PDE (2.1), we also have the following formulas.

Lemma 2.9. $\nabla_t \tau_t = (D_s w_t) n_t$, $\nabla_t n_t = -(D_s w_t)\tau_t$.

Proof. Let us compute the covariant derivative $\nabla_t\tau_t$. Since $\|\tau_t\| = \|n_t\| \equiv 1$, we have

$$g(\nabla_t\tau_t, \tau_t) = g(\nabla_t n_t, n_t) = g(\nabla_u n_t, n_t) = g(\nabla_u\tau_t, \tau_t) = 0.$$

Moreover

$$g(\nabla_t\dot\gamma_t, n_t) = g(\nabla_u\partial_t\gamma_t, n_t) = g(\nabla_u(w_t n_t), n_t) = \dot w_t,$$

which readily implies

$$\begin{aligned}\nabla_t\tau_t &= g(\nabla_t\tau_t, n_t)n_t = \frac{\dot w_t}{\|\dot\gamma_t\|}n_t = (D_s w_t)n_t,\\ \nabla_t n_t &= g(\nabla_t n_t, \tau_t)\tau_t = -\frac{g(n_t, \nabla_t\dot\gamma_t)}{\|\dot\gamma_t\|}\tau_t = -\frac{\dot w_t}{\|\dot\gamma_t\|}\tau_t = -(D_s w_t)\tau_t.\end{aligned}$$

□

Lemma 2.10. $\partial_t\|\dot\gamma_t(u)\| = -\kappa_t(u)w_t(u)\|\dot\gamma_t(u)\|$.

Proof. By means of the commutativity $\nabla_t\partial_u = \nabla_u\partial_t$ and of the PDE (2.1), we compute

$$\begin{aligned}\partial_t\|\dot\gamma_t(u)\| &= \frac{g(\nabla_t\dot\gamma_t(u), \dot\gamma_t(u))}{\|\dot\gamma_t(u)\|} = \frac{g(\nabla_u\partial_t\gamma_t(u), \dot\gamma_t(u))}{\|\dot\gamma_t(u)\|} = \frac{g(\nabla_u(w_t n_t), \dot\gamma_t(u))}{\|\dot\gamma_t(u)\|}\\ &= w_t(u)\frac{g(\nabla_u n_t, \dot\gamma_t(u))}{\|\dot\gamma_t(u)\|} = -w_t(u)\frac{g(\nabla_u\dot\gamma_t, n_t(u))}{\|\dot\gamma_t(u)\|}\\ &= -\kappa_t(u)w_t(u)\|\dot\gamma_t(u)\|.\end{aligned}$$

□

Lemma 2.11. *The curvature κ_t evolves according to the PDE*

$$\partial_t\kappa_t(u) = D_s^2 w_t(u) + w_t(u)\kappa_t^2(u) + w_t(u)k_g(\gamma_t(u)),$$

where k_g denotes the Gaussian curvature of (M, g), i.e., $k_g(x) = g(R(v, Jv)v, Jv)$ for all $v \in S_xM$.

Proof. The lemma follows by direct computation:

$$\begin{aligned}\partial_t\kappa_t &= \partial_t\frac{g(\nabla_u\tau_t, n_t)}{\|\dot\gamma_t\|}\\ &= \left(\partial_t\frac{1}{\|\dot\gamma_t\|}\right)\kappa_t\|\dot\gamma_t\| + \frac{1}{\|\dot\gamma_t\|}g(\nabla_t\nabla_u\tau_t, n_t) + \frac{1}{\|\dot\gamma_t\|}\underbrace{g(\nabla_u\tau_t, \nabla_t n_t)}_{=0}\\ &= \frac{\kappa_t w_t\|\dot\gamma_t\|}{\|\dot\gamma_t\|^2}\kappa_t\|\dot\gamma_t\| + \frac{1}{\|\dot\gamma_t\|}g(\nabla_u\nabla_t\tau_t, n_t) + \frac{1}{\|\dot\gamma_t\|}g(R(\dot\gamma_t, \partial_t\gamma_t)\tau_t, n_t)\\ &= w_t\kappa_t^2 + D_s^2 w_t + w_t k_g \circ \gamma_t.\end{aligned}$$

□

We set $\ell_t := L(\gamma_t)$, and denote by $\Gamma_t : \mathbb{R}/\ell_t\mathbb{Z} \to M$ the reparametrized γ_t with unit speed with respect to the auxiliary Riemannian metric g. Namely $\Gamma_t(s) = \gamma_t \circ \sigma_t^{-1}(s)$, where

$$\sigma_t(u) = \int_0^u \|\dot\gamma_t(r)\|\,\mathrm{d}r,$$

and therefore, $\ell_t = \sigma_t(1)$. We also set

$$\begin{aligned} W_t(s) &:= w_t \circ \sigma_t^{-1}(s), \\ N_t(s) &:= n_t \circ \sigma_t^{-1}(s), \\ K_t(s) &:= \kappa_t \circ \sigma_t^{-1}(s). \end{aligned} \tag{2.17}$$

Notice that

$$K_t(s) = g(\nabla_s \dot{\Gamma}_t, N_t), \quad W_t(s) = V(K_t(s), \dot{\Gamma}_t(s)).$$

Moreover

$$\dot{W}_t \circ \sigma_t = D_s w_t, \quad \dot{K}_t \circ \sigma_t = D_s \kappa_t.$$

If $f : SM \to \mathbb{R}$ is a smooth function, we will denote by $\nabla^{\mathrm{h}} f$ and $\nabla^{\mathrm{v}} f$ the duals of the horizontal and vertical projections, respectively, of its gradient with respect to the Sasaki metric of SM induced by g. These operators allow to express $\partial_t w_t$ as

$$\begin{aligned} \partial_t w_t &= (\partial_\kappa V)\partial_t \kappa_t + (\nabla^{\mathrm{h}} V)\partial_t \gamma_t + (\nabla^{\mathrm{v}} V)\nabla_t \tau_t \\ &= (\partial_\kappa V)\big(D_s^2 w_t + w_t \kappa_t^2 + w_t k_g(\gamma_t)\big) + (\nabla^{\mathrm{h}} V) n_t \, w_t + (\nabla^{\mathrm{v}} V) n_t \, D_s w_t. \end{aligned}$$

We set

$$A(x,v) := \partial_\kappa V(x,v) = F_{vv}(x,v)[Jv, Jv].$$

Notice that A is uniformly bounded from below by a positive constant. From now on, we will consider it evaluated at $(\gamma_t(u), \tau_t(u))$. Notice that $A\, D_s^2 w_t = D_s(A\, D_s w_t) - (D_s A)(D_s w_t)$, and

$$D_s A = \tfrac{1}{\|\dot{\gamma}_t\|}\big((\nabla^{\mathrm{h}} A)\dot{\gamma}_t + (\nabla^{\mathrm{v}} A)\nabla_u \tau_t\big) = (\nabla^{\mathrm{h}} A)\tau_t + (\nabla^{\mathrm{v}} A) n_t \, \kappa_t.$$

Therefore, $\partial_t w_t$ can be written as

$$\begin{aligned} \partial_t w_t &= A\, D_s^2 w_t + A w_t \kappa_t^2 + B\, D_s w_t + C\, w_t \\ &= D_s(A\, D_s w_t) + A w_t \kappa_t^2 + E\, D_s w_t + H\, \kappa_t D_s w_t + C\, w_t, \end{aligned}$$

where B, C, E, and H are smooth functions on SM evaluated at $(\gamma_t(u), \tau_t(u))$.

We are now going to employ the open sets $\mathcal{U}(\ell, \epsilon)$ defined in (1.4).

Lemma 2.12. *For all $\ell > 2\rho_0$, there exists a constant $c \geq 1$ with the following properties: for all $\epsilon > 0$ small enough, $\gamma_0 \in \mathrm{Emb}(S^1, M)$, and $t \geq 0$ such that*

$$\|W_0\|_{L^2} \leq \epsilon, \quad \ell - \epsilon^2 \leq \ell_t \leq \ell_0 \leq \ell + \epsilon^2,$$

we have $\|W_t\|_{L^2} \leq c\,\epsilon$.

Proof. We require $\epsilon > 0$ to be small enough so that $\ell - \epsilon^2 > 2\rho_0$, where $\rho_0 > 0$ is the constant fixed in (2.2). We consider an arbitrary $\gamma_0 \in \mathrm{Emb}(S^1, M)$ such that $\|W_0\|_{L^2} \leq \epsilon$ and $\ell_0 \in (\ell - \epsilon^2, \ell + \epsilon^2)$, and its evolution γ_t. We recall that the corresponding w_t has the form $w_t = V(\kappa_t, \gamma_t, \tau_t)$, where $\tau_t := \dot{\gamma}_t / \|\dot{\gamma}_t\|$. We compute

$$\partial_t \|W_t\|_{L^2}^2 = \int_0^1 \Big(2 w_t \,(\partial_t w_t)\, \dot{\sigma}_t + w_t^2 (\partial_t \dot{\sigma}_t)\Big) \mathrm{d}u$$

$$= \int_0^1 \Big(2w_t\, D_s(AD_s w_t) + 2Aw_t^2\kappa_t^2 + 2E\, w_t\, D_s w_t + 2H\, \kappa_t w_t D_s w_t$$
$$+ 2Cw_t^2 - \kappa_t w_t^3\Big)\dot\sigma_t\, \mathrm{d}u$$
$$= \int_0^{\ell_t}\Big(- 2A(\dot W_t)^2 + 2AW_t^2K_t^2 + 2EW_t\dot W_t + 2HK_tW_t\dot W_t$$
$$+ 2CW_t^2 - K_tW_t^3\Big)\mathrm{d}s.$$

From now on, we will denote by $c \geq 1$ a positive constant (independent of γ_t), that may increase along the different inequalities. The above expression for $\partial_t\|W_t\|^2_{L^2}$ readily implies

$$\partial_t\|W_t\|^2_{L^2} \leq -c^{-1}\|\dot W_t\|^2_{L^2} + c\left(\|W_t\dot W_t\|_{L^1} + \|K_tW_t\dot W_t\|_{L^1} + \|W_t^2K_t^2\|_{L^1}\right.$$
$$\left. +\|K_tW_t^3\|_{L^1} + \|W_t\|^2_{L^2}\right). \tag{2.18}$$

By the Peter–Paul inequality, for every $\rho > 0$, the term $\|W_t\dot W_t\|_{L^1}$ can be bounded as

$$\|W_t\dot W_t\|_{L^1} \leq \rho^2\|\dot W_t\|^2_{L^2} + \tfrac{1}{4\rho^2}\|W_t\|^2_{L^2},$$

and the term $\|K_tW_t\dot W_t\|_{L^1}$ as

$$\|K_tW_t\dot W_t\|_{L^1} \leq \rho^2\|\dot W_t\|^2_{L^2} + \tfrac{1}{4\rho^2}\|K_tW_t\|^2_{L^2}$$
$$\leq \rho^2\|\dot W_t\|^2_{L^2} + \tfrac{1}{4\rho^2}\|K_t\|^2_{L^\infty}\|W_t\|^2_{L^2}.$$

We will fix a sufficiently small constant $\rho > 0$ so that, in the inequality (2.18), the term $-c^{-1}\|\dot W_t\|^2_{L^2}$ will be able to absorb the terms $\rho^2\|\dot W_t\|^2_{L^2}$, still producing a negative factor in front of $\|\dot W_t\|^2_{L^2}$.

Equation (2.16) readily implies that the curvature K_t is related to W_t by $K_t = A^{-1}W_t + P$, where, once again, P is a smooth function on SM evaluated at $(\Gamma_t(s), \dot\Gamma_t(s))$. Therefore, $\|K_t\|_{L^\infty} \leq c\,(\|W_t\|^2_{L^\infty} + 1)$.

Inserting these estimates in (2.18), we obtain

$$\partial_t\|W_t\|^2_{L^2} \leq c\|W_t\|^2_{L^2} + c\|W_t\|^2_{L^\infty}\|W_t\|^2_{L^2} - c^{-1}\|\dot W_t\|^2_{L^2}.$$

We require $c > 0$ to be large enough so that, since $\ell - \epsilon^2 \leq \ell_t \leq \ell + \epsilon^2$, we have

$$c^{-1} \leq \ell_t \leq c.$$

If we bound from above the term $-c^{-1}\|\dot W_t\|^2_{L^2}$ by means of the inequality

$$\|W_t\|^2_{L^\infty} \leq 2\ell_t^{-1}\|W_t\|^2_{L^2} + 2\ell_t\|\dot W_t\|^2_{L^2} \leq c\|W_t\|^2_{L^2} + c\|\dot W_t\|^2_{L^2},$$

we further obtain

$$\partial_t\|W_t\|^2_{L^2} \leq c\|W_t\|^2_{L^2} + c\|W_t\|^2_{L^\infty}(\|W_t\|^2_{L^2} - c^{-1}).$$

We claim that, if $\epsilon > 0$ is small enough (independently of γ), then $\|W_t\|^2_{L^2} < c^{-1}$ for all $t \geq 0$ such that $\ell_t \geq \ell - \epsilon^2$. Indeed, assume that this is not the case. If $\epsilon^2 < 1/c$, since $\|W_0\|^2_{L^2} \leq \epsilon^2 < c^{-1}$, there must be $\tau > 0$ such that $\|W_t\|^2_{L^2} < c^{-1}$ for all $t \in [0, \tau)$, $\|W_\tau\|^2_{L^2} = c^{-1}$, and $\ell_\tau \geq \ell - \epsilon^2$. For all $t \in [0, \tau]$ we have the inequality $\partial_t \|W_t\|^2_{L^2} \leq c\|W_t\|^2_{L^2}$, and thus

$$c^{-1} = \|W_\tau\|^2_{L^2} \leq e^{c\tau}\|W_0\|^2_{L^2} \leq e^{c\tau}\epsilon^2.$$

If $\epsilon^2 \leq e^{-c}c^{-1}$, then $\tau \geq 1$. Therefore, since $\|W_\tau\|^2_{L^2} \leq e^{c(\tau-t)}\|W_t\|_{L^2}$ for all $t \in [0, \tau]$, by (2.9) we have

$$c^{-1} = \|W_\tau\|^2_{L^2} \leq e^c \int_{\tau-1}^{\tau} \|W_t\|^2_{L^2} \mathrm{d}t = e^c(\ell_{\tau-1} - \ell_\tau) \leq 2\epsilon^2 e^c,$$

which is impossible if $\epsilon^2 < 1/(2e^c c)$.

Summing up, we showed that $\partial_t \|W_t\|^2_{L^2} \leq c\|W_t\|^2_{L^2}$ provided $\ell_t \geq \ell - \epsilon^2$, and therefore

$$\|W_t\|^2_{L^2} \leq e^{ct} \int_0^t \|W_r\|^2_{L^2}\, \mathrm{d}r \leq e^c(\ell_0 - \ell_t) \leq 2\epsilon^2 e^c,$$
$$\forall t \geq 0 \text{ such that } \ell_t \geq \ell - \epsilon^2.$$

□

Lemma 2.13. *For all $\ell > 2\rho_0$, there exists a constant $c \geq 1$ with the following properties: for all $\epsilon > 0$ small enough, $\gamma_0 \in \mathrm{Emb}(S^1, M)$, and $t \geq c\log(\|\dot{W}_0\|^2_{L^2}\epsilon^{-2})$ such that*

$$\|W_0\|_{L^2} \leq \epsilon, \quad \ell - \epsilon^2 \leq \ell_t \leq \ell_0 \leq \ell + \epsilon^2,$$

we have $\|\dot{W}_t\|_{L^2} \leq c\,\epsilon$.

Proof. We require $\epsilon > 0$ to be small enough so that Lemma 2.12 holds, and in particular so that $\ell - \epsilon^2 > 2\rho_0$. We consider an arbitrary $\gamma_0 \in \mathrm{Emb}(S^1, M)$ such that $\|W_0\|_{L^2} \leq \epsilon$ and $\ell_0 \in (\ell - \epsilon^2, \ell + \epsilon^2)$, and its evolution γ_t. Once again, we will denote by $c \geq 1$ a large enough constant independent of γ_t and ϵ, possibly growing throughout the computations.

We estimate

$$\begin{aligned}
\partial_t \|\dot{W}_t\|^2_{L^2} &= \int_0^1 \Big(2(D_s w_t)\partial_t(D_s w_t)\,\dot{\sigma}_t - (D_s w_t)^2\kappa_t w_t \dot{\sigma}_t\Big)\mathrm{d}u \\
&= \int_0^1 \Big(2(D_s w_t)\partial_t \dot{w}_t + (D_s w_t)^2 \kappa_t w_t \dot{\sigma}_t\Big)\mathrm{d}u \\
&= \int_0^1 \Big(2(D_s w_t) D_s\big(A\, D_s^2 w_t + A w_t \kappa_t^2 + B\, D_s w_t + C\, w_t\big) \\
&\quad + (D_s w_t)^2 \kappa_t w_t\Big)\dot{\sigma}_t \mathrm{d}u \\
&= \int_0^{\ell_t} \Big(-2A(\ddot{W}_t)^2 - 2A\ddot{W}_t W_t K_t^2 - 2B\ddot{W}_t\dot{W}_t - C\ddot{W}_t W_t \\
&\quad + (\dot{W}_t)^2 K_t W_t\Big)\mathrm{d}s \\
&\leq -c^{-1}\|\ddot{W}_t\|^2_{L^2} + c\big(\|\ddot{W}_t W_t K_t^2\|_{L^1} + \|\ddot{W}_t \dot{W}_t\|_{L^1} + \|\ddot{W}_t W_t\|_{L^1} \\
&\quad + \|(\dot{W}_t)^2 K_t W_t\|_{L^1}\big).
\end{aligned}$$

By employing the Peter–Paul inequality and expressing K_t as an affine function of W_t, we have

$$\begin{aligned}\partial_t\|\dot W_t\|_{L^2}^2 &\leq -c^{-1}\|\ddot W_t\|_{L^2}^2 + c\big(\|W_t^3\|_{L^2}^2+\|W_t^2\|_{L^2}^2+\|W_t\|_{L^2}^2+\|\dot W_t\|_{L^2}^2\\ &\qquad +\|(\dot W_t)^2W_t^2\|_{L^1}+\|(\dot W_t)^2W_t\|_{L^1}\big)\\ &= -c^{-1}\|\ddot W_t\|_{L^2}^2 + c\big(\|W_t^6\|_{L^1}+\|W_t^4\|_{L^1}+\|W_t\|_{L^2}^2+\|\dot W_t\|_{L^2}^2\\ &\qquad +\|(\dot W_t)^2W_t^2\|_{L^1}+\|(\dot W_t)^2W_t\|_{L^1}\big). \end{aligned}\tag{2.19}$$

We require $c>1$ to be large enough so that $c^{-1}\leq \ell_t\leq c$.

By Lemma 2.12, we have $\|W_t\|_{L^2}^2 < c\,\epsilon^2$ for all $t>0$ such that $\ell_t>\ell-\epsilon^2$. We introduce a large constant $d\geq 1$ that we will fix later. We introduce the set

$$I := \big\{t\in[0,\infty)\ \big|\ \ell_t\geq \ell-\epsilon^2,\ \|\dot W_t\|_{L^2}^2\geq d\|W_t\|_{L^2}^2\big\}.$$

and consider $t\in I$. By means of an integration by parts and Cauchy–Schwarz's inequality, we have

$$\|\dot W_t\|_{L^2}^2 \leq -\int_0^{\ell_t} W_t\ddot W_t\,\mathrm{d}s \leq \|W_t\|_{L^2}\|\ddot W_t\|_{L^2}\leq d^{-1}\|\dot W_t\|_{L^2}\|\ddot W_t\|_{L^2},$$

and thus

$$\|\dot W_t\|_{L^2}\leq d^{-1}\|\ddot W_t\|_{L^2}.$$

We employ this inequality to bound from above the positive terms in (2.19) as follows.

$$\begin{aligned}\|W_t^4\|_{L^1} &\leq \|W_t\|_{L^2}^2\|W_t\|_{L^\infty}^2\leq c\,\epsilon^2\big(\|W_t\|_{L^1}^2+\|\dot W_t\|_{L^1}^2\big)\leq c\,\epsilon^2\|\dot W_t\|_{L^2}^2,\\ \|W_t^6\|_{L^1} &\leq \|W_t\|_{L^2}^2\|W_t\|_{L^\infty}^4\leq c\,\epsilon^2\big(\|W_t\|_{L^1}^4+\|\dot W_t\|_{L^1}^4\big)\\ &\leq c\,\epsilon^2\big(\|W_t\|_{L^1}^4+\|W_t\ddot W_t\|_{L^1}^2\big)\leq c\,\epsilon^2\big(\|W_t\|_{L^1}^4+\|W_t\|_{L^2}^2\|\ddot W_t\|_{L^2}^2\big)\\ &\leq c\,\epsilon^4\big(\|\dot W_t\|_{L^2}^2+\|\ddot W_t\|_{L^2}^2\big),\\ \|(\dot W_t)^2W_t\|_{L^1} &\leq \|W_t\|_{L^1}\|\dot W_t\|_{L^\infty}^2\leq c\|W_t\|_{L^2}\|\ddot W_t\|_{L^2}^2\leq c\,\epsilon\|\ddot W_t\|_{L^2}^2,\\ \|(\dot W_t)^2W_t^2\|_{L^1} &\leq \|W_t\|_{L^2}^2\|\dot W_t\|_{L^\infty}^2\leq c\,\epsilon^2\|\ddot W_t\|_{L^2}^2. \end{aligned}$$

We require $\epsilon>0$ to be small enough so that the negative term $-c^{-1}\|\ddot W_t\|_{L^2}^2$ can absorb the terms $c\,\epsilon\|\ddot W_t\|_{L^2}^2$, $c\,\epsilon^2\|\ddot W_t\|_{L^2}^2$, $c\,\epsilon^4\|\ddot W_t\|_{L^2}^2$, thus obtaining

$$\partial_t\|\dot W_t\|_{L^2}^2\leq -c^{-1}\|\ddot W_t\|_{L^2}^2 + c\|\dot W_t\|_{L^2}^2\leq (-c^{-1}d+c)\|\dot W_t\|_{L^2}^2,\quad \forall t\in I.$$

We now fix $d>c^2$, so that $-b:=(cd^{-1}-c^{-1})<0$ and

$$\partial_t\|\dot W_t\|_{L^2}^2\leq -b\|\dot W_t\|_{L^2}^2,\quad \forall t\in I.\tag{2.20}$$

Now, consider a time value $t\geq b^{-1}\log(\|\dot W_0\|_{L^2}^2\epsilon^{-2})$ such that $\ell_t\geq\ell-\epsilon^2$. If $[0,t]\subset I$, then (2.20) and Gronwall's lemma imply

$$\|\dot W_t\|_{L^2}^2\leq e^{-bt}\|\dot W_0\|_{L^2}^2\leq\epsilon^2.$$

If instead $[0,t]\setminus I\neq\varnothing$, we set $t_0:=\sup([0,t]\setminus I)$, and we have

$$\|\dot W_t\|_{L^2}^2\leq e^{-b(t-t_0)}\|\dot W_{t_0}\|_{L^2}^2\leq\|\dot W_{t_0}\|_{L^2}^2\leq d\|W_{t_0}\|_{L^2}^2\leq dc^2\epsilon^2.$$

□

The analogue of Theorem 1.2(iv) is well known in the classical setting of Lusternik–Schnirelmann theory: it is based on the fact that a smooth function drops of at least a fixed amount along a gradient flow line that crosses a given shell around a critical set at a given level. In the context of the Riemannian curve shortening semi-flow, the analogous property is claimed[1] by Grayson [20]. We now employ the bounds of Lemmas 2.12 and 2.13 to provide a complete proof of Theorem 1.2(iv) in our general reversible Finsler setting.

Proof of Theorem 1.2(iv). Let $\ell > 2\rho_0$ and $\epsilon > 0$ be given. For any embedded circle $\gamma_0 \in \mathrm{Emb}(S^1, M)$, we denote by $\gamma_t = \phi_t(\gamma_0)$ the corresponding evolution under the curve shortening semi-flow, by $\ell_t = L(\gamma_t)$ the length, and by W_t the associated functions as defined in (2.17). We denote by $c \geq 1$ the maximum among the two constants c given by Lemmas 2.12 and 2.13.

Assume now that $\gamma_0 \in \mathrm{Emb}(S^1, M)$ has length $\ell_0 \in (\ell - \delta, \ell + \delta)$ for some constant $\delta \in (0, \ell - 2\rho_0)$ that we will fix later. We claim that there exists $t_0 \in [0, 2]$ such that

$$\ell_{t_0} < \ell - \delta \quad \text{or} \quad \|W_{t_0}\|_{L^2} \leq \sqrt{\delta}.$$

Indeed, if $\|W_t\|_{L^2} > \sqrt{\delta}$ for all $t \in [0, 2]$, we have

$$\ell_2 = \ell_0 - \int_0^2 \|W_t\|_{L^2}^2 \mathrm{d}t < \ell_0 - 2\delta < \ell - \delta.$$

By Lemma 2.12, for each $t \geq t_0$ such that $\ell_t \geq \ell - \delta$, we have

$$\|W_t\|_{L^2} \leq c\sqrt{\delta}.$$

In particular, since $t_0 \leq 2$, this holds for $t = 2$. Therefore, we can apply Lemma 2.13: for each $t \geq 2 + c\log(\|\dot{W}_2\|_{L^2}^2 c^{-2}\delta^{-1})$ such that $\ell_t \geq \ell - \delta$, we have

$$\|\dot{W}_t\|_{L^2} \leq c^2\sqrt{\delta},$$

and in particular

$$\begin{aligned} \|W_t\|_{L^\infty} &\leq \|\dot{W}_t\|_{L^1} + \min|W_t| \leq \|\dot{W}_t\|_{L^1} + \ell_t^{-1}\|W_t\|_{L^1} \\ &\leq \ell_t^{1/2}\|\dot{W}_t\|_{L^2} + \ell_t^{-1/2}\|W_t\|_{L^2} \leq \big((\ell+\delta)^{1/2} + (\ell-\delta)^{-1/2}\big)c^2\sqrt{\delta}; \end{aligned}$$

we now fix $\delta \in (0, \ell - 2\rho_0)$ small enough so that $\epsilon \geq \big((\ell-\delta)^{-1/2} + (\ell+\delta)^{1/2}\big)c^2\sqrt{\delta}$, which together with the previous L^∞ bound implies $\gamma_t \in \mathcal{U}(\ell, \epsilon)$. The desired positive function τ is, therefore, given by

$$\tau : \mathrm{Emb}(S^1, M)^{<\ell+\delta} \to (0, \infty), \quad \tau(\gamma_0) = 2 + c\log(\|\dot{W}_2\|_{L^2}^2 c^{-2}\delta^{-1}).$$

□

[1]Quoted from the last sentence in [20, page 109]: *"Any curve leaving a small neighborhood of a geodesic shortens a fixed amount before moving very far."*

2.6. Compactness

Finally Theorem 1.2(v) will be a consequence of the following compactness result. We denote by $\mathbb{P}\mathrm{T}M \to M$ the projectivized tangent bundle of M, whose fiber over any $x \in M$ is the real projective space

$$\mathbb{P}(\mathrm{T}_x M) = \frac{\mathrm{T}_x M \setminus \{0\}}{\sim},$$

where $v \sim \lambda v$ for all $v \in \mathrm{T}_x M \setminus \{0\}$ and $\lambda \in \mathbb{R}$.

Lemma 2.14. *Let $K \subseteq \mathbb{P}\mathrm{T}M$ be a compact subset. If no element of K is tangent to a simple closed geodesic of (M, F) of length ℓ, then for all $\epsilon > 0$ small enough no element in K is tangent to some curve $\gamma \in \mathcal{U}(\ell, \epsilon)$.*

Proof. If the Lemma does not hold, then there exists a sequence $\gamma_n \in \mathcal{U}(\ell, 1/n)$ such that $F(\gamma_n, \dot\gamma_n) \equiv L(\gamma_n)$ and $[\dot\gamma_n(0)] \in K$. The lifted curves $(\gamma_n, \dot\gamma_n/L(\gamma_n))$ are contained in the Finsler unit tangent bundle $\{(x, v) \in \mathrm{T}M \mid F(x, v) = 1\}$, which is a compact subset of $\mathrm{T}M$. We consider the function V_γ defined in (1.2). Since $\|V_{\gamma_n}\|_{L^\infty} < 1/n \to 0$ as $n \to \infty$, the sequence γ_n is bounded in the C^2-topology. Therefore, up to a subsequence, γ_n converges in the C^1-topology to some γ such that $F(\gamma, \dot\gamma) \equiv L(\gamma) = \ell$ and $\dot\gamma(0) \in K$. Now, consider the Finsler energy

$$E : W^{1,2}(S^1, M) \to [0, \infty), \quad E(\zeta) = \int_{S^1} F(\zeta(u), \dot\zeta(u))^2 \, \mathrm{d}u,$$

Since each γ_n has constant speed, we have

$$\begin{aligned} \mathrm{d}E(\gamma_n)X &= 2 \int_{S^1} F(\gamma_n(u), \dot\gamma_n(u)) \big(F_x(\gamma_n(u), \dot\gamma_n(u)) \\ &\quad - \tfrac{\mathrm{d}}{\mathrm{d}u} F_v(\gamma_n(u), \dot\gamma_n(u)) \big) \, X(u) \, \mathrm{d}u \\ &= 2L(\gamma_n) \, \mathrm{d}L(\gamma_n)X = 2L(\gamma_n) \\ &\quad \int_{S^1} V_{\gamma_n}(u) \, g(N_{\gamma_n}(u), X(u)) \, \|\dot\gamma_n(u)\|_g \, \mathrm{d}u. \end{aligned}$$

This, together with $\|V_{\gamma_n}\|_{L^\infty} \to 0$ and the fact that E is a $C^{1,1}$ function, readily implies that the limit curve γ is a critical point of E, and therefore, a closed geodesic. To reach a contradiction, we simply have to show that γ is simple closed.

On an orientable reversible Finsler surface, a closed geodesic that is the C^1-limit of embedded circles is itself simple. Indeed, γ cannot have a transverse self-intersection, for the same would be true for γ_n for n large enough. Moreover, γ cannot have a self-tangency with opposite orientation, i.e., of the form $\gamma(u_1) = \gamma(u_2)$ and $\dot\gamma(u_1) = -\dot\gamma(u_2)$ for some $u_1 < u_2$; indeed, since F is reversible, we would have $\dot\gamma(u_1 + r) = \dot\gamma(u_2 - r)$ for all $r > 0$, and then $\dot\gamma(\frac{u_1+u_2}{2}) = 0$, contradicting the fact that γ is a geodesic. Finally, γ cannot be an iterated curve, i.e., of the form $\gamma(u) = \zeta(mu)$ for some simple closed geodesic $\zeta : S^1 \to M$ and $m \geq 2$; otherwise, since M is an orientable surface, a tubular neighborhood of ζ would be diffeomorphic to the annulus $S^1 \times (-1, 1)$, ζ being its zero section $S^1 \times \{0\}$; any closed curve sufficiently

C^1-close to γ would wind $m \geq 2$ times around the annulus $S^1 \times (-1,1)$, and therefore, would have self-intersections. $\square$

Proof of Theorem 1.2(v). By choosing $K = \mathbb{P}\mathrm{T}M$ in Lemma 2.14, we infer that, for each $\ell \in [\ell_1, \ell_2]$, there exists $\epsilon > 0$ such that $\mathcal{U}(\ell, \epsilon) = \varnothing$; notice that this readily implies that $\mathcal{U}(\ell', \epsilon/\sqrt{2}) = \varnothing$ for all $\ell' \in [\ell - \frac{1}{2}\epsilon^2, \ell + \frac{1}{2}\epsilon^2]$. Theorem 1.2(v) is a direct consequence of this fact, together with Theorem 1.2(iv) and the compactness of the interval $[\ell_1, \ell_2]$.

3. Existence of simple closed geodesics

3.1. Lusternik–Schnirelmann theory

Let (M, F) be a closed, orientable, reversible, Finsler surface. We consider the space of embedded loops $\mathrm{Emb}(S^1, M)$ and the space of circle diffeomorphisms $\mathrm{Diff}(S^1)$, both endowed with the C^∞ topology. We introduce the space of unparametrized embedded loops

$$\Pi = \frac{\mathrm{Emb}(S^1, M)}{\mathrm{Diff}(S^1)},$$

endowed with the quotient topology. Here, $\mathrm{Diff}(S^1)$ acts by reparametrization on $\mathrm{Emb}(S^1, M)$. From now on, the length functional (1.1) will be considered as a continuous function on Π, i.e.,

$$L : \Pi \to [0, \infty).$$

For any subset $\mathcal{W} \subset \Pi$ and $\ell \in (0, \infty]$, we set

$$\mathcal{W}^{<\ell} := \{\gamma \in \mathcal{W} \mid L(\gamma) < \ell\}. \tag{3.1}$$

Throughout this paper, we shall denote by $H_*(\,\cdot\,; \mathbb{F})$ the singular homology with coefficients in a field $\mathbb{F}$; we shall remove $\mathbb{F}$ from the notation whenever the arguments will not require a specific field. If σ is a singular chain in Π, we denote by $\mathrm{supp}(\sigma)$ its support, which is a compact subset of Π. Each non-zero homology class $h \in H_*(\Pi^{<b}, \Pi^{<a})$, where $0 < a < b \leq \infty$, defines a min-max value

$$\ell(h) := \inf_{[\sigma]=h} \max L|_{\mathrm{supp}(\sigma)} \in [a, b).$$

Such value turns out to be the (positive) length of a simple closed geodesic of (M, F). This will be a rather direct consequence of the existence of the semi-flow of Theorem 1.2 and of the following statement. We will employ the open subsets $\mathcal{U}(\ell, \epsilon) \subset \mathrm{Emb}(S^1, M)$ defined in (1.4), which depend on an auxiliary Riemannian metric g on M. Since such open subsets are invariant under the action of $\mathrm{Diff}(S^1)$, we can consider their quotients

$$\mathcal{W}(\ell, \epsilon) := \frac{\mathcal{U}(\ell, \epsilon)}{\mathrm{Diff}(S^1)},$$

which are open subset in Π.

Lemma 3.1. *For each non-zero $h \in H_*(\Pi^{<b}, \Pi^{<a})$, the associated min-max $\ell = \ell(h)$ is the length of a simple closed geodesic of (M, F). For each $\epsilon > 0$, there exists $\delta \in (0, \epsilon^2)$ such that h can be represented by a relative cycle σ with*

$$\operatorname{supp}(\sigma) \subset \Pi^{<\ell-\delta} \cup \mathcal{W}(\ell, \epsilon). \tag{3.2}$$

Moreover, if there are only finitely many simple closed geodesics with length in $(\ell - \epsilon^2, \ell + \epsilon^2)$, there exists a simple closed geodesic γ of length ℓ such that, if we denote by $\mathcal{V}(\gamma, \epsilon)$ the connected component of $\mathcal{W}(\ell, \epsilon)$ containing γ, the inclusion induces a non-zero homomorphism $H_*(\mathcal{V}(\gamma, \epsilon), \mathcal{V}(\gamma, \epsilon)^{<\ell-\delta}) \to H_*(\Pi, \Pi^{<\ell})$.

Proof. We set $\rho_0 := a/3$, and consider the semi-flow ψ_t of Theorem 1.2. Since, by Theorem 1.2(ii), ψ_t is equivariant with respect to the action of $\mathrm{Diff}(S^1)$, it induces a well-defined continuous semi-flow on the quotient of its domain, which we still denote by $\psi_t : \Pi \to \Pi$. Given $\epsilon > 0$, we consider the associated $\delta \in (0, \epsilon^2)$ provided by Theorem 1.2(iv). By the definition of the min-max value $\ell := \ell(h)$, we can find a relative cycle σ representing h and such that $\max L|_{\operatorname{supp}(\sigma)} < \ell + \delta$. For each $t > 0$, the relative cycle $(\psi_t)_*\sigma$ still represents h. Since $\operatorname{supp}(\sigma)$ is compact, Theorem 1.2(iv) implies that, if we choose $t > 0$ large enough, we have $\operatorname{supp}((\psi_t)_*\sigma) \subset \Pi^{<\ell-\delta} \cup \mathcal{W}(\ell, \epsilon)$. This proves (3.2).

Now, assuming by contradiction that ℓ is not the length of a simple closed geodesic of (M, F), by choosing $K = \mathbb{P}\mathrm{T}M$ in Lemma 2.14 we would have that $\mathcal{W}(\ell, \epsilon) = \varnothing$ for all $\epsilon > 0$ small enough. However, by the result of the previous paragraph, this would allow us to find a relative cycle σ representing h and such that $\operatorname{supp}(\sigma) \subset \Pi^{<\ell-\delta}$, contradicting the definition of $\ell = \ell(h)$.

We are left to prove the moreover part of the statement. For that, notice that we can assume that $\epsilon > 0$ is arbitrarily small (if the theorem holds for some ϵ, it also holds for larger values of ϵ). In particular, we assume that ϵ is small enough so that, by our assumptions, there are only finitely many simple closed geodesics $\gamma_1, \ldots, \gamma_k$ with length in the interval $[\ell - \epsilon^2, \ell + \epsilon^2]$, and they all have length ℓ. We denote by $\mathcal{V}_{\epsilon,i}$ the connected component of $\mathcal{W}_\epsilon := \mathcal{W}(\ell, \epsilon)$ containing γ_i. If needed, we further lower $\epsilon > 0$, so that $\mathcal{V}_{\epsilon,i} \cap \mathcal{V}_{\epsilon,j} = \varnothing$ if $i \neq j$. We set

$$\mathcal{V}_\epsilon := \mathcal{V}_{\epsilon,1} \cup \cdots \cup \mathcal{V}_{\epsilon,k}.$$

We can also lower $\delta > 0$ so that $a \leq \ell - \delta$. The inclusions induce the commutative diagram

$$\begin{tikzcd}
H_*(\Pi^{<\ell-\delta} \cup \mathcal{W}_\epsilon, \Pi^{<a}) \arrow[r, "i_*"] \arrow[d] & H_*(\Pi, \Pi^{<a}) \arrow[d, "j_*"] \\
H_*(\Pi^{<\ell-\delta} \cup \mathcal{W}_\epsilon, \Pi^{<\ell-\delta}) \arrow[r] & H_*(\Pi, \Pi^{<\ell}) \\
H_*(\mathcal{W}_\epsilon, \mathcal{W}_\epsilon^{<\ell-\delta}) \arrow[u, "\cong"] \arrow[ur, "k_*"'] &
\end{tikzcd}$$

The homology class h is contained in the image of i_* according to (3.2), and the lower vertical arrow is an isomorphism by excision. Moreover, by the very definition of the min-max value $\ell = \ell(h)$, we have that $j_*(h) \neq 0$. Overall, this shows that the homomorphism k_* is non-zero.

We set $\mathcal{W}'_\epsilon := \mathcal{W}_\epsilon \setminus (\mathcal{V}_{\epsilon,1} \cup \cdots \cup \mathcal{V}_{\epsilon,k})$, and claim that

$$\epsilon' := \inf \left\{ \|V_\gamma\|_{L^\infty} \;\middle|\; \gamma \in \mathcal{W}'_\epsilon \right\} > 0,$$

where V_γ is the function defined in (1.2). Otherwise, we could find a sequence $\gamma_n \in \mathcal{W}'_\epsilon$ with $\|V_{\gamma_n}\|_{L^\infty} < 1/n$. As in the proof of Lemma 2.14, one can easily show that, up to extracting a subsequence, γ_n converges to a simple closed geodesic γ with length $L(\gamma) \in [\ell - \epsilon^2, \ell + \epsilon^2]$. But this would imply that $V_\gamma \equiv 0$ and $L(\gamma) = \ell$, and thus that $\gamma \in \mathcal{V}_{\epsilon,1} \cup \cdots \cup \mathcal{V}_{\epsilon,k}$, which is impossible since $\gamma_n \in \mathcal{W}'_\epsilon$ for all $n \in \mathbb{N}$.

Notice that $\mathcal{W}_{\epsilon'} \subset \mathcal{V}_\epsilon$, and once again the inclusion induces a non-zero homomorphism $H_*(\mathcal{W}_{\epsilon'}, \mathcal{W}_{\epsilon'}^{<\ell-\delta}) \to H_*(\Pi, \Pi^{<\ell})$, and therefore, a non-zero homomorphism

$$\bigoplus_{i=1}^{k} H_d(\mathcal{V}_{\epsilon,i}, \mathcal{V}_{\epsilon,i}^{<\ell-\delta}) \to H_d(\Pi, \Pi^{<\ell}).$$

We denote by $I_\epsilon \subseteq \{1, \ldots, k\}$ the subset of those i such that the homomorphism $H_d(\mathcal{V}_{\epsilon,i}, \mathcal{V}_{\epsilon,i}^{<\ell-\delta}) \to H_d(\Pi, \Pi^{<\ell})$ is non-zero. Notice that $I_{\epsilon_1} \subseteq I_{\epsilon_2}$ if $0 < \epsilon_1 < \epsilon_2$. Therefore, there exists

$$i \in \bigcap_{\epsilon \in (0, \epsilon_0]} I_\epsilon,$$

and the simple closed geodesic γ_i satisfies the desired properties. $\square$

Assume now to have a homology class $h \in H_{d+i}(\Pi, \Pi^{<\rho})$ and a cohomology class $w \in H^i(\Pi)$ whose cap product $w \frown h \in H_d(\Pi, \Pi^{<\rho})$ is non-zero. Given any relative cycle σ representing h we can produce a relative cycle σ' representing $w \frown h$ and such that $\mathrm{supp}(\sigma') \subset \mathrm{supp}(\sigma)$. This readily implies that

$$\ell(w \frown h) \leq \ell(h).$$

We can now state a version of the classical Lusternik–Schnirelmann theorem for the length functional.

Theorem 3.2. *If $w \frown h \neq 0$ and $\ell(w \frown h) = \ell(h)$, then for every $\epsilon > 0$ we have $w|_{\mathcal{W}(\ell(h),\epsilon)} \neq 0$ in $H^*(\mathcal{W}(\ell(h), \epsilon))$.*

Proof. Let $\epsilon > 0$ be given, and set $\ell := \ell(h)$ and $\mathcal{W} := \mathcal{W}(\ell, \epsilon)$. By Lemma 3.1, h can be represented by a relative cycle σ such that $\mathrm{supp}(\sigma) \subset \mathcal{W} \cup \Pi^{<\ell}$. By applying sufficiently many barycentric subdivisions to the singular simplexes in σ, we can assume that the relative cycle decomposes as $\sigma = \sigma' + \sigma''$, where σ' and σ'' are chains with $\mathrm{supp}(\sigma') \subset \Pi^{<\ell}$ and $\mathrm{supp}(\sigma'') \subset \mathcal{W}$. Let $w \in H^*(\Pi)$ be a cohomology class such that $w|_{\mathcal{W}} = 0$ in $H^*(\mathcal{W})$ and $w \frown h \neq 0$. The cohomology long exact sequence of the pair $\mathcal{W} \subset \Pi$ provides a relative cocycle

μ representing w that vanishes on all singular simplexes contained in $\mathcal{W}$. This implies

$$w \frown h = [\mu \frown (\sigma' + \sigma'')] = [\mu \frown \sigma'].$$

Namely, $w \frown h$ is represented by the relative cycle $\mu \frown \sigma'$ whose support is contained in the sublevel set $\Pi^{<\ell}$, which implies that $\ell(w \frown h) < \ell$.

3.2. Topology of the space of embedded circles on the 2-sphere

Once the results of the previous subsection are established, the proofs of points (i) and (ii) in Theorem 1.3 are analogous to ones of the Riemannian case in [37]. In this subsection, we provide the arguments for the reader's convenience. We will adopt the notation of the previous section, with M equal to the unit sphere $S^2 \subset \mathbb{R}^3$. It will be crucial to consider the singular homology H_* with coefficients in $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$, and therefore, we will specify the coefficients in the notation.

We first recall, from [6,37], some basic information concerning the topology of the space of its unparametrized embedded loops Π. It is convenient to slightly enlarge this space as follows: we denote by Π_0 the space of constant loops on S^2, and set $\overline{\Pi} := \Pi \cup \Pi_0$. We endow $\overline{\Pi}$ with the quotient C^∞-topology as a subspace of $C^\infty(S^1, M)/\mathrm{Diff}(S^1)$. The relevant topology of $\overline{\Pi}$, at least for what concerns the application to Theorem 1.3, is provided by the subspace of round circles. More precisely, let

$$E = \big\{([x], \lambda x) \in \mathbb{RP}^2 \times \mathbb{R}^3 \mid x \in S^2,\ \lambda \in [-1,1]\big\}.$$

Namely, E is the total space of the canonical line bundle $\pi : E \to \mathbb{RP}^2$ with fiber $[-1,1]$. We consider the embedding

$$\iota : E \to \overline{\Pi}, \quad \iota(e) = \gamma_e,$$

where, if $e = ([x], y)$, $\gamma_e \in \overline{\Pi}$ is the (possibly constant) loop in the intersection of S^2 with the affine plane orthogonal to x and passing through y. The fundamental group of this space is

$$\pi_1(E) \cong \pi_1(\mathbb{RP}^2) \cong \mathbb{Z}_2.$$

Its cohomology ring with $\mathbb{Z}_2$ coefficient is given by

$$H^*(E; \mathbb{Z}_2) = \mathbb{Z}_2[u]/(u^3),$$

where u is the generator of $H^1(E; \mathbb{Z}_2) \cong H^1(\mathbb{RP}^2; \mathbb{Z}_2) \cong \mathbb{Z}_2$. Moreover, by the Thom isomorphism theorem,

$$H^*(E, \partial E; \mathbb{Z}_2) \cong H^{*-1}(E; \mathbb{Z}_2) = \langle v, v \smile u, v \smile u^2 \rangle,$$

where $v \in H^1(E, \partial E)$ denotes the Thom class of the bundle $\pi : E \to \mathbb{RP}^2$. Since we work with $\mathbb{Z}_2$ coefficients, the homology is simply the dual of the cohomology, and in particular, there exists $k_3 \in H_3(E, \partial E; \mathbb{Z}_2)$ such that $(v \smile u^2)k_3 = 1$. Therefore, we also have the classes $k_2 := u \frown k_3$ and $k_1 := u \frown k_2$, and overall we have

$$H_*(E, \partial E; \mathbb{Z}_2) = \langle k_1, k_2, k_3 \rangle.$$

Lemma 3.3. *The map ι induces injective homomorphisms*

$$\iota_* : \pi_1(E) \hookrightarrow \pi_1(\overline{\Pi}), \quad \iota_* : H_1(E; \mathbb{Z}_2) \hookrightarrow H_1(\overline{\Pi}; \mathbb{Z}_2),$$

and a surjective homomorphism

$$\iota^* : H^1(\overline{\Pi}; \mathbb{Z}_2) \twoheadrightarrow H^1(E; \mathbb{Z}_2).$$

Proof. We introduce the following double covering map $p : C \to \overline{\Pi}$. The preimage of an embedded circle $\gamma \in \Pi$ is the two-element set $p^{-1}(\gamma) = \{D_1, D_2\}$, where D_1 and D_2 are the connected components of $S^2 \backslash \gamma$; namely, $p^{-1}(\Pi)$ is the space of the interiors of embedded compact disks in S^2. The preimage of a constant $\gamma \in \Pi_0$ is the two-element set $p^{-1}(\gamma) = \{\varnothing_\gamma, S^2 \backslash \gamma\}$; intuitively, $\varnothing_\gamma$ is the "empty filling disk" of the constant γ. The space C, endowed with the obvious topology, makes $p : C \to \overline{\Pi}$ a covering map.

Notice that both E and $\overline{\Pi}$ are path-connected. We fix an arbitrary base-point $e_0 = ([x_0], 0) \in E$ and a corresponding base-point $\gamma_{e_0} := \iota(e_0) \in \Pi$ for the fundamental groups of E and $\overline{\Pi}$, respectively. We define a homomorphism

$$A : \pi_1(\overline{\Pi}) \to \mathbb{Z}_2$$

as follows. For any continuous path $\mu : [0, 1] \to \overline{\Pi}$ such that $\mu(0) = \mu(1) = \gamma_{e_0}$, we consider an arbitrary continuous lift $\widetilde{\mu} : [0, 1] \to C$, i.e., $p \circ \widetilde{\mu} = \mu$. The homomorphism A is defined by

$$A([\mu]) = \begin{cases} 0, & \text{if } \widetilde{\mu}(0) = \widetilde{\mu}(1), \\ 1, & \text{if } \widetilde{\mu}(0) \neq \widetilde{\mu}(1). \end{cases}$$

For the generator $k \in \pi_1(E) \cong \mathbb{Z}_2$, we readily see that $A \circ \iota_*(k) = 1$. Namely, the composition

$$A \circ \iota_* : \pi_1(E) \xrightarrow{\cong} \mathbb{Z}_2$$

is an isomorphism. In particular, ι induces an injective homomorphism

$$\iota_* : \pi_1(E) \hookrightarrow \pi_1(\overline{\Pi}).$$

Since $\mathbb{Z}_2$ is abelian, the commutator subgroup $[\pi_1(\overline{\Pi}), \pi_1(\overline{\Pi})]$ is contained in the kernel of A. This readily implies that ι induces an injective homomorphism

$$\iota_* : H_1(E; \mathbb{Z}_2) \hookrightarrow H_1(\overline{\Pi}; \mathbb{Z}_2).$$

Indeed, we have

$$H_1(E; \mathbb{Z}_2) \cong \pi_1(E, e_0) \cong \mathbb{Z}_2, \quad H_1(\overline{\Pi}; \mathbb{Z}_2) \cong \frac{\pi_1(\overline{\Pi})}{[\pi_1(\overline{\Pi}), \pi_1(\overline{\Pi})]} \otimes \mathbb{Z}_2.$$

Assume by contradiction that $\iota_*(k) = 0$ in $H_1(\overline{\Pi}; \mathbb{Z}_2)$, where k is the generator of $\pi_1(E, e_0) \cong H_1(E; \mathbb{Z}_2)$. This is equivalent to $\iota_*(k) = a * a * b$ in $\pi_1(\overline{\Pi})$ for some $a \in \pi_1(\overline{\Pi})$ and $b \in [\pi_1(\overline{\Pi}), \pi_1(\overline{\Pi})]$; here, $*$ denotes the group multiplication in $\pi_1(\overline{\Pi})$. However, this implies $A \circ \iota_*(k) = 2A(a) + A(b) = 0$, contradicting the fact that $A \circ \iota_*$ is an isomorphism.

Finally, since the homomorphism $\iota_* : H_1(E; \mathbb{Z}_2) \hookrightarrow H_1(\overline{\Pi}; \mathbb{Z}_2)$ is injective, its dual $\iota^* : H^1(\overline{\Pi}; \mathbb{Z}_2) \twoheadrightarrow H^1(E; \mathbb{Z}_2)$ is surjective. □

We fix, once for all, a cohomology class $w \in H^1(\overline{\Pi};\mathbb{Z}_2)$ such that $\iota^* w = u$. Since $\partial E = \Pi_0 \cong S^2$ is simply connected, the long exact sequence of homology groups readily implies that $H_1(E,\partial E;\mathbb{Z}_2) \cong H_1(E;\mathbb{Z}_2)$ and $H_1(\overline{\Pi},\Pi_0;\mathbb{Z}_2) \cong H_1(\overline{\Pi};\mathbb{Z}_2)$. This, together with Lemma 3.3, implies that ι induces an injective homomorphism of relative homology groups

$$\iota_* : H_1(E,\partial E;\mathbb{Z}_2) \hookrightarrow H_1(\overline{\Pi},\Pi_0;\mathbb{Z}_2).$$

Therefore, $\iota_* k_2$ and $\iota_* k_3$ are both non-zero in $H_*(\overline{\Pi},\Pi_0;\mathbb{Z}_2)$, since

$$w^2 \frown \iota_* k_3 = w \frown \iota_*(u \frown k_3) = w \frown \iota_* k_2 = \iota_*(u \frown k_2) = \iota_* k_1 \neq 0.$$

Now, let us get rid of the space of constant curves Π_0. We recall that the systole $\mathrm{sys}(S^2,F)$ is the length of the shortest closed geodesic of (S^2,F).

Lemma 3.4. *For all $\rho \in (0,\mathrm{sys}(S^2,F))$, the inclusion $\Pi_0 \subset \overline{\Pi}^{<\rho}$ is a homotopy equivalence. Therefore, the inclusions induce the homology isomorphisms*

$$H_*(\overline{\Pi},\Pi_0;\mathbb{Z}_2) \xrightarrow[\cong]{j_*} H_*(\overline{\Pi},\overline{\Pi}^{<\rho};\mathbb{Z}_2) \xleftarrow[\cong]{l_*} H_*(\Pi,\Pi^{<\rho};\mathbb{Z}_2).$$

Proof. We fix $\rho = \rho_2 \in (0,\mathrm{sys}(S^2,F))$. We claim that, for any $\rho_1 \in (0,\rho_2)$, the inclusion $\overline{\Pi}^{<\rho_1} \subset \overline{\Pi}^{<\rho_2}$ is a homotopy equivalence. Indeed, since there are no simple closed geodesics of (S^2,F) with length in $[\rho_1,\rho_2]$, Theorem 1.2(v) with choice of parameter $\rho_0 \in (0,\rho_1/2)$ implies that there exists a continuous function $\tau : \overline{\Pi}^{<\rho_2} \to (0,\infty)$ and a continuous map

$$\kappa : \overline{\Pi}^{<\rho_2} \to \overline{\Pi}^{<\rho_1}, \quad \kappa(\gamma) := \psi_{\tau(\gamma)}(\gamma).$$

The map κ is a homotopy inverse of the inclusion $\overline{\Pi}^{<\rho_1} \subset \overline{\Pi}^{<\rho_2}$.

We consider S^2 as the unit sphere in $\mathbb{R}^3$, equipped with its round metric g_0 induced by the Euclidean metric of $\mathbb{R}^3$. We denote by $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^3$, and by $\exp_x$ the exponential maps of (S^2,g_0). We fix a sufficiently large constant $a \geq 1$ so that

$$a^{-1}\|v\| \leq F(x,v) \leq a\|v\|, \quad \forall (x,v) \in \mathrm{T}S^2.$$

If $\rho_1 > 0$ is small enough, each $\gamma \in \overline{\Pi}^{<\rho_2}$ has average outside the origin; namely, if we parametrize γ with constant speed $F(\gamma,\dot\gamma) \equiv L(\gamma)$, we have

$$\int_{S^1} \gamma(u)\,\mathrm{d}u \neq 0.$$

We denote by $\pi : \mathbb{R}^3 \setminus \{0\} \to S^2$ the radial projection $\pi(x) = x/\|x\|$, and set

$$\widehat{\gamma} := \pi\left(\int_{S^1} \gamma(u)\,\mathrm{d}u\right) \in S^2.$$

Notice that

$$\max_{u \in S^1} \|\gamma(u) - \widehat{\gamma}\| \leq aL(\gamma) < a\rho_1. \tag{3.3}$$

For each $x \in S^2$, we set

$$B_x := \{ v \in T_x S^2 \mid \|v\| < \pi \}, \quad U_x := \exp_x(B_x) \subset S^2.$$

From now on we will restrict the exponential map of (S^2, g_0) as a diffeomorphism of the form $\exp_x : B_x \to U_x$. We require $\rho_1 < \pi/a$ so that, by (3.3), we have $\gamma \subset U_{\widehat{\gamma}}$ whenever $L(\gamma) < \rho_1$.

We define the continuous homotopy

$$r_t : \overline{\Pi}^{<\rho_1} \to \overline{\Pi}, \quad r_t(\gamma)(u) = \exp_{\widehat{\gamma}}\big((1-t)\exp_{\widehat{\gamma}}^{-1}(\gamma(u))\big).$$

Notice that the time-1 map is a retraction

$$r_1 : \overline{\Pi}^{<\rho_1} \to \Pi_0.$$

Moreover, if $b > 0$ is a constant larger than $\|\mathrm{d}\exp_x(v)\|$ and $\|\mathrm{d}\exp_x^{-1}(y)\|$ for all $x \in S^2$, $v \in B_x$, and $y \in U_x$, we have

$$L(r_t(\gamma)) \leq a^2 b^2 \int_{S^1} \|\dot{\gamma}(u)\|\,\mathrm{d}u \leq a^2 b^2 L(\gamma) < a^2 b^2 \rho_1.$$

We require $\rho_1 < \rho_2 a^{-2} b^{-2}$, so that every r_t is a map of the form

$$r_t : \overline{\Pi}^{<\rho_1} \to \overline{\Pi}^{<\rho_2}.$$

Overall, this implies that the inclusion $\Pi_0 \subset \overline{\Pi}^{<\rho}$ is a homotopy equivalence. The homology long exact sequence of the triple $\Pi_0 \subset \overline{\Pi}^{<\rho} \subset \overline{\Pi}$ readily implies that j_* is an isomorphism. Finally, the excision property implies that l_* is an isomorphism as well. $\square$

We consider the isomorphisms j_* and l_* provided by Lemma 3.4, and define the non-zero relative homology classes

$$h_i := l_*^{-1} j_* \iota_* k_i \in H_*(\Pi, \Pi^{<\rho}; \mathbb{Z}_2), \quad i = 1, 2, 3. \tag{3.4}$$

Notice that

$$w|_\Pi \frown h_{i+1} = h_i.$$

We denote by $E_0 \subset E$ the zero section of the line bundle $\pi : E \to \mathbb{RP}^2$. Notice that ι restricts as a map of the form $\iota_0 := \iota|_{E_0} : E_0 \to \Pi$.

Lemma 3.5. *For each $z \in H^2(\Pi; \mathbb{Z}_2)$ such that $\iota_0^* z \neq 0$ in $H^2(E_0; \mathbb{Z}_2)$, we have*

$$z \frown h_3 = h_1.$$

Proof. For each $r \in [0, 1]$, we consider the subset

$$E_r = \big\{([x], \lambda x) \in E \mid \lambda \in [-r, r]\big\}. \subset E,$$

Notice that this notation agrees with the definition of E_0 as the zero section of the line bundle $\pi : E \to \mathbb{RP}^2$. We fix $r \in (0, 1)$ sufficiently close to 1 so that $\iota(\partial E_r) \subset \Pi^{<\rho}$. By deformation and excision, we have that the inclusions induce isomorphisms

$$\begin{aligned} &H_*(E, \partial E; \mathbb{Z}_2) \xrightarrow{\cong} H_*(E, E \backslash E_0; \mathbb{Z}_2) \\ &\xleftarrow{\cong} H_*(E_r, E_r \backslash E_0; \mathbb{Z}_2) \xleftarrow{\cong} H_*(E_r, \partial E_r; \mathbb{Z}_2). \end{aligned}$$

We denote by $k_i' \in H_*(E_r, \partial E_r; \mathbb{Z}_2)$ the image of $k_i \in H_*(E, \partial E; \mathbb{Z}_2)$ under the composition of the above isomorphisms. Notice that $k_1' = u|_{E_r} \frown k_3'$. The restriction $\iota_r = \iota|_{E_r} : E_r \to \Pi$ induces a homomorphism

$$(\iota_r)_* : H_*(E_r, \partial E_r; \mathbb{Z}_2) \to H_*(\Pi, \Pi^{<\rho}; \mathbb{Z}_2),$$

which allows to express the homology classes (3.4) as $h_i = (\iota_r)_* k_i'$. Since the inclusion $E_0 \subset E_r$ is a homotopy equivalence, a cohomology class $z \in H^2(\Pi; \mathbb{Z}_2)$ satisfies $\iota_0^* z \neq 0$ if and only if $\iota_r^* z \neq 0$, and thus if and only if $\iota_r^* z \frown k_3' = k_1'$. Therefore, if this is the case, we have

$$z \frown h_3 = z \frown (\iota_r)_* k_3' = (\iota_r)_*(\iota_r^* z \frown k_3') = (\iota_r)_* k_1' = h_1.$$

□

We set

$$\ell_i := \ell(h_i), \quad i = 1, 2, 3. \tag{3.5}$$

Since every ℓ_i is the length of a simple closed geodesic of (S^2, F), if the simple length spectrum $\sigma_{\mathrm{s}}(S^2, F)$ is a singleton we have $\ell_1 = \ell_2 = \ell_3$. In this case Theorem 1.3(i) is a consequence of the following statement.

Theorem 3.6. *If $\ell_1 = \ell_2 = \ell_3$, then (S^2, F) is simple Zoll.*

Proof. We consider a circle bundle $\mathrm{pr} : P \to \Pi$, whose total space is given by $P = \{(\gamma, x) \in \Pi \times S^2 \mid x \in \gamma\}$ and whose projection is $\mathrm{pr}(\gamma, x) = \gamma$. We consider the projectivized tangent bundle

$$\mathbb{P}\mathrm{T}S^2 = \big\{ V_x \bigm| x \in S^2,\ V_x \text{ 1-dimensional vector subspace of } \mathrm{T}_x S^2 \big\},$$

and define the continuous evaluation map $\mathrm{ev} : P \to \mathbb{P}\mathrm{T}S^2$, $\mathrm{ev}(\gamma, x) = \mathrm{T}_x\gamma$. Since $\mathbb{P}\mathrm{T}S^2$ is a closed 3-manifold, we have $H^3(\mathbb{P}\mathrm{T}S^2; \mathbb{Z}_2) \cong \mathbb{Z}_2$, and we denote by m a generator of $H^3(\mathbb{P}\mathrm{T}S^2; \mathbb{Z}_2)$. We consider the pull-back bundle

$$P_0 = \iota_0^* P = \big\{ (e, p) \in E_0 \times P \mid \iota_0(e) = \mathrm{pr}(p) \big\},$$

and the commutative diagram

$$\begin{tikzcd}
P_0 \arrow[r, hook, "\tilde\iota_0"] \arrow[d, "\mathrm{pr}|_{P_0}"] & P \arrow[r, "\mathrm{ev}"] \arrow[d, "\mathrm{pr}"] & \mathbb{P}\mathrm{T}S^2 \\
E_0 \arrow[r, hook, "\iota_0"] & \Pi &
\end{tikzcd}$$

Here, $\tilde\iota_0(e, p) = p$ is the projection onto the second factor. Notice that $\mathrm{ev} \circ \tilde\iota_0$ is a homeomorphism. Moreover, since $H^3(E_0; \mathbb{Z}_2)$ and $H^4(E_0; \mathbb{Z}_2)$ are trivial, the Gysin sequence of the pull-back bundle $\mathrm{pr}|_{P_0} : P_0 \to E_0$ readily implies that

$$(\mathrm{pr}|_{P_0})_* : H^3(P_0; \mathbb{Z}_2) \to H^2(E_0; \mathbb{Z}_2)$$

is an isomorphism. This implies that $(\mathrm{pr}|_{P_0})_* \tilde\iota_0^* \mathrm{ev}^* m \neq 0$ in $H^2(E_0; \mathbb{Z}_2)$. We set

$$z := \mathrm{pr}_* \mathrm{ev}^* m \in H^2(\Pi; \mathbb{Z}_2).$$

Since $\iota_0^* z = (\mathrm{pr}|_{P_0})_* \tilde\iota_0^* \mathrm{ev}^* m \neq 0$, Lemma 3.5 implies that $h_1 = z \frown h_3$.

Now, assume by contradiction that $\ell_1 = \ell_2 = \ell_3 =: \ell$, but there exists $(x, v) \in SS^2$ such that the geodesic $\gamma_{x,v}(t) := \exp_x(tv)$ is not a simple closed geodesic of minimal period ℓ (namely, $\gamma_{x,v}$ is not a closed geodesic, or it is closed but not simple closed, or it is simple closed but its length is not ℓ). By Lemma 2.14, there exists $\epsilon > 0$ small enough so that v is not tangent to any curve $\gamma \in \mathcal{W} := \mathcal{W}(\ell, \epsilon)$ passing through x. Namely, if we set $P' := \mathrm{pr}^{-1}(\mathcal{W})$, the restriction $\mathrm{ev}|_{P'} : P' \to \mathbb{P}\mathrm{T}S^2$ is not surjective. Since $\ell_1 = \ell_3$ and $h_1 = z \frown h_3$, Theorem 3.2 implies that $z|_{\mathcal{W}} \neq 0$ in $H^2(\mathcal{W}; \mathbb{Z}_2)$. However, since $z|_{\mathcal{W}} = (\mathrm{pr}|_{P'})_* \mathrm{ev}|^*_{P'} m$, this implies that the homomorphism

$$\mathrm{ev}|^*_{P'} : H^3(\mathbb{P}\mathrm{T}S^2; \mathbb{Z}_2) \to H^3(P'; \mathbb{Z}_2)$$

is non-zero, which is impossible since $\mathrm{ev}|_{P'}$ is not surjective. □

If the simple length spectrum $\sigma_{\mathrm{s}}(S^2, F)$ contains exactly two elements, we must have $\ell_1 = \ell_2$ or $\ell_2 = \ell_3$. In this case Theorem 1.3(i) is a consequence of the following statement.

Theorem 3.7. *If $\ell_i = \ell_{i+1}$ for some $i \in \{1, 2\}$, then every point of S^2 lies on a simple closed geodesic of (S^2, F) of length ℓ_i.*

Proof. Assume by contradiction that $\ell := \ell_i = \ell_{i+1}$, but that some point $x \in S^2$ does not lie on a simple closed geodesic of length ℓ. We consider the subset $\mathcal{U} = \{\gamma \in \overline{\Pi} \mid x \notin \gamma\}$. It is easy to see that $\mathcal{U}$ is contractible: if we denote by $B^2 \subset \mathbb{R}^2$ the unit open ball, and we consider a homeomorphism $\theta : S^2 \setminus \{x\} \to B^2$, the homotopy $r_t : \mathcal{U} \to \mathcal{U}$, $t \in [0, 1]$, given by

$$r_t(\gamma) = \theta^{-1}((1 - t)\theta(\gamma))$$

defines a contraction of $\mathcal{U}$ onto a point curve in $\Pi_0 \cap \mathcal{U}$. In particular $H^1(\mathcal{U}; \mathbb{Z}_2)$ is trivial.

By applying Lemma 2.14 with $K = \mathbb{P}(\mathrm{T}_x S^2)$, we infer that there exists $\mathcal{W} = \mathcal{W}(\ell, \epsilon)$, for $\epsilon > 0$ small enough, such that none of the curves $\gamma \in \mathcal{W}$ passes through x. Since $h_i = w|_{\Pi} \frown h_{i+1}$ and $\ell_i = \ell_{i+1}$, Theorem 3.2 implies that $w|_{\mathcal{W}} \neq 0$ in $H^1(\mathcal{W}; \mathbb{Z}_2)$. However, since $\mathcal{W} \subset \mathcal{U}$, we have $w|_{\mathcal{U}} \neq 0$ in $H^1(\mathcal{U}; \mathbb{Z}_2)$ as well, contradicting the conclusion of the previous paragraph.

4. Critical point theory of the energy functional

In this section, we shall recall the background on the variational theory of Finsler closed geodesics. The reader can find more details and proofs in [1, 10, 14, 41] and references therein. Throughout the section, we shall consider a closed Finsler manifold (M, F) of arbitrary dimension, except in certain statements where we will assume M to be a surface. The Finsler metric F is not required to be reversible, unless specifically stated.

4.1. The energy functional

We denote by $\Lambda = W^{1,2}(S^1, M)$ the free loop space of M of regularity $W^{1,2}$, and consider the energy functional

$$E : \Lambda \to [0, \infty), \quad E(\gamma) = \int_{S^1} F(\gamma(u), \dot{\gamma}(u))^2 \, du.$$

Unlike in the Riemannian case, in the Finsler setting E is $C^{1,1}$, but possibly not C^2. Its critical points with positive critical value are precisely those $\gamma \in \Lambda$ that are closed geodesics of (M, F) parametrized with constant speed $F(\gamma, \dot\gamma) \equiv E(\gamma)^{1/2}$. For each $\gamma \in \Lambda$, we denote by

$$\gamma^m \in \Lambda, \quad \gamma^m(t) = \gamma(mt)$$

its m-th iterate, whose energy is $E(\gamma^m) = m^2 E(\gamma)$. Clearly, iterates of critical points of E are again critical points. Identifying different iterates of the same closed geodesic detected with global variational methods is the crux of the matter in the closed geodesics problem.

The C^1 regularity of E is actually enough to define a smooth pseudo-gradient flow of E on Λ. It is well known that E satisfies the Palais–Smale condition with respect to a suitable complete Riemannian metric on Λ, and therefore, we can perform the usual deformations of critical point theory. Since E is even $C^{1,1}$, it has a well-defined Gateaux Hessian $\mathrm{d}^2E(\gamma)$ at every critical point. However, the C^2 regularity would be needed to apply the classical Morse–Gromoll–Meyer lemma [18]. A simple way to circumvent the potential lack of C^2 regularity and, at the same time, work in a finite dimensional setting consists in employing Morse's finite dimensional approximations of Λ. We consider the (non-symmetric) Finsler distance

$$d : M \times M \to [0, \infty), \quad d(x, y) = \min_{\gamma} \int_0^1 F(\gamma(u), \dot\gamma(u))\,\mathrm{d}u, \tag{4.1}$$

where the minimum ranges over all absolutely continuous curves $\gamma : [0, 1] \to M$ joining x and y. For each integer $k \geq 2$, we consider the space

$$\Lambda_k = \left\{ \boldsymbol{x} = (x_0, \dots, x_{k-1}) \in M^{\times k} \;\middle|\; \sum_{i \in \mathbb{Z}_k} d(x_i, x_{i+1})^2 < \mathrm{injrad}(M, F)^2 \quad \forall i \in \mathbb{Z}_k \right\}.$$

There is a smooth embedding $\iota : \Lambda_k \hookrightarrow \Lambda$ defined as follows: every $\boldsymbol{x} \in \Lambda_k$ is mapped to the curve $\gamma_{\boldsymbol{x}} := \iota(\boldsymbol{x}) \in \Lambda$ such that each restriction $\gamma_{\boldsymbol{x}}|_{[i/k,(i+1)/k]}$, for $i \in \mathbb{Z}_k$, is the shortest geodesic parametrized with constant speed joining x_i and x_{i+1}. In the following, we will identify Λ_k with it image $\iota(\Lambda_k) \subset \Lambda$, and indistinctively write $\boldsymbol{x}$ or $\gamma_{\boldsymbol{x}}$ for the same object. The restriction of the energy to Λ_k has the form

$$E_k = E|_{\Lambda_k} : \Lambda_k \to \big[0, k\,\mathrm{injrad}(M, F)^2\big), \quad E_k(\boldsymbol{x}) = k \sum_{i \in \mathbb{Z}_k} d(x_i, x_{i+1})^2.$$

Since the distance d is smooth away from the diagonal, E_k is smooth on the subspace of those $\boldsymbol{x}$ with $x_i \neq x_{i+1}$ for all $i \in \mathbb{Z}_k$. The critical points of E_k are precisely those $\boldsymbol{x}$ such that $\gamma_{\boldsymbol{x}}$ is a closed geodesic of (M, F) parametrized with constant speed and having energy $E_k(\boldsymbol{x}) = E(\gamma_{\boldsymbol{x}}) < k\,\mathrm{injrad}(M, F)^2$. In particular E_k is smooth on a sufficiently small neighborhood of its critical points with positive energy. For each compact interval $[a, b] \subset \big(-\infty, k\,\mathrm{injrad}(M, F)^2\big)$, the preimage $E_k^{-1}[a, b]$ is compact, which allows us to apply the gradient flow deformations from critical point theory. For each $a > 0$, up to choosing $k \in \mathbb{N}$ large enough, the inclusion of the energy sublevel sets $E_k^{-1}(-\infty, a) \hookrightarrow E^{-1}(-\infty, a)$ admits the homotopy inverse

$$r : E^{-1}(-\infty, a) \to E_k^{-1}(-\infty, a), \quad r(\gamma) = (\gamma(0), \gamma(\tfrac{1}{k}), \ldots, \gamma(\tfrac{k-1}{k})).$$

4.2. The Morse index and nullity

Let $h : V \times V \to \mathbb{R}$ be a symmetric bilinear form on a vector space V. Its index $\mathrm{ind}(h)$ is defined as the supremum of the dimension of the subspaces $W \subset V$ such that $h|_W$ is negative definite. Its nullity $\mathrm{nul}(h)$ is defined as the dimension of $\ker(h) = \{v \in V \mid h(v, \cdot) = 0\}$. Notice that the sum $\mathrm{ind}(h) + \mathrm{nul}(h)$ is the supremum of the dimension of the subspaces $Z \subset V$ such that $h|_Z$ is negative semi-definite.

Let us consider a closed geodesic $\gamma \in \mathrm{crit}(E) \cap E^{-1}(0, \infty)$, and the associated Gateaux Hessian $h := \mathrm{d}^2 E(\gamma)$. The Morse index and nullity of γ are defined by

$$\mathrm{ind}(\gamma) := \mathrm{ind}(h), \quad \mathrm{nul}(\gamma) := \mathrm{nul}(h) - 1.$$

It is well known that the indices are always finite. The reason for the -1 appearing in the definition of the nullity is that $\mathrm{nul}(h)$ is always larger than or equal to 1, as the vector field $\dot\gamma$ belongs to $\ker(h)$. If $x_0 := \gamma(0)$, we denote by

$$\Omega := \{\zeta \in \Lambda \mid \zeta(0) = x_0\}$$

the space of loops based at x_0. The critical points of $E|_\Omega$ are the geodesic loops, that is, those $\zeta \in \Lambda$ whose restriction $\zeta|_{(0,1)}$ is a geodesic parametrized with constant speed. The Morse index and nullity of γ in the based loop space are defined as

$$\mathrm{ind}_\Omega(\gamma) := \mathrm{ind}(h|_{\mathrm{T}_\gamma\Omega}), \quad \mathrm{nul}_\Omega(\gamma) := \mathrm{nul}(h|_{\mathrm{T}_\gamma\Omega}).$$

The behavior of the Morse indices under iteration of the closed geodesic has been thoroughly studied since the seminal work of Bott [11]. Without invoking Bott's theory, one has the following properties, which are rather immediate or can be proved as an exercise.

Lemma 4.1. *Let (M, F) be a Finsler manifold with a closed geodesic $\gamma \in \mathrm{crit}(E)$. The Morse indices of γ satisfy the following properties.*

(i) $\mathrm{ind}(\gamma) \geq \mathrm{ind}_\Omega(\gamma)$.
(ii) $\mathrm{ind}(\gamma) + \mathrm{nul}(\gamma) \geq \mathrm{ind}_\Omega(\gamma) + \mathrm{nul}_\Omega(\gamma)$.
(iii) *If* $\mathrm{ind}(\gamma) > 0$, *then* $\mathrm{ind}(\gamma^m) \to \infty$ *as* $m \to \infty$.
(iv) $\mathrm{ind}(\gamma^m) \geq \mathrm{ind}(\gamma)$ *and* $\mathrm{nul}(\gamma^m) \geq \mathrm{nul}(\gamma)$ *for all* $m \in \mathbb{N}$.
(v) $\mathrm{ind}_\Omega(\gamma^m) \geq m\, \mathrm{ind}_\Omega(\gamma)$ *and* $\mathrm{ind}_\Omega(\gamma^m) + \mathrm{nul}_\Omega(\gamma^m) \geq m\, (\mathrm{ind}_\Omega(\gamma) + \mathrm{nul}_\Omega(\gamma))$ *for all* $m \in \mathbb{N}$.

□

The following proposition summarizes those subtler results concerning the Morse indices of closed geodesics that we will need in the proof of Theorem 1.5. In the literature, most of these results are proved in the Riemannian setting: points (i–iv) can be found in [12], point (vi) in [28], and point (vii) in

[8]. In the Finsler setting, the differences in the proofs are essentially cosmetic, but we include them for the reader's convenience.

Proposition 4.2. *Let* (M, F) *be an orientable Finsler manifold, and* $\gamma \in \mathrm{crit}(E) \cap E^{-1}(0, \infty)$ *a closed geodesic. The indices of* γ *satisfy the following properties.*

(i) $\mathrm{nul}(\gamma) \leq 2\dim(M) - 2$.
(ii) $\mathrm{nul}_\Omega(\gamma) \leq \dim(M) - 1$.
(iii) $\mathrm{ind}(\gamma) \leq \mathrm{ind}_\Omega(\gamma) + \dim(M) - 1$.
(iv) $\mathrm{ind}(\gamma) + \mathrm{nul}(\gamma) \leq \mathrm{ind}_\Omega(\gamma) + \mathrm{nul}_\Omega(\gamma) + \dim(M) - 1$.

Moreover, if M *is an orientable surface, they further satisfy the following properties.*

(v) *If* $\mathrm{nul}_\Omega(\gamma) = 1$ *then* $\mathrm{ind}_\Omega(\gamma^m) = m\,\mathrm{ind}_\Omega(\gamma) + m - 1$ *and* $\mathrm{nul}_\Omega(\gamma^m) = \mathrm{nul}_\Omega(\gamma)$ *for all* $m \in \mathbb{N}$,
(vi) *If* $\mathrm{nul}(\gamma) = 2$ *then* $\mathrm{nul}(\gamma^m) = 2$ *and* $\mathrm{ind}(\gamma^m)$ *is odd for all* $m \in \mathbb{N}$.
(vii) *If* $\mathrm{ind}_\Omega(\gamma^m) > 0$ *for some* $m \geq 1$, *then* $\mathrm{ind}(\gamma) > 0$ *and indeed there exists a nowhere-vanishing smooth vector field* ζ *along* γ *that is* 1*-periodic, everywhere transverse to* $\dot\gamma$, *and such that* $\mathrm{d}^2E(\gamma)(\zeta, \zeta) < 0$.

Proof. We can assume without loss of generality that $E(\gamma) = 1$, so that $F(\gamma, \dot\gamma) \equiv 1$. We set

$$G : \mathrm{T}M \to [0, \infty), \quad G(x, v) = \tfrac{1}{2}F(x, v)^2,$$

which is a $C^{1,1}$ function, smooth outside the zero section, and fiberwise positively homogeneous of degree 2. The function G defines a 1-form λ on $\mathrm{T}M$ by

$$\lambda_{(x,v)}(w) = G_v(x, v) \circ \mathrm{d}\pi(x, v)w, \quad \forall w \in \mathrm{T}_{(x,v)}(TM).$$

The 2-form $-\mathrm{d}\lambda$ is a symplectic form on the complement of the zero section in $\mathrm{T}M$. We treat G as a Hamiltonian, and consider its associated Hamiltonian vector field X defined by $-\mathrm{d}\lambda(X, \cdot) = \mathrm{d}G$. We denote by $\phi_t : \mathrm{T}M \to \mathrm{T}M$ the associated Hamiltonian flow of X. Its flow lines are the speed vectors of the geodesics of (M, F) parametrized with constant speed. In particular, the curve $\tilde\gamma(t) := (\gamma(t), \dot\gamma(t))$ is the periodic orbit of ϕ_t corresponding to the closed geodesic γ. Since G is autonomous, the Hamiltonian flow ϕ_t preserves each level set $G^{-1}(\ell^2)$. The energy level of $\tilde\gamma$ is

$$G(\tilde\gamma(t)) = \tfrac{1}{2}F(\tilde\gamma(t))^2 = 1/2,$$

and we denote by $SM := G^{-1}(1/2) = F^{-1}(1)$ the corresponding energy hypersurface, which is the unit tangent bundle of (M, F). The 1-form λ restricts to a contact form $\alpha := \lambda|_{SM}$, and X restricts to the Reeb vector field of (SM, α). Namely $\alpha(X) = 1$ and $\mathrm{d}\alpha(X, \cdot) = 0$. In particular $\phi_t^*\alpha = \alpha$. We denote by

$$\xi := \ker(\alpha) \subset \mathrm{T}(SM)$$

the contact distribution of α. Notice that

$$\mathrm{d}\pi(\tilde\gamma(t))\xi_{\tilde\gamma(t)} = \ker(G_v(\tilde\gamma(t))).$$

Let L be the vector field on TM defined by

$$L(x,v) = \tfrac{\mathrm{d}}{\mathrm{d}t}\big|_{t=1}(x,tv).$$

This is a Liouville vector field, meaning that $\mathrm{d}\lambda(L,\cdot) = \lambda$, and is transverse to SM. Over SM, the vector bundle $\mathrm{T}(\mathrm{T}M)$ splits as a direct sum

$$\mathrm{T}(\mathrm{T}M)|_{SM} = \mathrm{span}\{X,L\} \oplus \xi, \tag{4.2}$$

and this decomposition is symplectically orthogonal, i.e.,

$$\mathrm{d}\lambda(V,W) = 0, \quad \forall V \in \mathrm{span}\{X,L\},\ W \in \xi.$$

We recall that a Jacobi vector field $\zeta : \mathbb{R} \to \gamma^*\mathrm{T}M$ is a solution of the Jacobi equation, which is the linearization of the Hamiltonian equation $\dot{\tilde{\gamma}} = X \circ \tilde{\gamma}$ at $\tilde{\gamma}$. In local coordinates, the Jacobi equation reads

$$\partial_t(G_{vv}\,\dot{\zeta} + G_{xv}\,\zeta) - G_{xx}\,\zeta - G_{vx}\,\dot{\zeta} = 0.$$

Here and in the following, the second derivatives G_{xx}, G_{xv}, G_{vx}, G_{vv} are meant to be evaluated at $\tilde{\gamma}(t)$. We denote by $\Phi_t := \mathrm{d}\phi_t(\tilde{\gamma}(0)) : \mathrm{T}_{\tilde{\gamma}(0)}\mathrm{T}M \to \mathrm{T}_{\tilde{\gamma}(t)}\mathrm{T}M$ the linearized Hamiltonian flow along $\tilde{\gamma}$. Its flow lines are lifts of Jacobi vector fields ζ, that is, in local coordinates they can be written as

$$\tilde{\zeta}(t) = \Phi_t(\tilde{\zeta}(0)) = (\zeta(t), \dot{\zeta}(t)).$$

The linearized flow Φ_t preserves the splitting (4.2). Indeed, $\phi_t^*\alpha = \alpha$ implies that $\Phi_t(\xi) = \xi$. Moreover, $\Phi_t(X) = X$ and $\Phi_t(L) = tX + L$, that is, $\Phi_t|_{\mathrm{span}\{X,L\}}$ can be written in the frame X, L as the symplectic matrix

$$\Phi_t|_{\mathrm{span}\{X,L\}} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \in \mathrm{Sp}(2). \tag{4.3}$$

Let $W^{1,2}(S^1,\gamma^*\mathrm{T}M)$ be the Hilbert space of 1-periodic $W^{1,2}$-vector fields along γ. The Hessian $h := \mathrm{d}^2E(\gamma)$ is the symmetric bilinear form on $W^{1,2}(S^1,\gamma^*\mathrm{T}M)$ given by

$$h(\zeta,\eta) = 2\int_{S^1}\Big(\langle G_{xx}\,\zeta,\eta\rangle + \langle G_{vx}\,\dot{\zeta},\eta\rangle + \langle G_{xv}\,\zeta,\dot{\eta}\rangle + \langle G_{vv}\,\dot{\zeta},\dot{\eta}\rangle\Big)\mathrm{d}t. \tag{4.4}$$

In this expression, we adopt a common abuse of notation: we write the integrand in local coordinates (this can be made precise by splitting the domain of integration S^1 as a finite union of intervals over which the local coordinates are available). A bootstrap argument, together with an integration by parts, implies that the kernel of h is precisely given by the 1-periodic Jacobi vector fields. In particular $\mathrm{nul}(h) = \dim\ker(\Phi_1 - I)$, and therefore, $\mathrm{nul}(\gamma) = \mathrm{nul}(h) - 1 = \dim\ker(\Phi_1|_{\xi_{\tilde{\gamma}(0)}} - I)$, which implies point (i).

We consider the subspace

$$Z = \big\{\zeta \in W^{1,2}(S^1,\gamma^*\mathrm{T}M) \ \big|\ G_v(\gamma,\dot{\gamma})\zeta \equiv 0\big\}.$$

We claim that

$$\mathrm{ind}(h|_Z) = \mathrm{ind}(h), \quad \mathrm{nul}(h|_Z) = \mathrm{nul}(h) - 1.$$

Indeed, a straightforward computation shows that the h-orthogonal

$$Z^h := \big\{\zeta \in W^{1,2}(S^1,\gamma^*\mathrm{T}M) \ \big|\ h(\zeta,\cdot)|_Z = 0\big\}$$

is precisely the space of those $\zeta \in W^{1,2}(S^1, \gamma^*\mathrm{T}M)$ of the form $\zeta(t) = f(t)\dot\gamma(t)$ for some $f : S^1 \to \mathbb{R}$, and we have $W^{1,2}(S^1, \gamma^*\mathrm{T}M) = Z \oplus Z^h$, $\mathrm{ind}(h|_{Z^h}) = 0$, $\ker(h|_{Z^h}) = \mathrm{span}_{\mathbb{R}}\{\dot\gamma\}$, and $\mathrm{nul}(h|_{Z^h}) = 1$. From now on, we will simply write h for the restriction $h|_Z$, so that in particular

$$\mathrm{nul}(\gamma) = \mathrm{nul}(h) = \dim\ker(\Phi_1|_{\xi_{\tilde\gamma(0)}} - I).$$

Analogously, if we set

$$Z_0 = \big\{\zeta \in Z \;\big|\; \zeta(0) = \zeta(1) = 0\big\},$$

we have

$$\mathrm{ind}_\Omega(\gamma) = \mathrm{ind}(h|_{Z_0}), \quad \mathrm{nul}_\Omega(\gamma) = \mathrm{nul}(h|_{Z_0}).$$

The kernel of $h|_{Z_0}$ is the space of Jacobi fields ζ such that $\zeta(0) = \zeta(1) = 0$ and $G_v(\gamma, \dot\gamma)\zeta \equiv 0$, and thus

$$\mathrm{nul}(h|_{Z_0}) \le \dim\ker G_v(\gamma(t), \dot\gamma(t)) = \dim(M) - 1,$$

which proves point (ii).

Let us reduce the setting to finite dimension. Let $k \ge 2$ be a large enough integer such that no restriction $\gamma|_{[a,b]}$ with $b - a < k^{-1}$ contains conjugate points; namely, there are no Jacobi vector fields along γ vanishing on more than one point of $[a, b]$. We consider the finite dimensional vector space $V \subset Z$ of those vector fields $\zeta \in Z$ such that, for all $i = 0, \dots, k-1$, each restriction $\zeta|_{[i/k,(i+1)/k]}$ is a Jacobi vector field. The Morse indices of h and $h|_V$ are the same

$$\mathrm{ind}(h|_V) = \mathrm{ind}(h), \quad \mathrm{nul}(h|_V) = \mathrm{nul}(h).$$

Indeed, an integration by parts in (4.4) shows that the h-orthogonal to V is the subspace

$$V^h = \big\{\zeta \in Z \;\big|\; h(\zeta, \cdot)|_V = 0\big\} = \big\{\zeta \in Z \;\big|\; \zeta(\tfrac{i}{k}) = 0 \quad \forall i = 0, \dots, k-1\big\},$$

and we have $Z = V \oplus V^h$ and $\mathrm{ind}(h|_{V^h}) + \mathrm{nul}(h|_{V^h}) = 0$. Analogously, if we set $V_0 := V \cap Z_0$, we have

$$\begin{aligned} \mathrm{ind}(h|_{V_0}) &= \mathrm{ind}(h|_{Z_0}) \\ \mathrm{nul}(h|_{V_0}) &= \mathrm{nul}(h|_{Z_0}). \end{aligned} \tag{4.5}$$

From now on, we will simply write h for the restriction $h|_V$.

The h-orthogonal $V_0^h = \big\{\zeta \in V \;\big|\; h(\zeta, \cdot)|_{V_0} = 0\big\}$ is precisely the space of vector fields $\zeta \in Z$ such that $\zeta|_{(0,1)}$ is a Jacobi vector field. We denote by $\mathrm{Ver} := \ker d\pi \subset \mathrm{TT}M$ the vertical sub-bundle of $\mathrm{TT}M$. Each intersection

$$(\xi \cap \mathrm{Ver})_{\tilde\gamma(t)} = (SM \cap \mathrm{Ver})_{\tilde\gamma(t)}$$

has dimension $\dim(M) - 1$. For each $\zeta \in V_0^h$, we set

$$\tilde\zeta(t) := (\zeta(t), \dot\zeta(t)) = \Phi_t(\tilde\zeta(0^+)), \quad \forall t \in (0, 1).$$

Notice that there is an isomorphism

$$V_0^h \to (\Phi_1 - I)|^{-1}_{\xi_{\tilde\gamma(0)}}(\xi \cap \mathrm{Ver})_{\tilde\gamma(0)}, \quad \zeta \mapsto \tilde\zeta(0^+).$$

In particular

$$\begin{aligned}\dim(V_0^h) &\leq \dim\ker(\Phi_1 - I)|_{\xi_{\tilde\gamma(0)}} + \dim(\xi\cap \mathrm{Ver})_{\tilde\gamma(0)}\\ &= \mathrm{nul}(h) + \dim(M) - 1. \end{aligned} \tag{4.6}$$

Moreover, the evaluation map $V_0^h \to \ker(G_v(\tilde\gamma(0)))$, $\zeta \mapsto \zeta(0)$ is surjective, and its kernel is precisely $\ker(h|_{V_0})$. Therefore,

$$\begin{aligned}\dim(V_0^h) &= \dim\ker(h|_{V_0}) + \dim\ker(G_v(\gamma(0),\dot\gamma(0)))\\ &= \mathrm{nul}(h|_{V_0}) + \dim(M) - 1. \end{aligned} \tag{4.7}$$

The general formula relating the Morse indices of a quadratic form to the ones of its restriction to a subspace (see, e.g., [34, Section A.2]) gives

$$\mathrm{ind}(h) = \mathrm{ind}(h|_{V_0}) + \mathrm{ind}(h|_{V_0^h}) + \mathrm{nul}(h|_{V_0^h}) - \mathrm{nul}(h). \tag{4.8}$$

In particular, by (4.6), we have

$$\mathrm{ind}(h) \leq \mathrm{ind}(h|_{V_0}) + \dim(V_0^h) - \mathrm{nul}(h) \leq \mathrm{ind}(h|_{V_0}) + \dim(M) - 1,$$

which proves point (iii). By (4.7) and (4.8), we have

$$\begin{aligned}\mathrm{ind}(h) + \mathrm{nul}(h) &\leq \mathrm{ind}(h|_{V_0}) + \dim(V_0^h)\\ &\leq \mathrm{ind}(h|_{V_0}) + \mathrm{nul}(h|_{V_0}) + \dim(M) - 1,\end{aligned}$$

which proves point (iv).

From now on, let us now assume that M is an orientable surface. The classical index theorem of Morse [36] allows us to express $\mathrm{ind}(h|_{Z_0})$ and $\mathrm{nul}(h|_{Z_0})$ as

$$\begin{aligned}\mathrm{ind}(h|_{V_0}) &= \sum_{t\in(0,1)} \dim\big(\Phi_t(\mathrm{Ver}_{\tilde\gamma(0)}) \cap \mathrm{Ver}_{\tilde\gamma(t)}\big),\\ \mathrm{nul}(h|_{V_0}) &= \dim\big(\Phi_1(\mathrm{Ver}_{\tilde\gamma(0)}) \cap \mathrm{Ver}_{\tilde\gamma(1)}\big).\end{aligned}$$

Notice that the Liouville vector field L takes values in the vertical sub-bundle Ver, and Eq. (4.3) implies that

$$\Phi_t(\mathrm{Ver}_{\tilde\gamma(0)}) \cap \mathrm{Ver}_{\tilde\gamma(t)} = \Phi_t((\xi\cap\mathrm{Ver})_{\tilde\gamma(0)}) \cap (\xi\cap\mathrm{Ver})_{\tilde\gamma(t)}, \quad \forall t\neq 0.$$

Since the fibers of the bundle $\xi\cap\mathrm{Ver}$ have dimension 1, we can express these index formulas by means of a single vector field η, as follows. Let us fix an arbitrary non-zero vector $\tilde\eta_0 \in (\xi\cap\mathrm{Ver})_{\tilde\gamma(0)}$, and define

$$\tilde\eta(t) = (\eta(t), \dot\eta(t)) := \Phi_t(\tilde\eta_0),$$

so that η is a Jacobi field along γ such that $\eta(0) = 0$ and $G_v(\gamma,\dot\gamma)\eta \equiv 0$. Since M is an orientable surface, the normal bundle of γ is trivial, and we can find a nowhere-vanishing 1-periodic smooth vector field μ along γ such that $G_v(\gamma,\dot\gamma)\mu \equiv 0$, so that we can express η as

$$\eta(t) = f(t)\mu(t)$$

for some smooth function $f : \mathbb{R} \to \mathbb{R}$. Notice that, since η is a Jacobi vector field that does not vanish identically, it has isolated zeroes, and in particular

$\dot f(t)\neq 0$ whenever $f(t)=0$. The index theory of Morse reduces to

$$\mathrm{ind}(h|_{Z_0})=\#\{t\in(0,1)\mid f(t)=0\},\quad \mathrm{nul}(h|_{Z_0})=\begin{cases}1 & \text{if } f(1)=0,\\ 0 & \text{if } f(1)\neq 0.\end{cases}$$

If $\mathrm{nul}(h|_{Z_0})=1$, then $\eta(1)=0$, and therefore,

$$\tilde\eta(t+1)=\tfrac{\dot f(1)}{\dot f(0)}\tilde\eta(t),\quad \forall t\in\mathbb{R}$$

This readily implies that $\mathrm{nul}_\Omega(\gamma^m)=1$ and

$$\mathrm{ind}_\Omega(\gamma^m)=\#\{t\in(0,m)\mid f(t)=0\}=m\,\mathrm{ind}_\Omega(\gamma)+m-1.$$

This proves point (v).

With an integration by parts in (4.4), we readily see that the quadratic form h on the space V_0^h can be expressed in local coordinates as

$$\begin{aligned}h(\zeta,\zeta)&=\langle G_{vv}\dot\zeta(1^-)+G_{xv}\zeta(1)-G_{vv}\dot\zeta(0^+)-G_{xv}\zeta(0),\zeta(0)\rangle\\ &=\mathrm{d}\lambda((\Phi_1-I)\tilde\zeta(0^+),\tilde\zeta(0^+)).\end{aligned}\tag{4.9}$$

Let us assume that $\mathrm{nul}(\gamma)=2$, so that $\Phi_1|_{\xi_{\tilde\gamma(0)}}=I$, $(\Phi_1-I)\tilde\zeta(0^+)=0$ for all $\zeta\in V_0^h$, and $\mathrm{nul}(h|_{V_0})=1$. Equation (4.9) implies that $h|_{V_0^h}=0$. Since $\dim(V_0^h)=1+\mathrm{nul}(h|_{V_0})=2$, this implies that $\mathrm{ind}(h|_{V_0^h})=0$ and $\mathrm{nul}(h|_{V_0^h})=2$. Therefore, Eq. (4.8) becomes $\mathrm{ind}(h)=\mathrm{ind}(h|_{V_0})$. Since $\Phi_1|_{\xi_{\tilde\gamma(0)}}=I$, in particular the above Jacobi field η is (smoothly) 1-periodic, and so is the function f. Therefore, since f has non-zero derivative at its zeroes, it must vanish an odd number of times in the open interval $(0,1)$. Equation (4.5) allows to conclude that $\mathrm{ind}(\gamma)=\mathrm{ind}(h)=\mathrm{ind}(h|_{V_0})$ is odd. We can now repeat the same argument for all the iterates γ^m, since

$$\mathrm{nul}(\gamma^m)=\dim\ker(\Phi_1|^m_{\xi_{\tilde\gamma(0)}}-I)=2,$$

and conclude that $\mathrm{ind}(\gamma)^m$ is odd as well for all $m\in\mathbb{N}$. This proves point (vi).

Finally, let us assume that $\mathrm{ind}_\Omega(\gamma^m)>0$ for some $m\geq 1$, which is equivalent to the fact that the Jacobi field η introduced above vanishes at some positive time. Let $\tau>0$ be the minimum $t>0$ such that $\eta(t)=0$. Up to replacing μ with $-\mu$, we can assume that $f|_{(0,\tau)}>0$, so that $\dot f(0)>0$ and $\dot f(\tau)<0$. If $\tau\leq 1$, we consider the 1-periodic vector field along γ

$$\theta(t)=\begin{cases}\eta(t), & \text{if } t\in[0,\tau],\\ 0, & \text{if } t\in[\tau,1],\end{cases}$$

which satisfies $h(\theta,\theta)=0$ and

$$\begin{aligned}h(\theta,\mu)&=\langle G_{vv}\,\dot\eta(\tau)-G_{vv}\,\dot\eta(0),\mu(0)\rangle\\ &=\dot f(\tau)\,\langle G_{vv}\,\mu(\tau),\mu(\tau)\rangle-\dot f(0)\,\langle G_{vv}\,\mu(0),\mu(0)\rangle\\ &<0.\end{aligned}$$

For each $\epsilon>0$ the piecewise smooth vector field $\theta+\epsilon\mu$ is 1-periodic and everywhere transverse to $\dot\gamma$. Moreover,

$$h(\theta+\epsilon\mu,\theta+\epsilon\mu)=2\epsilon h(\mu,\theta)+\epsilon^2 h(\mu\mu)$$

which is negative if $\epsilon > 0$ is sufficiently small. Assume now that $\tau > 1$. In this case, there exists $t > 0$ such that $f(t) = f(t+1) > 0$, $\dot f(t) > 0$ and $\dot f(t+1) < 0$. We define θ to be the 1-periodic vector field along γ such that $\theta|_{[t,t+1]} = \eta|_{[t,t+1]}$. Notice that θ is everywhere transverse to $\dot\gamma$, and satisfies

$$\begin{aligned} h(\theta,\theta) &= \langle G_{vv}\, \dot\eta(t+1), \eta(t+1)\rangle - \langle G_{vv}\, \dot\eta(t), \eta(t)\rangle \\ &= \langle G_{vv}\, (\dot\eta(t+1) - \dot\eta(t)), \eta(t)\rangle \\ &= (\dot f(t+1) - \dot f(t))\, f(t)\, \langle G_{vv}\, \mu(t), \mu(t)\rangle \\ &< 0. \end{aligned}$$

In both cases, we can approximate θ with a C^0-close 1-periodic smooth vector field ζ. Such a ζ will still be everywhere transverse to $\dot\gamma$ and will still satisfy $h(\zeta,\zeta) < 0$. This completes the proof of point (vii). $\square$

4.3. Local homology

The last index that is usually employed in critical point theory is the local homology, whose construction we now recall for closed geodesics of Finsler manifolds (M, F). Actually, the theory of local homology is very general, and essentially does not see the difference between the Riemannian and the Finsler settings. We refer the reader to [1, 10, 41], and in particular to [1, Section 3], for a more comprehensive treatment.

For any $\mathcal{U} \subset \Lambda$, $U \subset \Lambda_k$, and $\ell > 0$, we set

$$\mathcal{U}^{<\ell} := \big\{\gamma \in \mathcal{U} \bigm| E(\gamma) < \ell^2\big\}, \quad U^{<\ell} := \big\{\boldsymbol{x} \in U \bigm| E_k(\boldsymbol{x}) < \ell^2\big\}.$$

Notice that $\mathcal{U}^{<\ell}$ and $U^{<\ell}$ are sublevel sets of the energy functional E, whereas in (3.1) we denoted by $\mathcal{W}^{<\ell}$ a sublevel set of the length functional L. Nevertheless, the notation is consistent: $\mathcal{W}$ was indeed a subset of the space of unparametrized loops Π, and if we parametrize any $\gamma \in \mathcal{W}$ with constant speed and period 1 we have $L(\gamma)^2 = E(\gamma)$.

The energy functional E is invariant under the circle action

$$u \cdot \gamma = \gamma(u + \cdot) \in \Lambda, \quad \forall u \in S^1,\ \gamma \in \Lambda.$$

Therefore, every closed geodesic $\gamma \in \mathrm{crit}(E) \cap E^{-1}(\ell^2)$ (with $\ell > 0$) belongs to a circle of critical points of E

$$S^1 \cdot \gamma := \big\{\gamma(u + \cdot) \in \Lambda \bigm| u \in S^1\big\}$$

A closed geodesic γ is said to be isolated when the critical circles of its iterates $S^1 \cdot \gamma^m$ are isolated in $\mathrm{crit}(E)$. Under this assumption, the local homology of γ and of $S^1 \cdot \gamma$ are the relative homology groups

$$C_*(\gamma) := H_*(\Lambda^{<\ell} \cup \{\gamma\}, \Lambda^{<\ell}), \quad C_*(S^1 \cdot \gamma) := H_*(\Lambda^{<\ell} \cup S^1 \cdot \gamma, \Lambda^{<\ell}).$$

As we already mentioned, we will specify the coefficient field in the notation only when we will need to employ a specific one.

Even though the energy function E may not be C^2, the local homology of an isolated closed geodesic $C_*(\gamma)$ is isomorphic to the local homology of a smooth function on a finite dimensional manifold at an isolated critical point of index $\mathrm{ind}(\gamma)$ and nullity $\mathrm{nul}(\gamma)$. Indeed, if $k \in \mathbb{N}$ is large enough so that the isolated closed geodesic $\gamma \in \mathrm{crit}(E)\cap E^{-1}(\ell^2)$ belongs to Λ_k, the inclusion induces the homology isomorphism

$$H_*(\Lambda_k^{<\ell} \cup \{\gamma\}, \Lambda_k^{<\ell}) \xrightarrow{\cong} C_*(\gamma).$$

The energy $E_k = E|_{\Lambda_k}$ is smooth in a neighborhood of the critical point γ (indeed, E_k is smooth at all those $\zeta \in \Lambda_k$ such that $\zeta(\frac{i}{k}) \neq \zeta(\frac{i+1}{k})$ for all $i \in \mathbb{Z}_k$). Let $\Sigma \subset M$ be an embedded hypersurface intersecting γ transversely at $x_0 := \gamma(0)$. We define the smooth hypersurface

$$\Sigma_k := \big\{\zeta \in \Lambda_k \;\big|\; \zeta(0) \in \Sigma\big\} \subset \Lambda_k.$$

It turns out that γ is an isolated critical point of $E_k|_{\Sigma_k}$ of index $\mathrm{ind}(\gamma)$ and nullity $\mathrm{nul}(\gamma)$, and the inclusion induces the homology isomorphism

$$H_*(\Sigma_k^{<\ell} \cup \{\gamma\}, \Sigma_k^{<\ell}) \xrightarrow{\cong} C_*(\gamma), \tag{4.10}$$

see [1, Prop. 3.1].

Since γ is an isolated critical point of $E_k|_{\Sigma_k}$, its local homology can also be expressed by means of the so-called Gromoll–Meyer neighborhoods [18]: these are suitable arbitrarily small compact path-connected neighborhoods $W \subset \Sigma_k$ of γ such that, for some $\delta' > 0$ and for all $\delta \in [0, \delta']$, the inclusion induces the homology isomorphisms

$$H_*(\Sigma_k^{<\ell} \cup \{\gamma\}, \Sigma_k^{<\ell}) \xleftarrow{\cong} H_*(W^{<\ell} \cup \{\gamma\}, W^{<\ell-\delta}) \xrightarrow{\cong} H_*(W, W^{<\ell-\delta}). \tag{4.11}$$

Indeed, a homotopy inverse of these inclusion can be built by suitably "pushing" in the direction given by a pseudo-gradient of the energy functional E, see e.g., [13, Theorem 5.2]. Gromoll–Meyer neighborhoods are particularly useful to prove certain technical statements concerning the local homology. For instance the following one that we will employ in the proof of Corollary 5.7.

Lemma 4.3. *Let $\gamma \in \mathrm{crit}(E) \cap E^{-1}(\ell^2)$, with $\ell > 0$, be an isolated closed geodesic. Assume that, for any sufficiently small open neighborhood $\mathcal{U} \subset \Lambda$ or $\mathcal{U} \subset \Lambda_k$ of γ, the open subset $\mathcal{U}^{<\ell}$ is not connected. Then, the local homology $C_1(\gamma)$ is non-zero.*

Proof. We prove the lemma in the infinite dimensional setting of Λ, the proof for the setting of Λ_k being entirely analogous. Let $\mathcal{U}_0 \subset \Lambda$ be an open neighborhood of γ such that, for every open neighborhood $\mathcal{U} \subset \mathcal{U}_0$ of γ, the open subset $\mathcal{U}^{<\ell}$ is not connected. Let $\mathcal{U} \subset \mathcal{U}_0$ be one such open neighborhood. We consider the connected components $\mathcal{U}_1, \dots, \mathcal{U}_r \subset \mathcal{U}^{<\ell}$ such that $\gamma \notin \overline{\mathcal{U}}_i$ for all $i = 1, \dots, r$. The subset $\mathcal{V} := \mathcal{U} \setminus (\overline{\mathcal{U}}_1 \cup \cdots \cup \overline{\mathcal{U}}_r)$ is still an open neighborhood of γ contained in $\mathcal{U}_0$, and therefore, $\mathcal{V}^{<\ell}$ is not connected.

Let $W \subset \Sigma_k$ be a Gromoll–Meyer neighborhood of γ that is small enough so that $W \subset \mathcal{V}$. We claim that, for each connected component $\mathcal{V}' \subset \mathcal{V}^{<\ell}$, we have

$$\mathcal{V}' \cap W \neq \varnothing. \tag{4.12}$$

This implies that $W^{<\ell}$ is not path-connected. Since W is path-connected, the long exact sequence

$$\begin{array}{ccccccccc} \dots & \longrightarrow & H_1(W,W^{<\ell}) & \longrightarrow & H_0(W^{<\ell}) & \longrightarrow & H_0(W) & \longrightarrow & \dots \\ & & & & & & \| & & \\ & & & & & & 0 & & \end{array}$$

implies that $H_1(W,W^{<\ell})$ is non-zero. This latter relative homology group is isomorphic to $C_1(\gamma)$, according to (4.10) and (4.11).

It remains to establish (4.12). Since $\gamma \in \partial\mathcal{V}'$, there exists $\zeta_0 \in \mathcal{V}'$ arbitrarily close to γ and such that $\zeta_0(0) \in \Sigma$. For each $s \in (0,1]$, we define $\zeta_s \in \Lambda$ to be the unique loop such that, for each $i = 0,\dots,k-1$, the segment $\zeta_s|_{[i/k,(i+s)/k]}$ is a length-minimizing geodesic, while $\zeta_s|_{[(i+s)/k,(i+1)/k]} = \zeta_0|_{[(i+s)/k,(i+1)/k]}$. Notice that $s \mapsto \zeta_s$ is a continuous path in Λ, and $\zeta_1 \in \Sigma_k$. Up to choosing the initial loop ζ_0 to be sufficiently close to γ, every ζ_s is contained in the neighborhood $\mathcal{V}$. Since $E(\zeta_s) \le E(\zeta_0) < \ell$, we actually have that every ζ_s is contained in $\mathcal{V}'$. Finally, if we choose ζ_0 to be sufficiently close to γ, we have that $\zeta_1 \in W$. □

The local homology groups of the critical circles of closed geodesics are the “building blocks” for the homology of the free loop space Λ. Indeed, if $\gamma \in \mathrm{crit}(E) \cap E^{-1}(\ell^2)$ is an isolated closed geodesic and the interval $(\ell, \ell+\epsilon)$ does not contain critical values of E, the inclusion induces an injective homomorphism

$$C_*(S^1 \cdot \gamma) \hookrightarrow H_*(\Lambda^{<\ell+\epsilon}, \Lambda^{<\ell}),$$

see, e.g., [19, proof of Lemma 4].

The local homology of an isolated closed geodesic γ does not vary (up to a shift in degree) under iterations that preserve the nullity. In particular, if $\mathrm{ind}(\gamma) = \mathrm{ind}(\gamma^m)$ and $\mathrm{nul}(\gamma) = \mathrm{nul}(\gamma^m)$, the iteration map $\psi^m : \Lambda \hookrightarrow \Lambda$, $\psi^m(\zeta) = \zeta^m$ induces the local homology isomorphisms

$$\psi^m_* : C_*(\gamma) \xrightarrow{\cong} C_*(\gamma^m), \quad \psi^m_* : C_*(S^1 \cdot \gamma) \xrightarrow{\cong} C_*(S^1 \cdot \gamma^m).$$

This is actually a consequence of a general Morse-theoretic result due to Gromoll–Meyer [18, Lemma 7].

The local homology of an isolated closed geodesic often embeds into the local homology of its critical circle. More precisely, the following statement holds. A closed geodesic $\gamma \in \mathrm{crit}(E) \cap E^{-1}(0,\infty)$ is said to be prime when $\gamma = \zeta^m$ if and only if $\zeta = \gamma$ and $m = 1$.

Lemma 4.4. *If $\gamma \in \mathrm{crit}(E) \cap E^{-1}(0,\infty)$ is an isolated prime closed geodesic, then for all odd numbers $m \in \mathbb{N}$ the inclusion induces an injective homomorphism*

$$C_*(\gamma^m;\mathbb{Q}) \hookrightarrow C_*(S^1 \cdot \gamma^m;\mathbb{Q}).$$

Proof. The proof of this fact is rather long, and we only provide its main steps. Let $m > 0$ be a positive integer. We denote by $\mu : \Lambda \to \Lambda$ the continuous map $\mu(\gamma) = \gamma(\frac{1}{m} + \cdot)$. There is an exact sequence

$$0 \longrightarrow (\mu_* - \mathrm{id})C_*(\gamma^m;\mathbb{Q}) \longrightarrow C_*(\gamma^m;\mathbb{Q}) \longrightarrow C_*(S^1 \cdot \gamma^m;\mathbb{Q}), \qquad (4.13)$$

where all the homomorphisms are induced by the inclusion, see [1, Lemma 3.4]. The homomorphism $\mu_* : C_*(\gamma^m;\mathbb{Q}) \to C_*(\gamma^m;\mathbb{Q})$ turns out to be equal to

$$\mu_* = (-1)^{\mathrm{ind}(\gamma^m)-\mathrm{ind}(\gamma)}\mathrm{id}, \tag{4.14}$$

see [1, Lemma 3.5]. Bott's iteration theory [11] implies that the Morse indices $\mathrm{ind}(\gamma)$ and $\mathrm{ind}(\gamma^m)$ have the same parity provided m is odd. This, together with (4.13) and (4.14), implies that the inclusion induces an injective homomorphism $C_*(\gamma^m;\mathbb{Q}) \hookrightarrow C_*(S^1\cdot\gamma^m;\mathbb{Q})$ for all odd integers $m>0$. □

We close this section by proving the following proposition relating the local homology in $\Lambda = W^{1,2}(S^1,M)$ to the one in $\Pi = \mathrm{Emb}(S^1,M)/\mathrm{Diff}(S^1)$ of Sect. 3. By applying the proposition to the three min-max values $\ell(h_i)$, for $i=1,2,3$, defined in (3.5), we will infer Theorem 1.3(iii).

Proposition 4.5. *Let (M,F) be a closed, orientable, reversible Finsler surface, $\rho>0$ a constant, and $h\in H_d(\Pi,\Pi^{<\rho})$ a non-trivial homology class. Assume that there are only finitely many simple closed geodesics of (M,F) having length in a neighborhood of $\ell(h)$. Then, there exists a simple closed geodesic $\gamma\in\mathrm{crit}(E)\cap E^{-1}(\ell(h)^{1/2})$ with non-zero local homology $C_d(\gamma)\neq 0$.*

Proof. We first apply Lemma 3.1 and obtain a simple closed geodesic γ of length $\ell:=\ell(h)$ and, for every $\epsilon>0$, an open neighborhood $\mathcal{V}(\gamma,\epsilon)\subset\Pi$ and a constant $\delta\in(0,\epsilon^2)$ such that the homomorphism

$$H_*(\mathcal{V}(\gamma,\epsilon),\mathcal{V}(\gamma,\epsilon)^{<\ell-\delta}) \to H_*(\Pi,\Pi^{<\ell})$$

induced by the inclusion is non-zero.

Let $\Sigma\subset M$ be an embedded open hypersurface (i.e., an open segment) intersecting γ transversely. We choose Σ and $\epsilon_0>0$ small enough so that every $\zeta\in\mathcal{V}(\gamma,\epsilon_0)$ intersect Σ in a single point and, by the implicit function theorem, the map $\mathcal{V}(\gamma,\epsilon_0)\to\Sigma$, $\zeta\mapsto\zeta\cap\Sigma$ is continuous. Throughout this section, we uniquely parametrize every $\zeta\in\mathcal{V}(\gamma,\epsilon_0)$ so that

$$\zeta : S^1\to M,\quad F(\zeta,\dot\zeta)\equiv L(\zeta),\quad \zeta(0)\in\Sigma.$$

With this choice of parametrizations, we identify $\mathcal{V}(\gamma,\epsilon_0)$ with a subset of the space of embeddings $\mathrm{Emb}(S^1,M)$, endowed as usual with the C^∞-topology. Notice that $\mathcal{V}(\gamma,\epsilon_0)$ is relatively compact in the C^1 topology. Moreover, every C^1-open neighborhood of γ contains $\mathcal{V}(\gamma,\epsilon)$ for a sufficiently small $\epsilon\in(0,\epsilon_0]$.

Let us consider an embedding $M\hookrightarrow\mathbb{R}^3$, which exists since M is an orientable closed surface. Let $U\subset\mathbb{R}^3$ be a tubular neighborhood of M with associated smooth retraction $\pi:U\to M$. We consider a family of mollifiers $\theta_s(u)=\theta(u/s)/s$, where $s\in(0,1)$ and $\theta:S^1\to[0,\infty)$ is a smooth function supported in $(-1/2,1/2)$ and with integral 1. We denote by $*$ the convolution operation. Since $\mathcal{V}(\gamma,\epsilon_0)$ is relatively C^1-compact and θ_s tends to the Dirac delta as $s\to0$, there exists $s_0>0$ and $\epsilon_1\in(0,\epsilon_0]$ such that we have a well-defined continuous map

$$c:[0,s_0]\times\mathcal{V}(\gamma,\epsilon_1)\to\mathrm{Emb}(S^1,M),\quad c(s,\zeta)(u)=c_s(\zeta)(u)=\pi(\zeta*\theta_s(u)).$$

Notice that $c_0(\zeta) = \zeta$ for all $\zeta \in \mathcal{V}(\gamma, \epsilon_1)$. Since the length function is continuous on the relatively C^1-compact subset $\mathcal{V}(\gamma, \epsilon_1)$, there exists $s_1 \in (0, s_0]$ such that

$$L(c_s(\zeta)) < L(\zeta) + \delta/2, \quad \forall s \in [0, s_1],\ \zeta \in \mathcal{V}(\gamma, \epsilon). \tag{4.15}$$

By the continuity of the convolution, there exists an open subset $\mathcal{U} \subset W^{1,2}(S^1, M)$ containing $\mathcal{V}(\gamma, \epsilon)$ such that c_{s_1} extends as a continuous map

$$c_{s_1} : \mathcal{U} \to \mathrm{Emb}(S^1, M), \quad c_{s_1}(\zeta)(u) = \pi(\zeta * \theta_{s_1}(u)),$$

and

$$L(c_{s_1}(\zeta)) < L(\zeta) + \delta, \quad \forall \zeta \in \mathcal{U}. \tag{4.16}$$

We consider an integer

$$k > \frac{\ell + \epsilon_1^2}{\mathrm{injrad}(M, F)}$$

that we will soon fix, and the space Σ_k of broken closed geodesics intersecting Σ at time 0. We define a continuous homotopy

$$r_t : \mathcal{V}(\gamma, \epsilon_1) \to W^{1,2}(S^1, M), \quad t \in [0, 1],$$

as follows: we uniquely parametrize every $\zeta \in \mathcal{V}(\gamma, \epsilon_0)$ so that

$$F(\zeta, \dot\zeta) \equiv L(\zeta), \quad \zeta(0) \in \Sigma;$$

for all $i = 0, \ldots, k-1$, we define

$$r_t(\zeta)|_{[i/k,(i+1-t)/k]} := \zeta|_{[i/k,(i+1-t)/k]},$$

and $r_t(\zeta)|_{[(i+1-t)/k,(i+1)/k]}$ as the shortest geodesic of (M, F) parametrized with constant speed and joining its endpoints. We require k to be large enough so that every r_t has image inside the open subset $\mathcal{U} \subset W^{1,2}(S^1, M)$. Clearly,

$$E(r_t(\zeta)) \le E(\zeta) = L(\zeta)^2, \quad \forall t \in [0, 1].$$

We consider a Gromoll–Meyer neighborhood $W \subset \Sigma_k \cap \mathcal{U}$ of $\gamma = r_1(\gamma)$. Notice that, by (4.16), we have

$$L(c_{s_1}(\zeta)) < L(\zeta) + \delta \le E(\zeta)^{1/2} + \delta, \quad \forall \zeta \in W,$$

and in particular $c_{s_1}(W^{<\ell-\delta}) \subset \Pi^{<\ell}$. We fix a constant $\epsilon_2 \in (0, \epsilon_1]$ small enough so that $r_1(\mathcal{V}(\gamma, \epsilon_1)) \subset W$. Overall, we have the homomorphisms

$$\begin{tikzcd} H_*(\mathcal{V}(\gamma,\epsilon_2), \mathcal{V}(\gamma,\epsilon_2)^{<\ell-\delta}) \arrow[rr, "i_*"] \arrow[dr, "(r_1)_*"'] & & H_*(\Pi, \Pi^{<\ell}) \\ & H_*(W, W^{<\ell-\delta}) \arrow[ur, "(c_{s_1})_*"'] & \end{tikzcd} \tag{4.17}$$

where i_* is the non-zero homomorphism induced by the inclusion (see the first paragraph of the proof). All we need to do to complete the proof is to show that the diagram (4.17) commutes. This is a consequence of the fact that the inclusion i is homotopic to the composition $c_{s_1} \circ r_1$ via the continuous homotopy

$$h_t : \mathcal{V}(\gamma, \epsilon_2) \to \Pi, \quad h_t = \begin{cases} c_{2ts_1}, & \text{if } t \in [0, 1/2], \\ c_{s_1} \circ r_{2t-1}, & \text{if } t \in [1/2, 1], \end{cases}$$

which satisfies $h_0 = i$, $h_1 = c_{s_1} \circ r_1$, and $h_t(\mathcal{V}(\gamma, \epsilon_2)^{<\ell-\delta}) \subset \Pi^{<\ell}$ for all $t \in [0, 1]$ according to (4.15) and (4.16). □

5. Infinitely many closed geodesics

5.1. The Birkhoff map

Let (S^2, F) be a reversible Finsler sphere, and $SS^2 = \{(x, v) \mid F(x, v) = 1\}$ its Finsler unit tangent bundle with base projection $\pi : SM \to M$, $\pi(x, v) = x$. As we already recalled in the proof of Proposition 4.2, SS^2 admits the contact form $\alpha = G_v \, \mathrm{d}\pi$, where $G(x, v) = \frac{1}{2} F(x, v)^2$, and the associated Reeb vector field X on SM defined by $\alpha(X) \equiv 1$ and $\mathrm{d}\alpha(X, \cdot) \equiv 0$. The flow $\phi_t : SM \to SM$ of X is precisely the geodesic flow of (S^2, F).

Let $\gamma : S^1 \hookrightarrow S^2$ a simple closed geodesic of (S^2, F). Without loss of generality, let us assume that $F(\gamma, \dot\gamma) \equiv 1$. The complement $S^2 \setminus \gamma$ is the disjoint union of two open balls $B_0, B_1 \subset S^2$. We consider the open annuli

$$A_i := \big\{(x, v) \in SS^2 \bigm| x \in \gamma(S^1),\ v \pitchfork \dot\gamma(t) \text{ and points inside } B_i\big\}, \quad i = 0, 1.$$

Since the Reeb vector field X is transverse to A_i, we readily see that

$$\mathrm{d}\alpha|_{A_i} = (X \lrcorner (\alpha \wedge \mathrm{d}\alpha))|_{A_i}$$

is a symplectic form on A_i. We assume that the first return time

$$\tau_i : A_i \to (0, \infty], \quad \tau_i(x, v) := \inf\big\{t > 0 \bigm| \phi_t(x, v) \in A_{1-i}\big\} \in (0, +\infty]$$

is finite for all $(x, v) \in A_i$ (here, we adopt the usual convention $\inf \varnothing = +\infty$). Under this assumption, there is a well-defined first return map

$$\psi_i : A_i \to A_{1-i}, \quad \psi_i(x, v) = \phi_{\tau_i(x,v)}(x, v),$$

which is a diffeomorphism. Since

$$\psi_i^* \alpha - \alpha = \phi_t^* \alpha|_{t=\tau_i} + \alpha(\partial_t \phi_t(z))|_{t=\tau_i} \mathrm{d}\tau_i - \alpha = \alpha + \alpha(X)\mathrm{d}\tau_i - \alpha = \mathrm{d}\tau_i,$$

the first return map is an exact symplectomorphism $\psi_i : (A_i, \mathrm{d}\alpha) \to (A_{1-i}, \mathrm{d}\alpha)$. Notice that

$$\partial A_0 = \partial A_1 = \{(\gamma(t), \dot\gamma(t)), (\gamma(t), -\dot\gamma(t)) \mid t \in S^1\},$$

and we readily see that $\mathrm{d}\alpha|_{\overline{A}_i}$ vanishes on ∂A_i.

For each $t \in [0, 1)$, we choose a non-zero $w_t \in \ker \mathrm{d}\pi(\tilde\gamma(0))$ depending smoothly on t, and we extend it to a vector field

$$\tilde\eta_t(s) = (\eta_t(s), \dot\eta_t(s)) := \mathrm{d}\phi_{s-t}(\tilde\gamma(t)) w_t. \tag{5.1}$$

Namely, η_t is a non-trivial Jacobi vector field along γ such that $G_v(\gamma, \dot\gamma)\eta_t \equiv 0$ and $\eta_t(t) = 0$. We recall that the points $\gamma(t), \gamma(s)$, with $t \neq s$, are conjugate when $\eta_t(s) = 0$. For each $t \in [0, 1)$, we set

$$t_{-1} := \sup\{s < t \mid \eta_t(s) = 0\}, \quad t_1 := \inf\{s > t \mid \eta_t(s) = 0\},$$

and $t_{\pm 2} := (t_{\pm 1})_{\pm 1}$ (here, once again, we set $\sup\varnothing = -\infty$ and $\inf\varnothing = +\infty$). Namely, t_i is the time of the $|i|$-th conjugate point to $\gamma(t)$ after t if $i > 0$, or before t if $i < 0$.

Lemma 5.1. *Assume that, for some* $t \in [0,1)$, t_1 *is finite. Then, for all* $t \in [0,1)$ *both* t_1 *and* t_{-1} *are finite, and the first return maps* ψ_i *can be extended as homeomorphisms*

$$\psi_i : \overline{A}_i \to \overline{A}_{1-i}, \quad \psi_i(\gamma(t), \pm\dot\gamma(t)) = (\gamma(t_{\pm 1}), \pm\dot\gamma(t_{\pm 1})). \tag{5.2}$$

Proof. Let μ be a nowhere-vanishing 1-periodic vector field along γ such that $G_v(\gamma,\dot\gamma)\mu \equiv 0$. The Jacobi fields η_t can be written as $\eta_t(s) = f(t,s)\mu(s)$ for some smooth function $f : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$. Up to replacing μ with $-\mu$, we can assume that $f(t, t+\epsilon) > 0$ for all $t \in S^1$ and $\epsilon > 0$ small enough. Since the Jacobi fields η_t are non-trivial, we have $\partial_s f(t,s) \neq 0$ whenever $f(t,s) = 0$. This readily implies that, if t_1 is finite for some $t \in \mathbb{R}$, the same is true for all $t \in \mathbb{R}$, and the function $t \mapsto t_1$ is continuous and monotone increasing. Since $(t_1)_{-1} = t$, we infer that the function $t \mapsto t_{-1}$ is well-defined, continuous and monotone increasing as well.

We fix an arbitrary $t \in [0,1)$ and $(x,v) := (\gamma(t), \dot\gamma(t))$. To complete the proof, we are left to show that, for each sequence $v_n \in S_xS^2$ of vectors pointing inside A_i and such that $v_n \to \pm v$, we have $\tau_i(x, v_n) \to \pm(t_{\pm 1} - t)$. Indeed, this implies that $\phi_{\tau_i(x,v_n)}(x, v_n) \to (\gamma(t_{\pm 1}), \pm\dot\gamma(t_{\pm 1}))$, and therefore, the extension (5.2) of ψ_i is continuous and bijective. Since the annuli $\overline{A}_i$ and $\overline{A}_{i-1}$ are compact and Hausdorff, such an extension is a homeomorphism.

Let us focus on the case $v_n \to v$, the other one being analogous. We set $\gamma_n(s) := \exp_x((s-t)\, v_n)$ and $\sigma_n := \tau_i(x, v_n)$. We claim that

$$\liminf_{n\to\infty} \sigma_n \geq t_1 - t.$$

Otherwise, we could extract a subsequence such that $\sigma_n \to \sigma \in (0, t_1 - t)$; however, since the geodesic $\gamma|_{[t,t+\sigma]}$ has no conjugate points, this would contradict the fact that the exponential map $\exp_x$ is a local diffeomorphism at σv. The fact that $f(t_1+\epsilon) < 0$ if $\epsilon > 0$ is small enough readily implies that $\gamma_n(t_1+\epsilon) \in A_{1-i}$ for all n large enough, and therefore, $\sigma_n < t_1 - t + \epsilon$. This implies that $\sigma_n \to t_1 - t$. □

We set $A := A_0$. The previous lemma implies that the annulus A is a surface of section for the geodesic flow: a surface that is transverse to the vector field X on its interior, and whose boundary is the union of periodic orbits of the flow. The composition $\psi := \psi_1 \circ \psi_0 : A \to A$ is the first return map of the surface of section A, and extends to a homeomorphism of $\overline{A}$ as

$$\psi(\gamma(t), \pm\dot\gamma(t)) = (\gamma(t_{\pm 2}), \pm\dot\gamma(t_{\pm 2})).$$

As customary in the Riemannian literature, we will call ψ the Birkhoff map of γ. With a suitable change of coordinates, A becomes the standard symplectic annulus.

Lemma 5.2. *There exists a smooth homeomorphism*

$$\sigma : S^1 \times [-1,1] \to \overline{A}$$

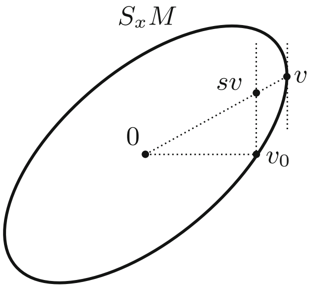


FIGURE 1. The value of $s \in [0,1]$ such that $v_0 - s\,v \in \ker G_v(x,v)$ can be found geometrically: the parallel to the tangent line $\mathrm{T}_v(S_xM)$ passing through v_0 intersects the segment joining the origin and v at sv

of the form $\sigma(t,s) = (\gamma(t), \nu(t,s))$, where $\nu(t,s) \in S_{\gamma(t)}M$ is the unique tangent vector (pointing inside B_0 or tangent to γ) such that

$$\dot\gamma(t) - s\,\nu(t,s) \in \ker G_v(\gamma(t), \nu(t,s)).$$

The map σ restrict to a diffeomorphism $\sigma : S^1 \times (-1,1) \to A$, and $\sigma^*\alpha = s\,\mathrm{d}t$.

Proof. We fix $x = \gamma(t)$ and $v_0 = \dot\gamma(t)$. For each $v \in S_xM$ there is a unique $s(v) \in [-1,1]$ such that $v_0 - s(v)v \in \ker G_v(x,v)$, see Fig. 1. Clearly, $s(v)$ depends smoothly on v. We choose an arbitrary parametrization of the fiber

$$v : [0,1] \xrightarrow{\cong} \overline{A} \cap S_xM$$

such that $v(0) = v_0$ and $v(1) = -v_0$. Notice that $\dot v(r) \in \ker G_v(x, v(r))$, and there exists $\lambda(r) \in \mathbb{R}$ such that $v_0 = s(v(r))v(r) + \lambda(r)\dot v(r)$. By the strict convexity of S_xM, we have $\lambda(r) = 0$ if and only if $r \in \{0,1\}$. Since

$$G_v(x,v)v_0 = G_v(x,v)s(v)v = s(v),$$

we have

$$\begin{aligned}
\mathrm{d}s(v(r))\dot v(r) = \tfrac{\mathrm{d}}{\mathrm{d}r}s(v(r)) = \tfrac{\mathrm{d}}{\mathrm{d}r}G_v(x,v(r))v_0 &= G_{vv}(x,v(r))\dot v(r)\,v_0\\
&= s(v(r))G_{vv}(x,v(r))\dot v(r)\,v(r) + \lambda(r)G_{vv}(x,v(r))\dot v(r)\,\dot v(r)\\
&= s(v(r))G_v(x,v(r))\dot v(r) + \lambda(r)G_{vv}(x,v(r))\dot v(r)\,\dot v(r)\\
&= \lambda(r)G_{vv}(x,v(r))\dot v(r)\,\dot v(r).
\end{aligned}$$

The last term is non-zero for all $r \in (0,1)$. Therefore, $s : \overline{A} \cap S_xM \to [-1,1]$ is a diffeomorphism that restricts to a diffeomorphism $s : A \cap S_xM \to (-1,1)$. We set $\sigma(x,\cdot)$ to be the inverse homeomorphism. The obtained map $\sigma : S^1 \times [-1,1] \to \overline{A}$ is thus a homeomorphism that restricts to a diffeomorphism $\sigma : S^1 \times (-1,1) \to A$. The pull-back of the contact form α by σ is

$$(\sigma^*\alpha)_{(t,s)} = G_v(\gamma(t),\nu(t,s))\dot\gamma(t)\,\mathrm{d}t = G_v(\gamma(t),\nu(t,s))s\,\nu(t,s)\,\mathrm{d}t = s\,\mathrm{d}t.$$

$\square$

From now on, the annulus $S^1 \times [-1,1]$ will be implicitly equipped with the Euclidean area form $\mathrm{d}s \wedge \mathrm{d}t$. By means of Lemma 5.2, we will always consider the Birkhoff map of a simple closed geodesic γ as a homeomorphism $\psi : S^1 \times [-1,1] \to S^1 \times [-1,1]$ that restricts to a symplectomorphism of $(S^1 \times (-1,1), \mathrm{d}s \wedge \mathrm{d}t)$ and acts on the boundary as $\psi(t, \pm 1) = (t_{\pm 2}, \pm 1)$.

5.2. Periodic points of twist maps

Let $\psi : S^1 \times [-1,1] \to S^1 \times [-1,1]$ be an area preserving homeomorphism preserving the boundary components $S^1 \times \{-1\}$ and $S^1 \times \{1\}$. Such a ψ is called a **twist map** when it admits a lift

$$\widetilde{\psi} : \mathbb{R} \times [-1,1] \to \mathbb{R} \times [-1,1], \quad \widetilde{\psi}(t,s) = (a(t,s), b(t,s)), \tag{5.3}$$

satisfying the twist conditions $a(t,1) < t$ and $a(t,-1) > t$ for all $t \in \mathbb{R}$. If ψ is the Birkhoff map of a simple closed geodesic γ of (S^2, F), the set of its periodic orbits in $S^1 \times (-1,1)$ is in one-to-one correspondence with the set of closed geodesics intersecting γ (other than γ itself). In particular, the existence of infinitely many periodic points of ψ implies the existence of infinitely many closed geodesics on (S^2, F)

By the celebrated Poincaré–Birkhoff theorem [9], any twist map has at least two fixed points in the interior of the annulus. Indeed, more is true: any lift (5.3) satisfying the twist condition has at least two fixed points. A simple argument due to Neumann [38] further implies that any twist map ψ has infinitely many periodic points. Indeed, consider the translation

$$\tau : \mathbb{R} \times [-1,1] \to \mathbb{R} \times [-1,1], \quad \tau(t,s) = (t+1, s).$$

For each integer $q > 0$ there exists another relatively prime integer $p > 0$ that is large enough so that

$$p \min_{x \in \mathbb{R}} \big(x - a(x,1)\big) > q.$$

This condition guarantees that $\widetilde{\phi} := \widetilde{\psi}^p \circ \tau^q$ is a lift of $\phi = \psi^p$ satisfying the twist condition, and therefore, has at least a fixed point $z \in \mathbb{R} \times [-1,1]$. Such a z projects to a p-periodic point $[z]$ of ψ, and since p, q are relatively prime the minimal period of $[z]$ is p.

Let us now apply this results to the Birkhoff map of a simple closed geodesic γ of (S^2, F). For each $t \in S^1$, we denote $t \cdot \gamma := \gamma(t + \cdot)$ the closed geodesic γ with the parametrization translated by t, and by $\mathrm{ind}_\Omega(t \cdot \gamma)$ the Morse index of $t \cdot \gamma$ in its corresponding based loop space $\Omega_t = \{\zeta \in \Lambda \mid \zeta(0) = \gamma(t)\}$ (see Sect. 4.2).

Theorem 5.3. *If the simple closed geodesic γ satisfies $\mathrm{ind}_\Omega(t \cdot \gamma) \geq 2$ for all $t \in S^1$ and has a well defined Birkhoff map ψ, then ψ is a twist map, and in particular (S^2, F) has infinitely many closed geodesics.*

Proof. Let us consider the family of Jacobi fields η_t introduced in (5.1). As we already mentioned in the proof of Proposition 4.2, the classical Morse index theorem [36] allows to relate $\mathrm{ind}_\Omega(t \cdot \gamma)$ to the zeros of η_t by

$$\mathrm{ind}_\Omega(t \cdot \gamma) = \#\{s \in (t, t+1) \mid \eta_t(s) = 0\} = \#\{s \in (t-1, t) \mid \eta_t(s) = 0\}.$$

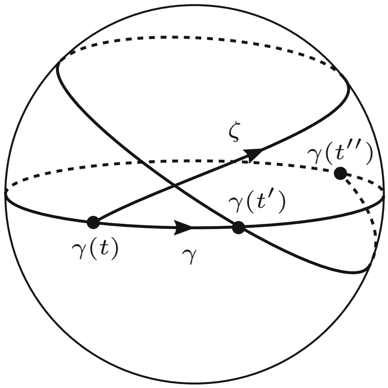


FIGURE 2. Example of geodesic ray ζ on a reversible Finsler sphere intersecting the simple closed geodesic γ at subsequent points $\gamma(t)$, $\gamma(t')$ and $\gamma(t'')$, with $t' = t + a' \in (t, t+1]$ and $t'' = t' + a'' \in (t', t' + 1]$, such that $i' = 1$ and $i'' = 0$

Therefore, $\mathrm{ind}_\Omega(t \cdot \gamma) \geq 2$ is equivalent to $t_2 - t < 1$ and $t - t_{-2} < 1$. If this holds for all $t \in \mathbb{R}$, we claim that the Birkhoff map ψ is a twist map. Indeed, ψ can be lifted to a continuous map

$$\widetilde{\psi} : \mathbb{R} \times [-1, 1] \to \mathbb{R} \times [-1, 1],$$

as follows. Let $\sigma : S^1 \times [-1, 1] \to \overline{A}$, $\sigma(t, s) = (\gamma(t), \nu(t, s))$ be the homeomorphism of Lemma 5.2. For each $(t, s) \in \mathbb{R} \times (-1, 1)$, we consider the geodesic ray ζ starting at $\zeta(0) = \gamma(t)$ with speed $\dot\zeta(0) = \nu(t, s)$. Let $a', a'' \in (0, 1]$ be such that the first intersection of ζ at positive time with γ is at $\gamma(t+a')$, and the second one is at $\gamma(t + a' + a'')$. Let $0 < b' < b''$ be the first positive times such that $\zeta(b') = \gamma(t+a')$ and $\zeta(b'') = \gamma(t+a'+a'')$. We denote by $i', i'' \in \mathbb{Z}$ the algebraic count of (transverse) self-intersections of the geodesics $\zeta|_{(0,b')}$ and $\zeta|_{(b',b'')}$, respectively (Fig. 2); here a double-point intersection is counted positively if and only if ζ crosses itself from left to right (up to isotoping $\zeta|_{[0,b'']}$ without moving $\zeta(0)$, $\zeta(b')$, and $\zeta(b'')$, we can assume that all the self-intersections of $\zeta|_{[0,b'']}$ are double points).

We define the lift

$$\widetilde{\psi}(t, s) = (a(t, s), b(t, s)),$$

by setting the first component to be

$$a(t, s) = t + a' + a'' + i' + i'' - 1.$$

It is straightforward to verify that such a function a is continuous. Since $\mathrm{ind}_\Omega(t \cdot \gamma) \geq 2$, if $|s|$ is close to 1 (that is, if $\dot\zeta(0)$ is close to $\pm\dot\gamma(0)$), we have $i' = i'' = 0$. Moreover, if s is close to 1 we have $a' + a'' \in (0, 1)$, whereas if s is close to -1 we have $a' + a'' \in (1, 2)$. Therefore,

$$a(t, 1) - t = t_2 - t - 1 \in (-1, 0), \quad a(t, -1) - t = 1 - (t - t_{-2}) \in (0, 1),$$

namely $\widetilde{\psi}$ satisfies the twist condition. Since ψ is a twist map, it has infinitely many periodic points corresponding to infinitely many closed geodesics of (S^2, F). $\square$

5.3. Hingston's theorems

A celebrated theorem due to Hingston [22], that extends previous results of Bangert [7,8], implies the existence of infinitely many closed geodesics on (S^2, F) when there is a simple closed geodesic with non-zero local homology in degree 3 and a Birkhoff map not of twist type. Hingston's original proof was phrased for Riemannian manifolds, but is valid as well in the Finsler setting, and indeed even in the non-reversible Finsler setting. We include the full argument here for the reader's convenience.

Theorem 5.4. *Let (M, F) be a closed Finsler manifold of dimension $d \geq 2$, and $\gamma \in \mathrm{crit}(E)$ a closed geodesic satisfying the following two conditions*:

(i) *The local homology $C_i(\gamma)$ with coefficient in some field is non-zero in degree $i = \mathrm{ind}(\gamma) + \mathrm{nul}(\gamma)$.*

(ii) $\mathrm{ind}(\gamma^m) + \mathrm{nul}(\gamma^m) \leq m(\mathrm{ind}(\gamma) + \mathrm{nul}(\gamma)) - (d-1)(m-1)$ *for all* $m \in \mathbb{N}$.

Then, (M, F) has infinitely many closed geodesics.

The proof of Theorem 5.4 will employ the following arithmetic statement, which is also due to Hingston.

Lemma 5.5. *Let (M, F) be a Finsler manifold, $\ell > 0$ a positive real number, and $\mathbb{K} \subseteq \mathbb{N}$ a k-dense subset for some $k > 0$, that is,*

$$(n-k, n+k) \cap \mathbb{K} \neq \varnothing, \quad \forall n \in \mathbb{N}.$$

Assume that for all $\epsilon > 0$, there exists $\overline{m} > 0$ and, for all $m \in \mathbb{K}$ with $m \geq \overline{m}$, a closed geodesic $\zeta_m \in \mathrm{crit}(E)$ whose length satisfies

$$m\ell < E(\zeta_m)^{1/2} \leq m\ell + \epsilon.$$

Then (M, F) has infinitely many closed geodesics.

Proof. Up to replacing the Finsler metric F with $\ell^{-1}F$, we can assume that

$$\ell = 1.$$

We recall that a closed geodesic $\gamma \in \mathrm{crit}(E)$ is called prime when it is not the iterate of another closed geodesic, that is, $\gamma = \zeta^m$ implies $m = 1$ and $\zeta = \gamma$. We prove the lemma by contradiction, by assuming that (M, F) has only finitely many prime closed geodesics $\gamma_1, \ldots, \gamma_r \in \mathrm{crit}(E)$. We denote by $\ell_i := E(\gamma_i)^{1/2}$ the length of such closed geodesics, and we reorder them so that $\ell_i \notin \mathbb{Q}$ for all $i = 1, \ldots, s$, and $\ell_i = p_i/q_i \in \mathbb{Q}$ for all $i = s+1, \ldots, r$. Here, p_i and q_i are positive integers.

If $s > 0$, that is, there are closed geodesics of irrational length, we set

$$c := \frac{(s+1)(2k-2)+1}{\min\{\ell_1, \ldots, \ell_s\}},$$

and

$$\delta_1 := \min\left\{ |m_1\ell_i - m_2| \;\middle|\; m_1, m_2 \in \mathbb{N} \text{ with } m_1 \leq c,\ i = 1, \ldots, s \right\} > 0.$$

If $s < r$, that is, there are closed geodesics of rational length, we have

$$\delta_2 := \min\Big\{ \big|m_1\ell_i - m_2\big| \ \Big|\ m_1, m_2 \in \mathbb{N} \text{ with } m_1\ell_i - m_2 \neq 0,\ i = s+1, \ldots, r \Big\} \\ \geq \frac{1}{\max\{q_{s+1}, \ldots, q_r\}} > 0.$$

If $s = 0$ we simply set $\delta_1 := +\infty$, and likewise if $s = r$ we set $\delta_2 := +\infty$.

Now, we fix $\epsilon \in (0, \min\{\delta_1, \delta_2, 1\})$. By the assumptions of the lemma, there exists $\overline{m}$ and, for all $m \in \mathbb{K}$ with $m \geq \overline{m}$, a closed geodesic $\zeta_m \in \mathrm{crit}(E)$ such that

$$m < E(\zeta_m)^{1/2} \leq m + \epsilon. \tag{5.4}$$

Since $\gamma_1, \ldots, \gamma_r$ are the only prime closed geodesics in (M, F), up to a shift in the parametrization each closed geodesic ζ_m must be of the form $\zeta_m = \gamma_i^\mu$ for some $i = i(m) \in \{1, \ldots, r\}$ and $\mu = \mu(m) \in \mathbb{N}$. In particular, the length of ζ_m is

$$E(\zeta_m)^{1/2} = \mu(m)\ell_{i(m)}.$$

We must have $i(m) \leq s$, that is, every $\gamma_{i(m)}$ must have irrational length $\ell_{i(m)}$; indeed, the inequality (5.4) implies

$$0 < \mu(m)\ell_{i(m)} - m < \epsilon < \delta_2.$$

Since $\mathbb{K}$ is k-dense in $\mathbb{N}$, for any $m_1 \in \mathbb{K}$, there exists $m_2 \in \mathbb{K}$ such that

$$m_1 < m_2 \leq m_1 + 2k - 2.$$

This, together with the fact that $\gamma_1, \ldots, \gamma_s$ are the only prime closed geodesics with irrational length, implies that we can find two integers $m_1, m_2 \in \mathbb{K}$ both larger than $\overline{m}$ and such that $m_1 < m_2 \leq m_1 + (s+1)(2k-2)$ and $i := i(m_1) = i(m_2)$. The inequalities in (5.4) applied to these two integers give

$$m_1 < \mu(m_1)\ell_i \leq m_1 + \epsilon, \quad m_2 < \mu(m_2)\ell_i \leq m_2 + \epsilon. \tag{5.5}$$

Therefore,

$$\big|(\mu(m_2) - \mu(m_1))\ell_i - (m_2 - m_1)\big| \leq \epsilon < \delta_1,$$

which implies

$$\mu(m_2) - \mu(m_1) > c$$

according to the definition of δ_1. This gives a contradiction, since the inequalities (5.5) imply

$$\mu(m_2) - \mu(m_1) < \frac{m_2 - m_1 + \epsilon}{\ell_i} < \frac{(s+1)(2k-2)+1}{\min\{\ell_1, \ldots, \ell_r\}} = c.$$

□

Proof of Theorem 5.4. Condition (ii), together with Lemma 4.1(ii,v), implies that

$$\mathrm{ind}_\Omega(\gamma) + \mathrm{nul}_\Omega(\gamma) \leq \tfrac{1}{m}\big(\mathrm{ind}_\Omega(\gamma^m) + \mathrm{nul}_\Omega(\gamma^m)\big) \\ \leq \mathrm{ind}(\gamma) + \mathrm{nul}(\gamma) - \tfrac{m-1}{m}(d-1).$$

In the limit $m \to \infty$ the latter term converges to $d-1$. This, together with the opposite inequality provided by Proposition 4.2(iv), implies

$$\mathrm{ind}_\Omega(\gamma) + \mathrm{nul}_\Omega(\gamma) = \mathrm{ind}(\gamma) + \mathrm{nul}(\gamma) - (d-1). \tag{5.6}$$

Let $\Sigma \subset M$ be an embedded hypersurface diffeomorphic to a compact $(d-1)$-dimensional disk intersecting in its interior the closed geodesic γ transversely at $\gamma(0)$. As in Sect. 4.3, we introduce the space

$$\Sigma_k := \{\zeta \in \Lambda_k \mid \zeta(0) \in \Sigma\},$$

for an integer k large enough so that $\gamma \in \Lambda_k$. Since we are looking for infinitely many closed geodesics, we can assume that γ is an isolated closed geodesic. Therefore, γ is an isolated critical point of the restricted energy functional $E|_{\Sigma_k}$ with non-trivial local homology

$$H_i(\Sigma_k^{<\ell} \cup \{\gamma\}, \Sigma_k^{<\ell}) \cong H_i(\Lambda^{<\ell} \cup \{\gamma\}, \Lambda^{<\ell}) = C_i(\gamma),$$

where $\ell^2 := E(\gamma)$.

Since $i = \mathrm{ind}(\gamma) + \mathrm{nul}(\gamma)$, by the Morse–Gromoll–Meyer lemma [18], there exists a smooth embedded ball $B \subset \Sigma_k$ of dimension i containing γ in its interior, and such that $E|_{B\setminus\{\gamma\}} < \ell^2$. Any such ball B represents a generator of the local homology $C_i(\gamma)$. This can be easily seen as follows. Let $N \subset \Sigma_k$ be a tubular neighborhood of B diffeomorphic to the normal bundle of B in Σ_k. The restriction of the energy functional E to any fiber F of N has a non-degenerate local minimizer at $F \cap B$. Thus, the local homology of E at γ is isomorphic to the local homology of $E|_B$ at its local maximizer γ, and the local homology at a local maximizer is generated by the relative cycle covering the whole domain (see, e.g., [33, Proposition 2.6] for a detailed proof of this general Morse-theoretic fact). This argument is independent of the choice of the coefficient field, and in particular $[B] \neq 0$ in $C_i(\gamma;\mathbb{Q})$ as well.

We consider the evaluation map $\mathrm{ev} : B \to \Sigma$, $\mathrm{ev}(\zeta) = \zeta(0)$, whose differential has the form

$$\mathrm{d}\,\mathrm{ev}(\gamma) : \mathrm{T}_\gamma B \to \mathrm{T}_{\gamma(0)}\Sigma, \quad \mathrm{d}\,\mathrm{ev}(\gamma)\xi = \xi(0).$$

We claim that $\mathrm{d}\,\mathrm{ev}(\gamma)$ is surjective. Indeed, if as usual $\Omega = \big\{\zeta \in \Lambda \mid \zeta(0) = \gamma(0)\big\}$ denotes the based loop space, we have

$$\ker(\mathrm{d}\,\mathrm{ev}(\zeta)) = \mathrm{T}_\gamma\Omega \cap \mathrm{T}_\gamma B.$$

Since the Hessian $\mathrm{d}^2E(\gamma)$ is negative semi-definite on $\mathrm{T}_\gamma B$, Eq. (5.6) implies

$$\begin{aligned}\dim\ker(\mathrm{d}\,\mathrm{ev}(\zeta)) &\leq \mathrm{ind}_\Omega(\gamma) + \mathrm{nul}_\Omega(\gamma)\\ &\leq \mathrm{ind}(\gamma) + \mathrm{nul}(\gamma) - (d-1)\\ &= \dim(B) - (d-1).\end{aligned}$$

Since $\dim(\Sigma) = d-1$, we infer that $\mathrm{d}\,\mathrm{ev}(\gamma)$ is surjective. By the implicit function theorem, up to shrinking B around γ, we find a diffeomorphism $\phi : \Sigma \times U \to B$ such that $\mathrm{ev} \circ \phi(x, y) = x$.

If $\zeta_i : [0, \tau_i] \to M$ are continuous paths such that $\zeta_i(\tau_i) = \zeta_{i+1}(0)$ for all $i \in \mathbb{Z}_m$, we define $\zeta := \zeta_0 * \cdots * \zeta_{m-1} \in \Lambda$ to be the 1-periodic curve obtained by first concatenating the ζ_i's with their original parametrization, and then by linearly reparametrizing the resulting curve so that it becomes 1-periodic. Namely,

$$\zeta(t) = \tilde{\zeta}((\tau_0 + \cdots + \tau_{m-1})t),$$

where

$$\tilde{\zeta}(\tau_0 + \cdots + \tau_{i-1} + u) = \zeta_i(u), \quad \forall u \in [0, \tau_i].$$

If the ζ_i's are $W^{1/2}$ paths, the energy of $\zeta_0 * \cdots * \zeta_{m-1}$ is

$$E(\zeta_0 * \cdots * \zeta_{m-1}) = (\tau_0 + \cdots + \tau_{m-1}) \sum_{i=0}^{m-1} \int_0^{\tau_i} F(\zeta_i, \dot{\zeta}_i)^2 \, \mathrm{d}t. \tag{5.7}$$

We now employ ϕ to construct a relative cycle representing a non-zero element of the local homology group of γ^m. We first define the smooth embedding

$$\phi_m : \Sigma \times U^{\times m} \hookrightarrow \Sigma_{mk}, \quad \phi_m(x, y_0, \ldots, y_{m-1}) = \phi(x, y_0) * \cdots * \phi(x, y_{m-1}),$$

where $U^{\times m} = U \times \cdots \times U$ denotes the m-fold Cartesian product. The fact that ϕ_m is a smooth embedding can be easily seen if we identify the loops $\zeta_i = \phi_i(x, y_i) \in \Sigma_k$ with the tuple $\boldsymbol{x}_i = (\zeta_i(0), \zeta_i(\frac{1}{k}), \ldots, \zeta_i(\frac{k-1}{k}))$ as explained in Sect. 4.1: indeed, the curve $\phi_m(x, y_0, \ldots, y_{m-1}) \in \Sigma_{mk}$ is then identified with the juxtaposition $(\boldsymbol{x}_0, \ldots, \boldsymbol{x}_{m-1})$. The image of ϕ_m is a smooth embedded ball

$$B_m := \phi_m(\Sigma \times U^{\times m}) \subset \Sigma_{mk}$$

containing γ^m in its interior. By assumption (ii) of the lemma, its dimension is bounded from below as

$$\dim(B_m) = d - 1 + m(i - (d-1)) \geq \mathrm{ind}(\gamma^m) + \mathrm{nul}(\gamma^m). \tag{5.8}$$

Since our ζ_i's are 1-periodic loops (that is, we consider them as closed paths parametrized on $[0, 1]$), Eq. (5.7) reduces to

$$E(\zeta_0 * \cdots * \zeta_{m-1}) = m\big(E(\zeta_0) + \cdots + E(\zeta_{m-1})\big).$$

Since $E(\zeta_i) < E(\gamma)$, we have

$$E|_{B_m \setminus \{\gamma^m\}} < E(\gamma^m) = m^2 \ell^2.$$

This, together with (5.8), implies that $\dim(B_m) = \mathrm{ind}(\gamma^m) + \mathrm{nul}(\gamma^m) =: i_m$. Therefore, as we explained above for B, the ball B_m represents a generator of the local homology group

$$H_{i_m}(\Sigma_{mk}^{<m\ell} \cup \{\gamma^m\}, \Sigma_{mk}^{<m\ell}; \mathbb{Q}) \cong C_{i_m}(\gamma^m; \mathbb{Q}).$$

We claim that, for each $\epsilon > 0$ sufficiently small, there exists $\overline{m} = \overline{m}_\epsilon \in \mathbb{N}$ such that, for all integers $m \geq \overline{m}$, the homomorphism

$$C_{i_m}(\gamma^m; \mathbb{Q}) \to H_{i_m}(\Lambda^{<m\ell + \epsilon/\ell}, \Lambda^{<m\ell}; \mathbb{Q})$$

induced by the inclusion is the zero one. Indeed, let $\epsilon \in (0, 1)$ be small enough so that

$$\max_{\Sigma \times \partial U} E \circ \phi < \ell^2 - \epsilon.$$

If needed, we shrink Σ around $\gamma(0)$ so that

$$\mathrm{diam}(\Sigma) := \max_{x_1,x_2\in\Sigma} d(x_1,x_2) < \frac{\epsilon}{2(\ell^2+2)},$$

where $d : M \times M \to [0,\infty)$ denotes the (possibly non-symmetric) distance (4.1) induced by the Finsler metric F. Let $\delta > 0$ be such that

$$\max_{\partial\Sigma\times U} E\circ\phi = \ell^2 - \delta,$$

and notice that

$$\max_{\partial\Sigma\times U^{\times m}} E\circ\phi_m = m^2(\ell^2-\delta).$$

We define the continuous map

$$\psi_m : \Sigma\times\Sigma\times U^{\times\lfloor m/2\rfloor}\times U^{\times\lceil m/2\rceil} \to \Lambda,$$
$$\psi_m(x_1,x_2,y_1,y_2) = \phi_{\lfloor m/2\rfloor}(x_1,y_1) * \gamma_{x_1x_2} * \phi_{\lceil m/2\rceil}(x_2,y_2) * \gamma_{x_2x_1},$$

where $\gamma_{x_ix_j} : [0, d(x_i,x_j)] \to M$ is the shortest geodesic parametrized with unit speed joining x_i and x_j. Let us compute the composition $E\circ\psi_m$. If we set

$$\zeta_1 := \phi_{\lfloor m/2\rfloor}(x_1,y_1), \quad \zeta_2 := \phi_{\lceil m/2\rceil}(x_2,y_2), \quad \zeta := \psi_m(x_1,x_2,y_1,y_2),$$

we have

$$\begin{aligned} E(\zeta) &= \big(m + d(x_1,x_2) + d(x_2,x_1)\big)\left(\frac{E(\zeta_1)}{\lfloor\frac{m}{2}\rfloor} + d(x_1,x_2) + \frac{E(\zeta_2)}{\lceil\frac{m}{2}\rceil} + d(x_2,x_1)\right) \\ &\le \big(m + 2\,\mathrm{diam}(\Sigma)\big)\big(m\ell^2 + 2\,\mathrm{diam}(\Sigma)\big) \\ &< m^2\ell^2 + 2\,\mathrm{diam}(\Sigma)(m + \ell^2 m + 2\,\mathrm{diam}(\Sigma)) \\ &< m^2\ell^2 + m\epsilon < (m\ell + \epsilon/\ell)^2. \end{aligned}$$

If $y_1 \in \partial U^{\times\lfloor m/2\rfloor}$ or $y_2 \in \partial U^{\times\lceil m/2\rceil}$, we have the estimate

$$\begin{aligned} E(\zeta) &< \big(m + 2\,\mathrm{diam}(\Sigma)\big)\big(m\ell^2 - \epsilon + 2\,\mathrm{diam}(\Sigma)\big) \\ &= m^2\ell^2 + m\big(2\,\mathrm{diam}(\Sigma)(1+\ell^2) - \epsilon\big) + 2\,\mathrm{diam}(\Sigma)\big(2\,\mathrm{diam}(\Sigma) - \epsilon\big) \\ &< m^2\ell^2. \end{aligned}$$

If instead $x_1 \in \partial\Sigma$ or $x_2 \in \partial\Sigma$, we have

$$\begin{aligned} E(\zeta) &\le \big(m + 2\,\mathrm{diam}(\Sigma)\big)\big(m\ell^2 - \lfloor\tfrac{m}{2}\rfloor\delta + 2\,\mathrm{diam}(\Sigma)\big) \\ &\le m^2\ell^2 + \big(2\,\mathrm{diam}(\Sigma)(2+\ell^2) - \lfloor\tfrac{m}{2}\rfloor\delta\big)m \\ &\le m^2\ell^2 + \underbrace{\big(\epsilon - \lfloor\tfrac{m}{2}\rfloor\delta\big)}_{(*)}m, \end{aligned}$$

and the term $(*)$ is negative for $m \ge \overline{m} = 2\epsilon/\delta + 2$. Summing up, our map ψ_m satisfies

$$E\circ\psi_m|_{\partial(\Sigma\times\Sigma\times U^m)} < m^2\ell^2. \tag{5.9}$$

The relative cycle $\mathrm{diag}_\Sigma \times U^m$ is null-homologous in

$$H_*(\Sigma\times\Sigma\times U^m, \partial(\Sigma\times\Sigma\times U^m); \mathbb{Q}),$$

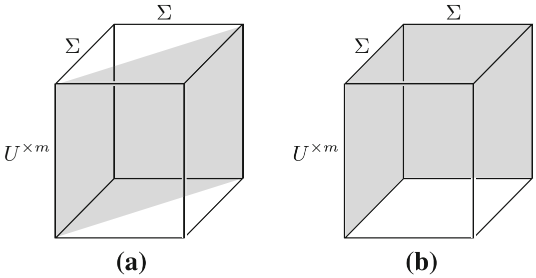


FIGURE 3. The shaded region in **a** is the relative cycle $\mathrm{diag}_\Sigma \times U^m$, which is null-homologous in $H_*(\Sigma \times \Sigma \times U^{\times m}, \partial(\Sigma \times \Sigma \times U^{\times m}); \mathbb{Q})$, for instance because it is homologous to the shaded region in **b**

since it is homologous to a relative cycle contained in $\partial(\Sigma \times \Sigma \times U^m)$, see Fig. 3. Therefore, (5.9) implies that the relative cycle $B_m = \phi_m(\Sigma \times U^m) = \psi_m(\mathrm{diag}_\Sigma \times U^m)$ is null-homologous in $H_{i_m}(\Lambda^{<m\ell+\epsilon/\ell}, \Lambda^{<m\ell}; \mathbb{Q})$, i.e.,

$$[B_m] = 0 \text{ in } H_{i_m}(\Lambda^{<m\ell+\epsilon/\ell}, \Lambda^{<m\ell}; \mathbb{Q}). \tag{5.10}$$

From now on, we assume that our integer $m \geq \overline{m}$ is odd, and employ Morse theory. Since m is odd, Lemma 4.4 implies that the inclusion induces an injective homomorphism $C_*(\gamma^m; \mathbb{Q}) \hookrightarrow C_*(S^1 \cdot \gamma^m; \mathbb{Q})$. Therefore, the commutative diagram

$$\begin{tikzcd} C_{i_m}(\gamma^m; \mathbb{Q}) \arrow[r, hook] \arrow[dr, "0"'] & C_{i_m}(S^1 \cdot \gamma^m; \mathbb{Q}) \arrow[d, "\iota_*"] \\ & H_{i_m}(\Lambda^{<m\ell+\epsilon/\ell}, \Lambda^{<m\ell}; \mathbb{Q}) \end{tikzcd}$$

whose homomorphisms are all induced by the inclusion, implies that ι_* is not injective. This, in turn, implies that there exists a closed geodesic $\zeta_m \in \mathrm{crit}(E)$ of length $E(\zeta_m)^{1/2} \in (m\ell, m\ell + \epsilon/\ell]$.

Summing up, we showed that for every $\epsilon > 0$ small enough there exists $\overline{m} > 0$ and, for all odd integers $m \geq \overline{m}$, a closed geodesic $\zeta_m \in \mathrm{crit}(E)$ such that $m\ell < E(\zeta_m)^{1/2} \leq m\ell + \epsilon$. We can now invoke Lemma 5.5 with $\mathbb{K}$ being the set of odd positive integers, and conclude that (M, F) has infinitely many closed geodesics. $\square$

We now derive the two corollaries that we will need for proving Theorem 1.5.

Corollary 5.6. *Let (M, F) be an orientable Finsler surface, and $\gamma \in \mathrm{crit}(E)$ a closed geodesic such that $\mathrm{ind}_\Omega(\gamma) = \mathrm{nul}_\Omega(\gamma) = 1$ and the local homology $C_3(\gamma)$ with some coefficient field is non-zero. Then, (M, F) has infinitely many closed geodesics.*

Proof. Since $C_3(\gamma)$ is non-trivial, we have

$$\mathrm{ind}(\gamma) \leq 3 \leq \mathrm{ind}(\gamma) + \mathrm{nul}(\gamma).$$

We now employ Proposition 4.2. Since $\mathrm{nul}_\Omega(\gamma)=1$, we have

$$\mathrm{ind}_\Omega(\gamma^m)=m\,\mathrm{ind}_\Omega(\gamma)+(m-1)\mathrm{nul}_\Omega(\gamma)=2m-1,\quad \mathrm{nul}_\Omega(\gamma^m)=1.$$

Moreover

$$\mathrm{ind}(\gamma)\le \mathrm{ind}_\Omega(\gamma)+1=2,\quad \mathrm{nul}(\gamma)\ge 3-\mathrm{ind}(\gamma)\ge 1.$$

If $\mathrm{nul}(\gamma)=2$, then $\mathrm{nul}(\gamma^m)=2$ and $\mathrm{ind}(\gamma^m)$ is odd for all $m\in\mathbb{N}$; since $\mathrm{ind}_\Omega(\gamma^m)\le \mathrm{ind}(\gamma^m)\le \mathrm{ind}_\Omega(\gamma^m)+1$, we infer

$$\mathrm{ind}(\gamma^m)=\mathrm{ind}_\Omega(\gamma^m)=2m-1.$$

If instead $\mathrm{nul}(\gamma)=1$, then $\mathrm{ind}(\gamma)=2$, and

$$\begin{aligned}\mathrm{ind}(\gamma^m)+\mathrm{nul}(\gamma^m)&\le \mathrm{ind}_\Omega(\gamma^m)+\mathrm{nul}_\Omega(\gamma^m)+1\\&=2m+1\\&=m(\mathrm{ind}(\gamma)+\mathrm{nul}(\gamma))-(m-1).\end{aligned}$$

In both cases, γ satisfies the assumptions of Theorem 5.4, and we infer that (S^2,F) has infinitely many closed geodesics. □

The second corollary of Theorem 5.4 was established in the Riemannian case by Bangert [7,8]. Even though we present it here as a corollary of Theorem 5.4, Bangert's proof came historically earlier than [22]. Let us recall, once again, the classical notion of conjugate points: two points $\gamma(t)$ and $\gamma(s)$ along a geodesic $\gamma:[t,s]\to M$ are conjugate when there exists a Jacobi field along γ that is not identically zero, but vanishes at both $\gamma(t)$ and $\gamma(s)$. When $\dim(M)=2$ this condition can be expressed in terms of the Jacobi field η_t introduced in Eq. (5.1): $\gamma(t)$ and $\gamma(s)$ are conjugate points if and only if $\eta_t(s)=0$. A closed geodesic γ on a Finsler surface has no conjugate points if and only if $\mathrm{ind}_\Omega(t\cdot\gamma^m)=\mathrm{nul}_\Omega(t\cdot\gamma^m)=0$ for all $t\in S^1$ and $m\in\mathbb{N}$; equivalently, $\mathrm{ind}(\gamma^m)=0$ for all $m\in\mathbb{N}$, according to Lemma 4.1(iii) and Proposition 4.2(iii).

Corollary 5.7. *Any reversible Finsler 2-sphere with a simple closed geodesic without conjugate points possesses infinitely many closed geodesics.*

Proof. Let $\gamma\in\mathrm{crit}(E)\cap E^{-1}(\ell^2)$ be a simple closed geodesic without conjugate points in the reversible Finsler 2-sphere (S^2,F). We claim that there exists a neighborhood $\mathcal{U}\subset\Lambda$ of the critical circle $S^1\cdot\gamma:=\{t\cdot\gamma\mid t\in S^1\}$ such that every $\zeta\in\mathcal{U}$ that intersects the curve γ has energy $E(\zeta)\ge E(\gamma)$. Indeed, if this were not true, we could find a sequence $\zeta_n\in\Lambda$ such that $\zeta_n(0)=\gamma(t_n)$, $E(\zeta_n)<E(\gamma)$, $t_n\to t$ and $\zeta_n\to t\cdot\gamma$ as $n\to\infty$. We consider the based loop spaces

$$\Omega_s:=\{\zeta\in\Lambda\mid \zeta(0)=\gamma(s)\},\quad s\in S^1,$$

and the space of broken closed geodesics Λ_k introduced in Sect. 4.1. Here $k\in\mathbb{N}$ must be large enough so that $\gamma\in\Lambda_k$. Since γ has no conjugate points, $\mathrm{ind}_\Omega(s\cdot\gamma)=\mathrm{nul}_\Omega(s\cdot\gamma)=0$. Therefore, every $s\cdot\gamma$ is a non-degenerate local minimizer of $E|_{\Lambda_k\cap\Omega_s}$. Since $E_k:=E|_{\Lambda_k}$ is smooth in a neighborhood of the critical circle of γ, we can apply the parametric Morse lemma, which provides an $\epsilon>0$ and an open neighborhood $U\subset\Lambda_k$ of γ such that, for all

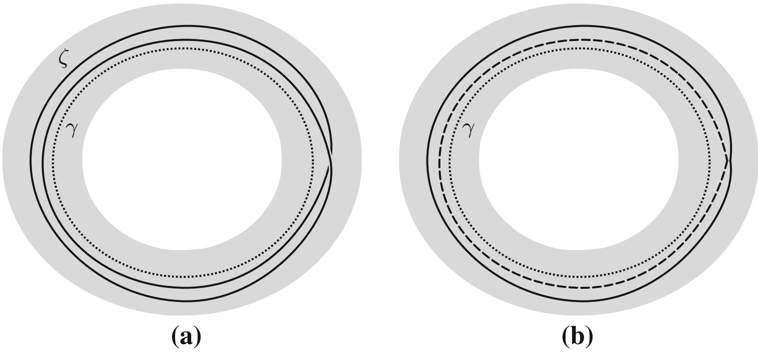


FIGURE 4. **a** The simple closed geodesic γ (dotted), and a curve $\zeta \in \Lambda$ close to γ^2 with a self-intersection. **b** The support of the curve ζ can be decomposed as the union of ζ_1 (dashed curve) and ζ_2 (solid curve), both close to γ

$t \in (-\epsilon, \epsilon)$, $t \cdot \gamma$ is the unique global minimizer of $E|_{U\cap\Omega_t}$. Let $\gamma_n \in \Lambda_k \cap \Omega_{t_n}$ be the sequence of broken closed geodesics such that $\gamma_n(i/k) = \zeta_n(i/k)$ for all $i \in \mathbb{Z}_k$, which have energy $E(\gamma_n) \le E(\zeta_n) < E(\gamma)$. Since $\zeta \to t\cdot\gamma$ in Λ, we would have that $\gamma_n \to \gamma$ in Λ_k, and in particular $\gamma_n \in U$ for all $n \in \mathbb{N}$ large enough, contradicting the fact that $E|_{U\cap\Omega_{t_n}}$ has a strict global minimizer at $t_n \cdot \gamma$.

We denote by B_0 and B_1 the connected components of $S^2\setminus\gamma$, and by $\mathcal{B}_i \subset \Lambda$ the open subset of those $\zeta \in \Lambda$ such that $\zeta(S^1) \subset B_i$. Since we are looking for infinitely many closed geodesics, we can assume that γ is an isolated closed geodesic (i.e., the critical circle of each iterate γ^m is isolated in $\mathrm{crit}(E)$). We set $\ell^2 := E(\gamma)$. We have two possible cases, which we deal with separately.

Case 1: For every open neighborhood $\mathcal{V} \subset \Lambda$ of γ, the intersections $\mathcal{V}\cap\mathcal{B}_0\cap\Lambda^{<\ell}$ and $\mathcal{V}\cap\mathcal{B}_1\cap\Lambda^{<\ell}$ are both non-empty. If we choose $\mathcal{V}$ contained in the above open subset $\mathcal{U}$, we infer that every connected component of $\mathcal{V}^{<\ell} := \mathcal{V}\cap\Lambda^{<\ell}$ is contained in either $\mathcal{V}^{<\ell}\cap\mathcal{B}_0$ or $\mathcal{V}^{<\ell}\cap\mathcal{B}_1$. In particular $\mathcal{V}^{<\ell}$ is not connected. Since $\mathcal{V}$ can be chosen arbitrarily small, Lemma 4.3 implies that the local homology $C_1(\gamma;\mathbb{Q})$ is non-zero. Since $\mathrm{ind}(\gamma^m) = 0$, Proposition 4.2(vi) implies that $\mathrm{nul}(\gamma^m) < 2$ for all $m \in \mathbb{N}$. Since the local homology $C_1(\gamma;\mathbb{Q})$ is non-zero, we must have $\mathrm{ind}(\gamma) + \mathrm{nul}(\gamma) \ge 1$, and thus $\mathrm{nul}(\gamma) = 1$. Since $\mathrm{nul}(\gamma) \le \mathrm{nul}(\gamma^m) < 2$, we infer that $\mathrm{nul}(\gamma^m) = 1$ for all $m \in \mathbb{N}$. Therefore, γ satisfies the assumptions of Theorem 5.4, which implies that (S^2, F) has infinitely many closed geodesics.

Case 2: For some $i \in \{0, 1\}$, there exists an open neighborhood $\mathcal{V} \subset \Lambda$ of γ such that $\mathcal{V}\cap\mathcal{B}_i\cap\Lambda^{<\ell} = \varnothing$. This implies the analogous property for γ^m: there exists an open neighborhood $\mathcal{V}_m \subset \Lambda$ of γ^m such that $\mathcal{V}_m\cap\mathcal{B}_i\cap\Lambda^{<\ell} = \varnothing$. Indeed, since we are on an orientable surface, a tubular neighborhood of the simple closed geodesic γ is diffeomorphic to the annulus $S^1 \times (-1, 1)$, γ being its zero section $S^1 \times \{0\}$. Therefore, any curve ζ sufficiently close to the iterated curve γ^m has at least $m - 1$ self-intersections counted with multiplicity (see Fig. 4). The support of ζ can be decomposed as the union of the supports of $\zeta_1, \ldots, \zeta_m \in \mathcal{V}$, and $E(\zeta) = m(E(\zeta_1) + \cdots + E(\zeta_m))$, see for instance [2, Lemma 4.2]. If $E(\zeta) < E(\gamma^m) = m^2 E(\gamma)$, we would have $E(\zeta_j) < E(\gamma)$ for some j, contradicting the fact that $\mathcal{V} \cap \mathcal{B}_i \cap \Lambda^{<\ell} = \varnothing$.

Since γ is an isolated closed geodesic, we can choose $\mathcal{V}_m$ to be small enough so that $\mathcal{V}_m \cap \mathcal{B}_i \cap \mathrm{crit}(E) = \varnothing$. This implies that

$$E(\zeta) > E(\gamma^m), \quad \forall \zeta \in \mathcal{V}_m \cap \mathcal{B}_i.$$

Indeed, by the previous paragraph we know that $E(\zeta) \geq E(\gamma^m)$ for any $\zeta \in \mathcal{V}_m \cap \mathcal{B}_i$. If we had equality $E(\zeta) = E(\gamma^m)$ then ζ would be a local minimizer, and in particular $\zeta \in \mathrm{crit}(E)$.

Let us fix a homotopy $u : [0,1] \to \mathcal{B}_i \cup \{\gamma^m\}$, $u(t) = u_t$, such that $u_0 = \gamma^m$ and $E(u_1) = 0$; namely, u_t defines a contraction of γ to a point within the disk B_i. We choose an integer

$$k > \frac{\max\{E \circ u\}}{\mathrm{injrad}(S^2, F)^2},$$

and consider the space of broken closed geodesics Λ_k and the restricted energy functional $E_k = E|_{\Lambda_k}$. We have the associated retraction

$$r : \Lambda^{<\sqrt{k}\,\mathrm{injrad}(S^2,F)} \to \Lambda_k, \quad r(\zeta) = \tilde{\zeta},$$

where $\tilde{\zeta} \in \Lambda_k$ is the broken closed geodesics such that $\tilde{\zeta}(i/k) = \zeta(i/k)$ for all $i \in \mathbb{Z}_k$. We recall that $E_k \circ r \leq E$. Since the boundary of the disk B_i is geodesic, we readily see that r preserves $\mathcal{B}_i$. Consider an open neighborhood $V \subset \Lambda_k$ of γ^m with compact closure $\overline{V} \subset \mathcal{V}_m$. Since ∂V is compact, we have

$$b^2 := \min_{\partial V \cap \mathcal{B}_i} E_k > E_k(\gamma^m).$$

Let $W \subset V$ be a small enough open neighborhood of γ^m such that

$$E_k(\gamma^m) < a^2 := \max_{\overline{W}} E_k < b^2.$$

We consider

$$c^2 := \inf_v \max\{E_k \circ v\},$$

where the infimum ranges over all homotopies $v : [0,1] \to \Lambda_k \cap \mathcal{B}_i \cup \{\gamma^m\}$, $v(t) = v_t$, such that $v_0 = \gamma^m$ and $E_k(v_1) = 0$. Notice that the space of such homotopies is non-empty, as it contains $r \circ u$. We have $c \geq b$, since every such a homotopy v must eventually intersect $\partial V \cap \mathcal{B}_i$. We fix an arbitrary $d \in (c, k\,\mathrm{injrad}(S^2,F)^{1/2})$. Notice that $E_k^{-1}[a^2,d^2] \cap \mathcal{B}_i$ is compact, since it is a closed subset of the compact set $E_k^{-1}[a^2,d^2] \setminus W$. Therefore, the classical min-max theorem implies that c^2 is a critical value of E_k. Since we are looking for infinitely many closed geodesics, we can assume that (S^2, F) has only isolated closed geodesic (i.e., any critical circle is isolated in $\mathrm{crit}(E) \cap E^{-1}(0,\infty)$). Under this assumption, there exists at least one closed geodesic $\zeta_m \in \mathrm{crit}(E_k) \cap E_k^{-1}(c^2) \cap \mathcal{B}_i$ such that every open neighborhood $Z \subset \Lambda_k \cap \mathcal{B}_i$ of ζ_m has a non-connected intersection $Z^{<c} = Z \cap \Lambda_k^{<c}$. Indeed, if no closed geodesic in $\mathrm{crit}(E_k) \cap E_k^{-1}(c^2) \cap \mathcal{B}_i$ satisfied this property, we

could find a homotopy v as above such that $\max\{E_k \circ v\} < c^2$, contradicting the definition of the min-max value c^2. Lemma 4.3 implies that the local homology $C_1(\zeta_m; \mathbb{Q})$ is non-trivial.

Now, either the family $\{\zeta_m \mid m \in \mathbb{N}\}$ that we found contains infinitely many geometrically distinct closed geodesics, or there exists a non-iterated closed geodesic ζ, an infinite subset $\mathbb{K} \subset \mathbb{N}$, and a function $\mu : \mathbb{K} \to \mathbb{N}$ such that $\mu(m) \to \infty$ as $m \to \infty$ and $\zeta_m = \zeta^{\mu(m)}$ for all $m \in \mathbb{N}$. Since every iterate $\zeta^{\mu(m)}$ has non-trivial local homology $C_1(\zeta^{\mu(m)})$, we have $\mathrm{ind}(\zeta^{\mu(m)}) \leq 1$ for all $m \in \mathbb{K}$, and therefore, $\mathrm{ind}(\zeta^m) = 0$ for all $m \in \mathbb{N}$ according to Lemma 4.1(iii). We cannot have $\mathrm{nul}(\zeta) = 0$, for otherwise ζ would be a local minimizer of E_k, and the same would be true for all its iterates according to analogous argument of Fig. 4. Therefore, $1 \leq \mathrm{nul}(\zeta) \leq \mathrm{nul}(\zeta^m) \leq 2$ for all $m \in \mathbb{N}$. By Proposition 4.2(vi), since the Morse indices $\mathrm{ind}(\zeta^m)$ vanish, we must have $\mathrm{nul}(\zeta^m) = 1$ for all $m \in \mathbb{N}$.

Since $\mathrm{ind}(\zeta) = \mathrm{ind}(\zeta^m)$ and $\mathrm{nul}(\zeta) = \mathrm{nul}(\zeta^m)$ for all $m \in \mathbb{N}$, we have an isomorphism of local homology groups $C_*(\zeta; \mathbb{Q}) \cong C_*(\zeta^m; \mathbb{Q})$. In particular $C_1(\zeta; \mathbb{Q})$ does not vanish. Therefore, ζ satisfies the assumptions of Theorem 5.4, which implies that (S^2, F) has infinitely many closed geodesics. □

5.4. Bangert's theorem

As it turns out, the statements proved so far allow us to conclude the existence of infinitely many closed geodesics on any reversible (S^2, F), except when none of its simple closed geodesics has a well-defined Birkhoff map. We recall that a simple closed geodesic γ of a reversible (S^2, F) does not have a well-defined Birkhoff map when, for some $x = \gamma(t)$ and $v \in \mathrm{T}_x S^2$ transverse to $\dot\gamma(t)$, the geodesic $\zeta(t) = \exp_x(tv)$ does not intersect γ at any positive time $t > 0$. In this section, we show that this last case is covered by Corollary 5.7. For Riemannian 2-spheres, this is a theorem due to Bangert [8].

Theorem 5.8. *Any reversible Finsler 2-sphere having a simple closed geodesic without a well-defined Birkhoff map possesses infinitely many closed geodesics.*

The proof is based on the following lemma of independent interest.

Lemma 5.9. *Let (M, F) be a (not necessarily reversible) Finsler surface, and $\gamma : [-T, T] \to M$ a geodesic parametrized with constant speed. If there exists a sequence of geodesics $\gamma_n : [-T, T] \to M$ parametrized with constant speed, not intersecting γ, and such that $(\gamma_n(0), \dot\gamma_n(0)) \to (\gamma(0), \dot\gamma(0))$ in $\mathrm{T}M$, then $\gamma|_{(-T,T)}$ has no conjugate points.*

Proof. Since the problem is local to γ, we can assume without loss of generality that $M = \mathbb{R}^2$ and $\gamma(t) = (t, 0)$ for all $t \in [-T, T]$, so that we can write expressions in coordinates. Without loss of generality, we can assume that $F(\gamma, \dot\gamma) \equiv 1$. We reparametrize the geodesics γ_n so that they have speed $F(\gamma_n, \dot\gamma_n) \equiv 1$. By doing this, we change the interval of definition of γ_n: the reparametrized curve has the form $\gamma_n : [-T_n, T_n] \to M$ with $T_n \to T$ as $n \to \infty$.

We set $(x,v)=(\gamma(0),\dot\gamma(0))=(0,\dot\gamma(0))$, $G=\frac{1}{2}F^2$, and consider the line

$$\Sigma := \{y\in\mathbb{R}^2 \mid y\in\ker G_v(x,v)\}.$$

For all $n\in\mathbb{N}$ large enough, the geodesic $\gamma_n:[-T_n,T_n]\to\mathbb{R}^2$ intersects Σ in a unique point x_n, and clearly $x_n\to x$. We shift the parametrization of γ_n, so that we have a sequence of geodesics $\gamma_n:[-T_n+\epsilon_n,T_n+\epsilon_n]\to\mathbb{R}^2$ not intersecting γ and such that $\epsilon_n\to 0$ and $v_n:=\dot\gamma_n(0)\to v$. Up to extracting a subsequence, we can assume that each γ_n lies on the same side of γ. Therefore, we have a well-defined non-zero vector

$$w' := \frac{x_n-x}{\|x_n-x\|}\in\Sigma$$

independent of n. Here, $\|\cdot\|$ denotes the Euclidean norm. We consider the vectors

$$z_n := \frac{v_n-v}{\|x_n-x\|}.$$

Since the geodesics γ_n and γ do not intersect on the time interval $[-T/2,T/2]$ and the second derivative of $\gamma_n-\gamma$ is uniformly bounded on $[-T/2,T/2]$ independently of n, we readily obtain that the sequence $\|z_n\|$ is uniformly bounded from above. In particular, up to extracting a subsequence, we have $z_n\to z'$ as $n\to\infty$.

We set $\lambda:=(1+\|z'\|^2)^{-1/2}$, $w:=\lambda w'$, and $z:=\lambda z'$, so that

$$\lim_{n\to\infty}\frac{(x_n,v_n)-(x,v)}{\|(x_n,v_n)-(x,v)\|}=\lim_{n\to\infty}\frac{(x_n,v_n)-(x,v)}{\|x_n-x\|\sqrt{1+\|z_n\|^2}}=(w,z)\in \mathrm{T}_{(x,v)}S\mathbb{R}^2,$$

where $S\mathbb{R}^2=\{(x',v')\in\mathrm{T}\mathbb{R}^2 \mid F(x',v')=1\}$. Therefore,

$$\lim_{n\to\infty}\frac{\phi_t(x_n,v_n)-\phi_t(x,v)}{\|(x_n,v_n)-(x,v)\|}=\mathrm{d}\phi_t(x,v)(w,z).$$

The Jacobi field $\zeta:(-T,T)\to\mathbb{R}^2$ along γ defined by

$$(\zeta(t),\dot\zeta(t))=\mathrm{d}\phi_t(x,v)(w,z)$$

satisfies $G_v(\gamma(t),\dot\gamma(t))\zeta(t)=0$.

We claim that ζ is nowhere vanishing. Indeed, let $\nu:(-T,T)\to\mathbb{R}^2$ the smooth vector field along γ defined by $\nu(t)\in\ker G_v(\gamma(t),\dot\gamma(t))$, $\|\nu(t)\|=1$, and $\nu(t)$ pointing to the side of γ containing the γ_n's. We can write $\zeta(t)=z(t)\nu(t)$ for some continuous function $z=(-T,T)\to\mathbb{R}$. Notice that

$$\zeta(t)=\lim_{n\to\infty}\frac{\gamma_n(t)-\gamma(t)}{\|(x_n,v_n)-(x,v)\|}.$$

If $\zeta(t)=0$ for some t, since $\gamma_n(t)-\gamma(t)$ and $\nu(t)$ point to the same side of γ, we readily obtain that $z(t)=0$ and $\dot z(t)=0$. But this would imply that $\zeta(t)=\dot\zeta(t)=0$, and since ζ is a Jacobi field we would conclude that ζ vanishes identically, contradicting $\zeta(0)=w$.

If γ had conjugate points $\gamma(t_1),\gamma(t_2)$ for some $-T<t_1<t_2<T$, there would exists a Jacobi field $\eta:[t_1,t_2]\to\mathbb{R}^2$ such that $\eta(t_1)=\eta(t_2)=0$, $\dot\eta(t_1)\neq 0$, and $\eta(t)\in\ker G_v(\gamma(t),\dot\gamma(t))$ for all $t\in[t_1,t_2]$. Since we are on a surface, $\ker G_v(\gamma(t),\dot\gamma(t))$ is 1-dimensional. Therefore, by the Sturm

separation theorem, η and ζ would have alternating zeroes, contradicting the fact that ζ is nowhere vanishing. □

Proof of Theorem 5.8. Let $\gamma_0 \in \mathrm{crit}(E) \cap E^{-1}(0,\infty)$ be a simple closed geodesic that does not have a well-defined Birkhoff map. We only need to consider the case in which γ_0 has conjugate points (i.e., $\mathrm{ind}_\Omega(\gamma_0^m) > 0$ for some integer $m \geq 1$), for otherwise the existence of infinitely many closed geodesics is already provided by Corollary 5.7. The fact that γ_0 does not have a well-defined Birkhoff map means that, for some $x_0 = \gamma_0(t_0)$ and $v_0 \in S_x S^2$ transverse to $\dot\gamma_0(t_0)$, the geodesic $\zeta : (0,\infty) \to S^2$, $\zeta(t) = \exp_{x_0}(tv_0)$ does not intersect γ_0 in positive time, and therefore, stays trapped in a connected component $B \subset S^2 \backslash \gamma_0(S^1)$. We consider the compact subset

$$K := \bigcap_{t>0} \overline{\zeta[t,\infty)} \subset \overline{B}.$$

We claim that

$$K \cap \gamma_0(S^1) = \varnothing.$$

Otherwise we can find a sequence $t_n \to \infty$ and $s \in \mathbb{R}$ such that $\zeta(t_n) \to \gamma_0(s)$. Since ζ does not intersect γ_0 in positive time, up to extracting a subsequence we must have $\dot\zeta(t_n) \to \dot\gamma_0(s)$. Since γ_0 has conjugate points, there exists $\delta > 0$ such that $\gamma_0|_{(s-\delta,s+\delta)}$ has conjugate points. Lemma 5.9 thus provides a contradiction: since $\zeta|_{[t_n-\delta,t_n+\delta]}$ does not intersect $\gamma_0|_{[s-\delta,s+\delta]}$, $\gamma_0|_{(s-\delta,s+\delta)}$ cannot have conjugate points.

Let $U \subset B\backslash K$ be the connected component whose closure contains $\gamma_0(S^1)$. One would expect this open set to be locally geodesically convex. We prove a slightly weaker convexity: for all $x, y \in U$ that can be joined by means of an absolutely continuous curve in U of length strictly less than $\rho := \mathrm{injrad}(S^2, F)$, the shortest geodesic joining x and y is entirely contained in U. Indeed, let $\gamma_{x,y} : [0,1] \to S^2$, $\gamma_{x,y}(t) = \exp_x(t \exp_x^{-1}(y))$ be such a geodesic, and assume by contradiction that some $z = \gamma_{x,y}(s)$ belongs to K. Then, by the definition of K, there exists a sequence $t_n \to \infty$ such that $\zeta(t_n) \to z$. Up to extracting a subsequence, the sequence $\dot\zeta(t_n)$ converges to some $w \in S_z S^2$ that is transverse to $\dot\gamma_{x,y}(s)$, since the geodesic $\theta : \mathbb{R} \to S^2$, $\theta(t) = \exp_z(tw)$ is entirely contained in K. We denote the geodesic balls centered at z by

$$B(z,r) := \big\{ z' \in S^2 \bigm| d(z,z') < r \big\}, \quad r > 0,$$

where $d : S^2 \times S^2 \to [0,\infty)$ is the distance (4.1) induced by the Finsler metric F. The points x and y are contained in different connected components of $B(z,\rho)\backslash\zeta(-\rho,\rho)$. Therefore, every continuous curve $\theta : [0,1] \to U$ such that $\theta(0) = x$ and $\theta(1) = y$ must leave the geodesic ball $B(z,\rho)$; since $d(x,z) + d(z,y) < \rho$, we readily obtain that the length of such a θ is larger than ρ, contradicting our assumption on x, y.

We consider the space

$$W := \big\{ \gamma \in \Lambda \bigm| \gamma(S^1) \subset U,\ \gamma \text{ not contractible in } U,\ E(\gamma) < E(\gamma_0) \big\}.$$

We claim that W is not empty. Indeed, since γ_0 has conjugate points, by Proposition 4.2(vii) there exists a nowhere-vanishing 1-periodic vector field ξ along γ such that $\mathrm{d}^2E(\gamma)[\xi,\xi]<0$, and $\xi(t)$ points inside U for all $t\in S^1$. We define $\gamma_s\in\Lambda$ by

$$\gamma_s(t)=\exp_{\gamma_0(t)}(s\xi(t)).$$

If $s>0$ is small enough, then γ_s is contained in U, non-contractible in U (since it is homotopic to γ_0 within $U\cup\gamma_0(S^1)$), and since

$$E(\gamma_s)\le E(\gamma_0)+\tfrac{1}{2}s^2\,\mathrm{d}^2E(\gamma)[\xi,\xi]+o(s^2)$$

we have $E(\gamma_s)<E(\gamma_0)$. Thus, any such γ_s belongs to W.

We fix $k\in\mathbb{N}$ large enough so that γ_0 is contained in the space of broken closed geodesics $\Lambda_k\subset\Lambda$. We define the continuous map $r:W\to\Lambda_k$ by $r(\gamma)(\frac{i}{k})=\gamma(\frac{i}{k})$ for all $i\in\mathbb{Z}_k$. The above convexity property of U implies that, for each $\gamma\in W$, $r(\gamma)$ is a curve contained in U and homotopic to γ within U. Therefore, r is a retraction $r:W\to W\cap\Lambda_k$. Since $E(r(\gamma))\le E(\gamma)$, we have

$$\ell^2:=\inf_W E=\inf_{W\cap\Lambda_k}E.$$

We choose a sequence $\gamma_n\in W\cap\Lambda_k$ such that $E(\gamma_n)\to\ell^2$. We can assume that each γ_n is without self-intersection. Indeed, if γ_n has self-intersections, we can find an interval $[a,b]\subsetneq[0,1]$ such that $\gamma_n|_{[a,b]}$ is a non-contractible loop in U. If i_0,i_1 are positive integers such that

$$[\tfrac{i_0+1}{k},\tfrac{i_1-1}{k}]\subseteq[a,b]\subseteq[\tfrac{i_0}{k},\tfrac{i_1}{k}],$$

we define $\tilde\gamma_n\in W\cap\Lambda_k$ by setting $\tilde\gamma_n(\frac{i}{k})=\gamma_n(a)$ for all $i\in\{0,\dots,i_0\}\cup\{i_1,\dots,k-1\}$, and $\tilde\gamma_n(\frac{i}{k})=\gamma_n(\frac{i}{k})$ for all $i\in\{i_0+1,\dots,i_1-1\}$. The curve $\tilde\gamma_n$ has less self-intersections than γ_n, and energy $E(\tilde\gamma_n)\le E(\gamma_n)$. Since a broken closed geodesic has only finitely many self-intersections, by repeating this procedure a finite number of times we eliminate all of them.

Since $\overline{W\cap\Lambda_k}$ is compact, up to extracting a subsequence we have that

$$\gamma_n\to\gamma\in\overline{W\cap\Lambda_k},$$

and $E(\gamma)=\ell^2$. We claim that γ is a closed geodesic. This is clear if γ is contained in $W\cap\Lambda_k$, for in this case it would be a critical point of the energy functional E. Assume now that $\gamma\in\partial(W\cap\Lambda_k)$, and consider the unique $\theta,\theta_n\in\Lambda_k$ such that

$$\theta(\tfrac{i}{k})=\gamma(\tfrac{i+1/2}{k}),\quad \theta_n(\tfrac{i}{k})=\gamma_n(\tfrac{i+1/2}{k}),\quad \forall i\in\mathbb{Z}_k.$$

Clearly, $\theta_n\to\theta$. Moreover, $E(\theta)\le E(\gamma)$ with equality if and only if γ is a closed geodesic. The above convexity property of U implies that $\theta_n\in W\cap\Lambda_k$. Therefore, $E(\theta_n)\ge\inf E|_W=E(\gamma)$ and $E(\theta)=E(\gamma)$, and we conclude that γ is a closed geodesic.

Since the approximating loops γ_n are without self-intersections, γ is a simple closed geodesic. Therefore, the union $\gamma_0(S^1)\cup\gamma(S^1)$ bounds an open annulus $A\subset U$. Since $E(\gamma)=\inf E|_W$, in particular there is no $\tilde\gamma\in\Lambda$ freely homotopic to γ with energy $E(\tilde\gamma)<E(\gamma)$ and support $\tilde\gamma(S^1)\subset A$. Therefore,

by applying Proposition 4.2(vii) as above, we infer that γ has no conjugate points. Corollary 5.7 implies that (S^2, F) has infinitely many closed geodesics. □

Proof of Theorem 1.5. By Theorem 1.3, if (S^2, F) has only finitely many simple closed geodesics, there exists at least one simple closed geodesic $\gamma \in \mathrm{crit}(E)$ with non-zero local homology $C_3(\gamma; \mathbb{Z}_2)$. If γ does not have a well-defined Birkhoff map, Theorem 5.8 implies that there are infinitely many closed geodesics. Assume now that γ has a well-defined Birkhoff map. Since $C_3(\gamma; \mathbb{Z}_2)$ is non-zero, $C_3(t \cdot \gamma; \mathbb{Z}_2)$ is non-zero as well, and

$$\mathrm{ind}(t\cdot\gamma) \leq 3 \leq \mathrm{ind}(t\cdot\gamma) + \mathrm{nul}(t\cdot\gamma), \quad \forall t \in S^1.$$

By Proposition 4.2(iv), we have

$$\mathrm{ind}_\Omega(t\cdot\gamma) + \mathrm{nul}_\Omega(t\cdot\gamma) \geq \mathrm{ind}(t\cdot\gamma) + \mathrm{nul}(t\cdot\gamma) - 1 \geq 2, \quad \forall t \in S^1. \tag{5.11}$$

Since $\mathrm{nul}_\Omega(t\cdot\gamma) \leq 1$ according to Proposition 4.2(ii), the inequality in (5.11) implies that $\mathrm{ind}_\Omega(t\cdot\gamma) \geq 1$. If $\mathrm{ind}_\Omega(t\cdot\gamma) \geq 2$ for all $t \in S^1$, Theorem 5.3 implies that there are infinitely many closed geodesics. If instead $\mathrm{ind}_\Omega(t\cdot\gamma) = 1$ for some $t \in S^1$, the above inequality implies that $\mathrm{nul}_\Omega(t\cdot\gamma) = 1$, and Corollary 5.6 implies that there are infinitely many closed geodesics. □

Acknowledgements

We are grateful to Egor Shelukhin and Vukašin Stojisavljević for pointing out a mistake in the original proof of Theorem 5.4. We are also grateful to the anonymous referee, who did a tremendous work with the first draft of the manuscript: in particular, for pointing out a mistake in the original proof of Theorem 1.2(iv), and for providing many valuable corrections and suggestions throughout the whole manuscript. Part of this work was carried out while Marco Mazzucchelli was visiting the Scuola Internazionale Superiore di Studi Avanzati (Trieste, Italy) in May 2018, and the Ruhr-Universität Bochum (Germany) in September–November 2019. We wish to thank both institutions for providing an excellent working environment. Marco Mazzucchelli and Stefan Suhr are supported by the SFB/TRR 191 "Symplectic Structures in Geometry, Algebra and Dynamics", funded by the Deutsche Forschungsgemeinschaft.

References

[1] Asselle, L., Mazzucchelli, M.: Closed geodesics with local homology in maximal degree on non-compact manifolds. Differ. Geom. Appl. **58**, 17–51 (2018)

[2] Abbondandolo, A., Macarini, L., Paternain, G.P.: On the existence of three closed magnetic geodesics for subcritical energies. Comment. Math. Helv. **90**(1), 155–193 (2015)

[3] Angenent, S.: Parabolic equations for curves on surfaces. I. Curves with p-integrable curvature. Ann. Math. (2) **132**(3), 451–483 (1990)

[4] Angenent, S.: Parabolic equations for curves on surfaces. II. Intersections, blow-up and generalized solutions. Ann. Math. (2) **133**(1), 171–215 (1991)

[5] Angenent, S.: Self-intersecting geodesics and entropy of the geodesic flow. Acta Math. Sin. (Engl. Ser.) **24**(12), 1949–1952 (2008)

[6] Ballmann, W.: Der Satz von Lusternik und Schnirelmann, Beiträge zur Differentialgeometrie, Heft 1, Bonner Math. Schriften, vol. 102, pp. 1–25. University of Bonn, Bonn (1978)

[7] Bangert, V.: Closed geodesics on complete surfaces. Math. Ann. **251**(1), 83–96 (1980)

[8] Bangert, V.: On the existence of closed geodesics on two-spheres. Int. J. Math. **4**(1), 1–10 (1993)

[9] Birkhoff, G.D.: Dynamical Systems. American Mathematical Society Colloquium Publications, vol. IX. American Mathematical Society, Providence (1966)

[10] Bangert, V., Long, Y.: The existence of two closed geodesics on every Finsler 2-sphere. Math. Ann. **346**(2), 335–366 (2010)

[11] Bott, R.: On the iteration of closed geodesics and the Sturm intersection theory. Commun. Pure Appl. Math. **9**, 171–206 (1956)

[12] Ballmann, W., Thorbergsson, G., Ziller, W.: Closed geodesics on positively curved manifolds. Ann. Math. **116**(2), 213–247 (1982)

[13] Chang, K.-C.: Infinite-Dimensional Morse Theory and Multiple Solution Problems. Progress in Nonlinear Differential Equations and Their Applications, vol. 6. Birkhäuser Boston Inc., Boston (1993)

[14] Caponio, E., Javaloyes, M.Á., Masiello, A.: On the energy functional on Finsler manifolds and applications to stationary spacetimes. Math. Ann. **351**(2), 365–392 (2011)

[15] Frauenfelder, U., Lange, C., Suhr, S.: A Hamiltonian version of a result of Gromoll and Grove. Ann. Inst. Fourier (Grenoble) **69**(1), 409–419 (2019)

[16] Franks, J.: Geodesics on S^2 and periodic points of annulus homeomorphisms. Invent. Math. **108**(2), 403–418 (1992)

[17] Gromoll, D., Grove, K.: On metrics on S^2 all of whose geodesics are closed. Invent. Math. **65**(1), 175–177 (1981/1982)

[18] Gromoll, D., Meyer, W.: On differentiable functions with isolated critical points. Topology **8**, 361–369 (1969)

[19] Gromoll, D., Meyer, W.: Periodic geodesics on compact Riemannian manifolds. J. Differ. Geom. **3**, 493–510 (1969)

[20] Grayson, M..A.: Shortening embedded curves. Ann. Math. (2) **129**(1), 71–111 (1989)

[21] Hadamard, M.: Les surfaces à courbures opposées et leurs lignes géodésiques. J. Math. Pures Appl. **5**(4), 27–74 (1898)

[22] Hingston, N.: On the growth of the number of closed geodesics on the two-sphere. Int. Math. Res. Not. **1993**(9), 253–262 (1993)

[23] Hass, J., Scott, P.: Shortening curves on surfaces. Topology **33**(1), 25–43 (1994)

[24] Irie, K., Marques, F.C., Neves, A.: Density of minimal hypersurfaces for generic metrics. Ann. Math. (2) **187**(3), 963–972 (2018)

[25] Jost, J.: A nonparametric proof of the theorem of Lusternik and Schnirelman. Arch. Math. (Basel) **53**(5), 497–509 (1989)

[26] Jost, J.: Correction to: "A nonparametric proof of the theorem of Lusternik and Schnirelman". Arch. Math. (Basel) **56**(6), 624 (1991)

[27] Katok, A.B.: Ergodic perturbations of degenerate integrable Hamiltonian systems. Izv. Akad. Nauk SSSR Ser. Mat. **37**, 539–576 (1973)

[28] Klingenberg, W.P.A.: Riemannian Geometry, 2nd edn. De Gruyter Studies in Mathematics, vol. 1. Walter de Gruyter & Co., Berlin (1995)

[29] Ljusternik, L.A.: The Topology of the Calculus of Variations in the Large. Translated from the Russian by J. M. Danskin. Translations of Mathematical Monographs, vol. 16. American Mathematical Society, Providence (1966)

[30] Lusternik, L., Schnirelmann, L.: Existence de trois géodésiques fermées sur toute surface de genre 0. C. R. Acad. Sci. Paris **188**, 534–536 (1929)

[31] Lu, G.: Methods of infinite dimensional Morse theory for geodesics on Finsler manifolds. Nonlinear Anal. **113**, 230–282 (2015)

[32] Marques, F.C.: Abundance of minimal surfaces. Jpn. J. Math. **14**(2), 207–229 (2019)

[33] Mazzucchelli, M.: Symplectically degenerate maxima via generating functions. Math. Z. **275**(3–4), 715–739 (2013)

[34] Mazzucchelli, M.: The Morse index of Chaperon's generating families. Publ. Mat. Urug. **16**, 81–125 (2016)

[35] Mantegazza, C., Martinazzi, L.: A note on quasilinear parabolic equations on manifolds. Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **11**(4), 857–874 (2012)

[36] Morse, M.: The Calculus of Variations in the Large. American Mathematical Society Colloquium Publications, vol. 18. American Mathematical Society, Providence (1996). Reprint of the 1932 original

[37] Mazzucchelli, M., Suhr, S.: A characterization of Zoll Riemannian metrics on the 2-sphere. Bull. Lond. Math. Soc. **50**, 997–1006 (2018)

[38] Neumann, W.D.: Generalizations of the Poincaré Birkhoff fixed point theorem. Bull. Aust. Math. Soc. **17**(3), 375–389 (1977)

[39] Oaks, J.A.: Singularities and self-intersections of curves evolving on surfaces. Indiana Univ. Math. J. **43**(3), 959–981 (1994)

[40] Poincaré, H.: Sur les lignes géodésiques des surfaces convexes. Trans. Am. Math. Soc. **6**, 237–274 (1905)

[41] Rademacher, H.-B.: Morse-Theorie und geschlossene Geodätische, Bonner Math. Schriften, vol. 229. University of Bonn, Bonn (1992)

[42] Taimanov, I.A.: On the existence of three nonintersecting closed geodesics on manifolds that are homeomorphic to the two-dimensional sphere. Izv. Ross. Akad. Nauk Ser. Mat. **56**(3), 605–635 (1992)

[43] Vigué-Poirrier, M., Sullivan, D.: The homology theory of the closed geodesic problem. J. Differ. Geom. **11**(4), 633–644 (1976)

[44] Yau, S.T.: Problem Section, Seminar on Differential Geometry. Annals of Mathematics Studies, vol. 102, pp. 669–706. Princeton University Press, Princeton (1982)

[45] Ziller, W.: Geometry of the Katok examples. Ergod. Theory Dyn. Syst. **3**(1), 135–157 (1983)

Guido De Philippis
Courant Institute of Mathematical Sciences
New York University
251 Mercer Street
New York NY10012-1185
USA
e-mail: guido@cims.nyu.edu

Michele Marini
Scuola Internazionale Superiore di Studi Avanzati
via Bonomea 265
34136 Trieste
Italy
e-mail: mmarini@sissa.it

Marco Mazzucchelli
CNRS, École Normale Supérieure de Lyon, UMPA
46 allée d'Italie
69364 Lyon Cedex 07
France
e-mail: marco.mazzucchelli@ens-lyon.fr

Stefan Suhr
Fakultät für Mathematik
Ruhr Universität Bochum
Universitätsstraße 150
44801 Bochum
Germany
e-mail: stefan.suhr@rub.de

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:43
https://doi.org/10.1007/s11784-022-00964-7
Published online May 5, 2022

Journal of Fixed Point Theory and Applications

Check for updates

Families of Legendrians and Lagrangians with unbounded spectral norm

Georgios Dimitroglou Rizell

Abstract. Viterbo has conjectured that any Lagrangian in the unit co-disc bundle of a torus which is Hamiltonian isotopic to the zero-section satisfies a uniform bound on its spectral norm; a recent result by Shelukhin showed that this is indeed the case. The modest goal of our note is to explore two natural generalisations of this geometric setting in which the bound of the spectral norm fails: first, passing to Legendrian isotopies in the contactisation of the unit co-disc bundle (recall that any Hamiltonian isotopy can be lifted to a Legendrian isotopy) and, second, considering Hamiltonian isotopies but after modifying the co-disc bundle by attaching a critical Weinstein handle.

Mathematics Subject Classification. 53D40, 53D42.

Keywords. Spectral norm, Boundary depth, Spectral invariants for Legendrians, Floer homology for Legendrians.

Contents

The author is supported by the Grant KAW 2016.0198 from the Knut and Alice Wallenberg Foundation.

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder and Felix Schlenk.

1. Introduction and results

Spectral invariants were introduced in Viterbo's seminal work [31]. Since their appearance they have become one of the most fundamental tools of quantitative symplectic topology. We do not intend to give an overview of its development and many applications here; instead we direct the reader to work by Oh [23] for a thorough introduction to the subject from a modern perspective.

Very briefly, spectral invariants in the symplectic case consist of functions from the group of Hamiltonian diffeomorphisms

$$c\colon Ham(X,\omega) \to \mathbf{R}$$

that take values in the real numbers, and which satisfy a list of axioms that will be omitted. The spectral invariants that we consider here are constructed as follows. For a pair of exact Lagrangian submanifolds $L_0, L_1 \subset (X, d\lambda)$ (the symplectic manifold is thus necessarily exact) one can associate the Floer complex $CF(L_0, \phi(L_1))$ to any Hamiltonian diffeomorphism $\phi \in Ham(X,\omega)$ endowed with its canonical action filtration. Spectral invariants are certain real numbers that encode information about this filtered chain complex. To make this precise, we utilise the language of the **barcode** from the theory of persistent homology in topological data analysis, which goes back to work by Carlsson–Zomorodian–Collins–Guibas [8]. This theory has been proven to be very useful in quantitative symplectic topology, where it was introduced by Polterovich–Shelukhin [25] and Usher–Zhang [30]; also see the recent work [24] by Polterovich–Rosen–Samvelyan–Zhang for a systematic introduction. Here we give a quick definition of the barcode of a filtered complex that will suit our needs in Sect. 2.2.

The barcode can be defined for any chain complex $(C, \partial, \mathfrak{a})$ with a filtration by subcomplexes

$$C_*^{<c} := \mathfrak{a}^{-1}[-\infty, c) \subset C_*$$

defined by an "action" function

$$\mathfrak{a}\colon C \to \{-\infty\} \cup \mathbf{R},$$

where $\mathfrak{a}^{-1}(-\infty) = \{0\}$. Phrased in this language, the spectral invariants are the values of the starting points of the semi-infinite bars of the barcode associated to the Floer complex. In fact, the main interest here is not the spectral invariants themselves, but rather the following derived quantities (see Definition 2.7):

- The **spectral range** of a filtered complex, denoted by
$$\rho(C, \partial, \mathfrak{a}) \in \{-\infty\} \cup [0, +\infty].$$
This quantity is defined as the supremum of the distances between the starting points of two semi-infinite bars in the corresponding barcode. (It takes the value $-\infty$ if and only if there are no semi-infinite bars.)
- The **boundary depth** of a filtered complex, denoted by
$$\beta(C, \partial, \mathfrak{a}) \in \{-\infty\} \cup [0, +\infty].$$
This quantity is defined as the supremum of the lengths of a finite bar in the corresponding barcode. (It takes the value $-\infty$ if and only if there are no bars of finite length.)

For the Floer complex $CF(L, \phi^1_H(L))$ of a closed embedded exact Lagrangian and its Hamiltonian deformation, the spectral range coincides with a quantity called the **spectral norm**. This can be seen using Leclercq's results from [20, Corollary 1.7], after relating the spectral invariants used in that article to the endpoints of semi-infinite bars in the relevant barcode. Since we will not use any of the particular features satisfied by the spectral norm here, we will gloss over the difference between these two concepts and simply define the spectral norm as

$$\gamma(CF(\Lambda, \phi^1_H(\Lambda))) := \rho(CF(\Lambda, \phi^1_H(\Lambda))),$$

i.e. we prescribe it to be equal to the spectral range.

Remark 1.1. The correct way to define the spectral norm in the setting of Legendrians would be to define γ as the difference of action levels of the classes that correspond to the unit for the cup-product and its image under Poincaré duality. We do not go into details of products and Poincaré duality here, but when Λ is a Legendrian without Reeb chords, we again expect an equality between spectral norm and spectral range. In general there should be an inequality $\gamma \leq \rho$.

We also need a generalisation of the above spectral invariants to contact manifolds. Since we will only consider contact manifolds of a very particular type, namely contactisations

$$(Y, \alpha) = (X \times \mathbf{R}, dz + \lambda)$$

of exact symplectic manifolds $(X, d\lambda)$ (see Sect. 2.1), this can be done by relying on well-established techniques. From our point of view, the spectral invariants of a contact manifold are defined for the group of contactomorphisms which are contact-isotopic to the identity, and yield functions

$$c \colon Cont_0(Y, \alpha) \to \mathbf{R}.$$

Note that the value does depend on the choice of contact form α here, and not just on the contact structure $\ker\alpha \subset TY$. It should be noted that spectral invariants in the contact setting are much less studied and developed than the symplectic version. However, the original formulation of the spectral invariants, which appeared in [31] for symplectic cotangent bundles $(X, \omega) = (T^*M, d(p\,dq))$, admits a straightforward generalisation to the standard contact jet-space

$$(J^1M = T^*M \times \mathbf{R}, dz - p\,dq),$$

as shown by Zapolsky [33]. In fact, the spectral invariants in [31] are based on a version of Floer homology defined using generating families, and this theory can be generalised to invariants of Legendrian isotopies inside jet-spaces by work of Chekanov [7]. Note that jet-spaces are particular cases of contactisations.

The spectral invariants considered here can be defined either by using generating families as in [33], or using a Floer homology constructed using the Chekanov–Eliashberg algebra as first done in [13] by Ekholm–Etnyre–Sabloff; also see work [4] by the author together with Chantraine–Ghiggini–Golovko. The Chekanov–Eliashberg algebra is a Legendrian isotopy invariant in the form of a unital differential graded algebra (DGA) that is freely generated by the Reeb chords on the Legendrian, which are all assumed to be transverse. Given a pair of Legendrians Λ_0 and Λ_1, the spectral invariants that we consider are associated to the barcode of the Floer complex $CF(\Lambda_0, \phi(\Lambda_1))$ where ϕ is a contactomorphism that is contact isotopic to the identity. See Sect. 2.3 for the definition of this Floer complex.

Viterbo conjectured in [32] that the spectral norm $\gamma(CF(0_{\mathbf{T}^n}, \phi(0_{\mathbf{T}^n})))$ of the Floer complex of the zero section $0_{\mathbf{T}^n} \subset T^*\mathbf{T}^n$ satisfies a uniform bound whenever $\phi \in Ham(T^*\mathbf{T}^n)$ maps the zero section $\phi(0_{\mathbf{T}^n}) \subset DT^*\mathbf{T}^n$ into the unit-disc cotangent bundle. In recent work by Shelukhin [28,29] this property was finally shown to be the case, even for a wide range of cotangent bundles beyond the torus case. The main point of our work here is to give examples of geometric settings beyond symplectic co-disc bundles, where the analogous boundedness of the spectral norm fails. It should be stressed that, in the time of writing of this article, there are still many cases of cotangent bundles for which the problem remains open: does the spectral norm of an exact Lagrangian inside DT^*M which is Hamiltonian isotopic to the zero-section satisfy a uniform bound for an arbitrary closed smooth manifold M?

As a first result, in Part (1) of Theorem A, we show that the spectral norm of Legendrians inside the contactisation $D^*S^1 \times \mathbf{R} \subset J^1S^1$ which are Legendrian isotopic to the zero section does not satisfy a uniform bound. Recall that any Hamiltonian isotopy of $0_{S^1} \subset DT^*S^1$ lifts to a Legendrian isotopy of the zero section $j^10 \subset J^1S^1$ (see Lemma 2.1); consequently, one way to formulate Part (1) of Theorem A is by saying that Viterbo's conjecture cannot be generalised to Legendrian isotopies.

Below we denote by

$$F_{\theta_0,z_0} := \{\theta = \theta_0, z = z_0\} \subset J^1S^1$$

a Legendrian lift of the Lagrangian cotangent fibre $T^*_{\theta_0}S^1$.

Theorem A.

(1) There exists a contact isotopy $\phi^t\colon J^1S^1 \to J^1S^1$ that satisfies

$$\phi^t|_{j^10}\colon j^10 \hookrightarrow DT^*S^1 \times \mathbf{R} = (S^1 \times [-1,1]) \times \mathbf{R},$$

and for which $CF(j^10, \phi^t(j^10))$ all are generated by precisely two mixed Reeb chords, whose difference in length grows indefinitely as $t \to +\infty$. In particular, the spectral norm $\gamma(CF(j^10, \phi^t(j^10)))$ becomes arbitrarily large as $t \to +\infty$.

(2) There exists a contact isotopy $\phi^t\colon J^1S^1 \to J^1S^1$ that satisfies

$$\phi^t|_{\Lambda_{\mathrm{st}}}\colon \Lambda_{\mathrm{st}} \hookrightarrow DT^*S^1 \times \mathbf{R}_{>0} = (S^1 \times [-1,1]) \times \mathbf{R}_{>0} \subset J^1S^1,$$

where $\Lambda_{\mathrm{st}} \subset J^1S^1$ is the standard Legendrian unknot shown in Fig. 1, and for which the boundary depth $\beta(CF(\phi^t(\Lambda_{\mathrm{st}}), F_{\theta_0,0}))$ becomes arbitrarily large as $t \to +\infty$. In addition, we may assume that ϕ^t is supported inside some subset $\{z \geq c\}$, where $c > 0$, and for which the inclusion $\{z \geq c\} \cap \Lambda_{\mathrm{st}} \subsetneq \Lambda_{\mathrm{st}}$ is a strict subset.

In recent work [2, Section 6.2] Biran–Cornea showed that a bound $\gamma(CF(0^M, L)) \leq C$ on the spectral norm of the Floer complex of a Lagrangian $L \subset T^*M$, where L is Hamiltonian isotopic to the zero section, implies the bound $\beta(CF(L, T^*_{pt}M)) \leq 2C$ on the boundary depth of the Floer complex of L and a cotangent fibre. The Legendrians produced by Part (2) of Theorem A can be used to show that the analogous result cannot be generalised to Legendrian isotopies. More precisely,

Corollary B. *There exists a contact isotopy $\phi^t\colon J^1S^1 \to J^1S^1$ that satisfies*

$$\phi^t|_{j^10}\colon j^10 \hookrightarrow DT^*S^1 \times \mathbf{R} = (S^1 \times [-1,1]) \times \mathbf{R},$$

and for which the spectral norm $\gamma(CF(j^10, \phi^1(j^10)))$ is uniformly bounded for all $t \geq 0$, while the boundary depth $\beta(CF(\phi^1(0_{S^1}), F_{\theta_0,z_0}))$ becomes arbitrarily large as $t \to +\infty$.

Proof. Take a cusp-connected sum of a C^1-small perturbation of the zero-section j^10 and any unknot Λ^t_{st} from the family produced by Part (2) of Theorem A; the case of Λ_{st} is shown in Fig. 1. We refer to [9] for the definition of cusp-connected sum (also called ambient Legendrian 0-surgery) along a Legendrian arc (the so-called surgery disc). We perform the cusp-connected sum along a Legendrian arc contained inside the region $\{z < c\}$, and which is disjoint from the support of the Legendrian isotopy of the unknots. Note that the Legendrian resulting from the cusp-connected sum is Legendrian isotopic to the zero-section, as shown in Fig. 1. It follows that the same is true for the cusp-connected sum of j^10 and any Legendrian Λ^t_{st} from the family.

Finally, to evaluate the effect of the ambient surgery on the barcodes of the Floer complexes we apply Theorem C. To that end, the following two facts are needed. First, $CF(\Lambda_{\mathrm{st}}, j^10)$ is acyclic, and thus its barcode consists of only finite bars. The acyclicity of the Floer complex is a consequence of the

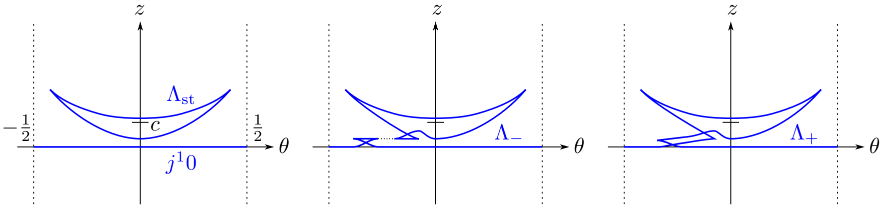


FIGURE 1. Left: the front projection of the zero section $j^1 0 \subset J^1(\mathbf{R}/\mathbf{Z}) = J^1 S^1$ and a standard Legendrian unknot Λ_{st}. Middle: the result of a Legendrian RI-move on each component, Λ_- denotes the union of the two components. Right: the Legendrian Λ_+ which is the result after a cusp-connected sum along the dotted arc shown in the middle picture. Λ_+ is Legendrian isotopic to the zero section (Λ_+ is obtained by performing two RI-moves on the zero-section)

invariance under Legendrian isotopy. (After a translation of Λ_{st} sufficiently far in the p-coordinate, all generators of the Floer complex disappear.) Second,

$$CF(j^1 f \cup \Lambda^t_{\mathrm{st}}, j^1 0) = CF(j^1 f, j^1 0) \oplus CF(\Lambda^t_{\mathrm{st}}, j^1 0)$$

is a direct sum of complexes. The barcode of the complex on the left-hand side is thus the union of the barcodes of the two complexes in the direct sum on the right-hand side. Here we have suppressed the choices of augmentations, since these Floer complexes do not depend on these choices (up to action preserving automorphism); see Remark 1.2. □

To define Floer homology for a pair of Legendrians Λ_0, Λ_1, it is necessary to also include the data of augmentations $\varepsilon_i \colon \mathcal{A}(\Lambda_i) \to \mathbf{k}$ of their Chekanov–Eliashberg algebras; these are unital DGA-morphisms onto the ground field. We write $CF((\Lambda_0, \varepsilon_0), (\Lambda_1, \varepsilon_1))$ for the induced complex, which in general does depend on the choices of augmentations; we refer to Sect. 2.4 for more details.

Remark 1.2. There are Legendrian isotopy classes for which different choices of augmentations always give rise to Floer complexes that are isomorphic as filtered chain complexes; this can be characterised using the invariance of the augmentation category of Chantraine–Bourgeois from [3]. Cases include the Legendrian isotopy class of the zero section $j^1 0$, the Legendrian fibre $F_{\theta,z}$, and the standard Legendrian unknot. The property is a consequence of the fact that these Legendrian isotopy classes admit representatives for which the Chekanov–Eliashberg algebra admits a *unique* augmentation.

Theorem C. *Let Λ_+ be a Legendrian obtained from Λ_- by a Legendrian ambient surgery. After making the surgery-region sufficiently small, we can assume that there is an action-preserving isomorphism*

$$CF((\Lambda_+, \varepsilon_+), (\Lambda, \varepsilon)) \to CF((\Lambda_-, \varepsilon_-), (\Lambda, \varepsilon))$$

of complexes, where (Λ, ε) is an arbitrary but fixed Legendrian equipped with an augmentation ε of its Chekanov–Eliashberg algebra $\mathcal{A}(\Lambda)$, and where the *augmentation ε_+ of $\mathcal{A}(\Lambda_+)$ is induced by pulling back the augmentation ε_- of $\mathcal{A}(\Lambda_-)$ under the unital DGA-morphism induced by the standard Lagrangian handle-attachment cobordism. In particular, the barcodes of the two Floer complexes coincide.*

In the setting of exact Lagrangian cobordisms in the sense of Arnol'd between exact Lagrangian submanifolds similar results were found in [2, Section 5.3] .

Finally we present a Hamiltonian isotopy of a closed exact Lagrangian inside a Liouville domain for which the spectral norm becomes arbitrarily large. The simplest examples of such a Liouville domain is the 2-torus with an open ball removed; we denote this space by $(\Sigma_{1,1}, d\lambda)$ and depict it in Fig. 13. The detailed construction is given in Sect. 2.1.2. As kindly pointed out to the author by the anonymous referee, this fact is not new. The same phenomenon was exhibited in, e.g. Zapolsky's work [34, Lemma 1.10], as well as the more recent [19, Remark 6] by Kislev–Shelukhin.

Theorem D. *There exists a closed exact Lagrangian submanifold $L \subset (\Sigma_{1,1}, \omega)$ and a compactly supported Hamiltonian $H\colon \Sigma_{1,1} \to \mathbf{R}$ for which the induced compactly supported Hamiltonian isotopy $\phi_H^t\colon (\Sigma_{1,1}, \omega) \to (\Sigma_{1,1}, \omega)$ satisfies the property that the spectral norm $\gamma(CF(L, \phi_H^t(L)))$ becomes arbitrarily large as $t \to +\infty$.*

1.1. Why the proofs of uniform bounds fail for Legendrians

The techniques that are used in [29] and [2] to prove the results in the case of the cotangent bundle are not yet fully developed in the case of Legendrians in contactisations. This includes the closed-open map, which is a crucial ingredient in [29], and a unital A_∞-structure on the Floer complex with relevant PSS-isomorphisms, which is crucial in [2]. Nevertheless, we still do expect that these operations can be defined also for the Floer homology of Legendrians in contactisations. In fact the A_∞-structure was recently extended to this setting by Legout [21]. Assuming the possibility to define these operations in the Legendrian setting, what goes wrong when one tries to generalise the proofs to the Legendrian case?

First we recall the properties of the Floer homology complex of a Legendrian and itself; see, e.g. [13] for the details. To define $CF(\Lambda, \Lambda)$ one must first make the mixed Reeb chords transverse by a Legendrian perturbation of the second copy of Λ. We do this by replacing Λ with a section $j^1 f$ in its standard contact jet-space neighbourhood, where $f\colon \Lambda \to \mathbf{R}$ is a C^1-small Morse function. In this manner, we obtain

$$CF(\Lambda, \Lambda) = C^{\mathrm{Morse}}(f; \mathbf{k}) \oplus \bigoplus_{c \in \mathcal{Q}(\Lambda)} \mathbf{k}p_c \oplus \mathbf{k}q_c$$

where $\mathcal{Q}(\Lambda)$ denotes the set of Reeb chords on Λ, and $C^{\mathrm{Morse}}(f; \mathbf{k})$ is the Morse homology complex with basis given by the critical points of the function $f\colon \Lambda \to \mathbf{R}$. The action of the former chords are approximately equal to $\mathfrak{a}(p_c) = \ell(c)$ and $\mathfrak{a}(q_c) = -\ell(c)$ while the action of the latter is equal to $\mathfrak{a}(x) = f(x)$. What is important to notice here is that the generators of

$C^{\text{Morse}}(f;\mathbf{k})$ may be assumed to have arbitrarily small action, while this is not the case for the generators that correspond to pure Reeb chords. When Λ is the Legendrian lift of a Lagrangian *embedding*, there are of course only generators of the type $C^{\text{Morse}}(f;\mathbf{k})$. This turns out to be the crucial difference between the symplectic and the contact case.

Example in Part (1) of Theorem A: The proof in [29] uses the closed-open map. More precisely, a crucial ingredient in the proof is the action-preserving property of the operations P'_a on the Floer homology $CF(0_M, \phi^1_H(0_M))$, which are defined using the length-0 part $\phi^0(a)$ and length-1 part $\phi^1(a,\cdot)$ of the closed open map for certain elements $a \in SH(T^*M)$ in symplectic cohomology (by this we mean the unital version of the symplectic invariant, which is contravariant with respect to inclusion of exact Liouville subdomains). In the case when the Legendrian has pure Reeb chords (i.e. it is not the lift of an exact Lagrangian embedding), the chain $\phi^0(a) \in CF(\Lambda,\Lambda)$ may consist of generators whose action does not vanish (since they do not correspond to Morse generators). In this case the action-preserving property of P'_a can no longer be determined from the aciton of $a \in SH(T^*M)$ alone.

Example in Part (2) of Theorem A: The proof in [2, Section 6.2] uses the fact that there are continuation elements $a \in CF(\phi^1_H(0_M), 0_M)$ and $b \in CF(0_M, \phi^1_H(0_M))$ for which $\mu_2(a,b) \in CF(\phi^1_H(0_M), \phi^1_H(0_M))$ is the unique maximum of a suitable Morse function. In the Legendrian case the element $\mu_2(a,b) \in CF(\phi^1(j^1 0), \phi^1(j^1 0))$ is still a homology unit; however, it not necessarily a sum of Morse chords, and can, therefore, have significant action. In particular, multiplication with the element $\mu_2(a,b)$ is not necessarily identity on the chain level, nor is it necessarily homotopic to the identity by a chain homotopy of small action. The geometrically induced chain homotopy $\mu_3(a,b,\cdot)$ between $\mu_2(a,\mu_2(b,\cdot))$ and $\mu_2(\mu_2(a,b),\cdot)$ increases action by at most the spectral norm, and is used in [2] for establishing the bound on the boundary depth. However, this chain homotopy does not do the job any more, since we also need an additional chain-homotopy (of unknown action properties) to take us from the map $\mu_2(\mu_2(a,b),\cdot)$ to the chain level identity.

2. Background

2.1. Contact geometry of jet-spaces and contactisations

An **exact symplectic manifold** is a smooth $2n$-dimensional manifold $(X^{2n}, d\lambda)$ equipped with a choice of a primitive one-form λ for an exact symplectic two-form $\omega = d\lambda$, i.e. ω is skew-symmetric, non-degenerate, and closed. Note that the primitive λ should be considered as part of the data describing the exact symplectic manifold. A compact exact symplectic manifold with boundary $(W, d\lambda)$ is a **Liouville domain** if the **Liouville vector field**, i.e. the vector field ζ given as the symplectic dual of λ via the equation $\iota_\zeta d\lambda = \lambda$, is transverse to the boundary ∂W. The flow generated by ζ is called the **Liouville flow** and satisfies $(\phi^t_\zeta)^*\lambda = e^t\lambda$. An open exact symplectic manifold $(\overline{W}, d\lambda)$ is a **Liouville manifold** if the all critical points of the Liouville vector field are contained inside some compact Liouville domain $W \subset (\overline{W}, d\lambda)$, and if the Liouville flow is complete.

A **Hamiltonian isotopy** is a smooth isotopy of X which is generated by a time-dependent vector field $V_t \in \Gamma(TX)$ that satisfies $\iota_{V_t} d\lambda = -dH_t$ for some smooth time-dependent function

$$H\colon X \times \mathbf{R}_t \to \mathbf{R}$$

which is called the **Hamiltonian**; a diffeomorphism of X which is the time-t flow generated by such a vector field preserves the symplectic form (but not necessarily the primitive) and is denoted by

$$\phi_H^t\colon (X, \omega) \to (X, \omega);$$

we call such a map a **Hamiltonian diffeomorphism**, and the corresponding flow a **Hamiltonian isotopy**. Conversely, any choice of Hamiltonian function induces a **Hamiltonian isotopy** ϕ_H^t in the above manner. Since we consider exact symplectic manifolds, a smooth isotopy $\phi^t\colon X \to X$ is a Hamiltonian isotopy if and only if $(\phi^t)^*\lambda = \lambda + dG_t \in \Omega^1(X)$ holds for some smooth function

$$G\colon X \times \mathbf{R}_t \to \mathbf{R}.$$

Note that the Hamiltonian function that corresponds to a Hamiltonian isotopy is determined only up to the addition of a function that only depends on t.

Any exact $2n$-dimensional symplectic manifold $(X^{2n}, d\lambda)$ gives rise to a $2n+1$-dimensional contact manifold $(X \times \mathbf{R}_z, dz + \lambda)$ called its **contactisation**, which is equipped with the canonical contact one-form $\alpha_{\mathrm{st}} := dz + \lambda$. The contactisations induced by choices of primitives of the symplectic form λ and $\lambda' = \lambda + df$ that differ by the exterior differential of $f\colon X \to \mathbf{R}$ are isomorphic via the coordinate change $z \mapsto z - f$. Recall that the contact condition is equivalent to $d\alpha_{\mathrm{st}}$ being non-degenerate on the contact planes $\ker \alpha_{\mathrm{st}} \subset T(X \times \mathbf{R})$. A **contact isotopy** is a smooth isotopy which preserves the distribution $\ker \alpha_{\mathrm{st}}$ (but not necessarily the contact form). The contraction $\iota_{V_t}\alpha_{\mathrm{st}}$ of the contact form and the infinitesimal generator gives a bijective correspondence between contact isotopies and smooth time-dependent functions on $X \times \mathbf{R}$, the latter are called **contact Hamiltonians**. We refer to [18] for more details.

Lemma 2.1. *A Hamiltonian isotopy $\phi_H^t\colon (X, d\lambda) \to (X, d\lambda)$ with a choice of Hamiltonian $H_t\colon X \to \mathbf{R}$ lifts to a contact isotopy*

$$\begin{aligned} X \times \mathbf{R} &\to X \times \mathbf{R}, \\ (x, z) &\mapsto (\phi_H^t(x), z - G_t(x)), \end{aligned}$$

where the function $G\colon X \times \mathbf{R}_t \to \mathbf{R}$ is defined by

$$G_t(x) = \int_0^t \lambda(V_s(\phi_H^s(x))) - H_s(\phi_H^s(x))ds$$

and satisfies the property

$$(\phi_H^t)^*\lambda = \lambda + d\,G_t.$$

Moreover, this contact isotopy preserves the contact form α_{st} *and is generated by the time-dependent contact Hamiltonian* $H_t \circ \mathrm{pr}_X : X \times \mathbf{R}_z \to \mathbf{R}$.

A smooth immersion of an n-dimensional manifold

$$\Lambda \looparrowright (X^{2n} \times \mathbf{R}, dz + \lambda)$$

in the contactisation is **Legendrian** if it is tangent to $\ker \alpha_{\mathrm{st}}$, while a smooth n-dimensional immersion $L \looparrowright (X^{2n}, \lambda)$ in an exact symplectic manifold is **exact Lagrangian** if λ pulls back to an *exact* one-form. The following relation between Legendrians and exact Lagrangians is immediate:

Lemma 2.2. *The canonical projection of a Legendrian immersion to* (X, λ) *is an exact Lagrangian immersion. Conversely, any exact Lagrangian immersion lifts to a Legendrian immersion of the contactisation* $X \times \mathbf{R}$*. Moreover, the lift is uniquely determined by the choice of a primitive* $f : L \to \mathbf{R}$ *of the pull-back* $\lambda|_{TL} = df$*, via the formula* $z = -f$*.*

Transverse double points of Lagrangian immersions are stable. On the other hand, generic Legendrian immersions are in fact *embedded*. However, there are stable self-intersections of Legendrians that appear in one-parameter families. Recall the following standard fact; again we refer to, e.g. [18] for details.

Lemma 2.3. *A compactly supported smooth isotopy* $\phi^t(\Lambda) \subset X \times \mathbf{R}$ *through Legendrian embeddings, also called a* **Legendrian isotopy***, can be generated by an ambient contact isotopy.*

2.1.1. The cotangent bundle and jet-space. There is a canonical exact symplectic two-form $-d(p\,dq)$ on any smooth cotangent bundle T^*M, whose primitive $-p\,dq$ is the tautological one-form with a minus sign. The cotangent bundle is a Liouville manifold and any co-disc bundle is a Liouville domain. The zero-section $0_M \subset T^*M$ is obviously an exact Lagrangian embedding.

The contactisation of T^*M is the one-jet space $J^1M = T^*M \times \mathbf{R}_z$ with the canonical contact one-form $dz - p\,dq$. The zero-section in T^*M lifts to the one-jet j^1c of any constant function c (obviously the one-jet j^1f of an arbitrary function $f : M \to \mathbf{R}$ is Legendrian isotopic to j^10). For us the most relevant example is actually the two-dimensional symplectic cotangent bundle $T^*S^1 = S^1 \times \mathbf{R}_p$ equipped with the exact symplectic two-form $-d(p\,d\theta)$, and its corresponding contactisation, i.e. the three-dimensional contact manifold

$$(J^1S^1 = T^*S^1 \times \mathbf{R}_z, dz - p\,d\theta)$$

(note the sign convention for the Liouville form).

To describe Legendrians in J^1M we will make use of the **front-projection**, by which one simply means the canonical projection

$$\Pi_F : J^1M \to M \times \mathbf{R}_z.$$

A Legendrian immersion can be uniquely determined by its post-composition with the front projection. A generic Legendrian knot in J^1S^1 has a front projection whose singular locus consists of

- non-vertical cubical cusps and

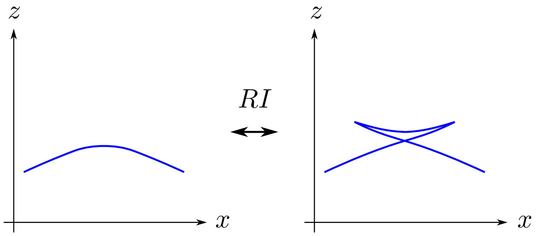


FIGURE 2. RI: the first Legendrian Reidemeister move in the front projection

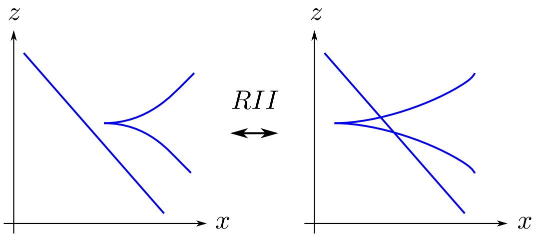


FIGURE 3. RII: the second Legendrian Reidemeister move in the front projection

- transverse self-intersections.

Note that the front projection cannot be tangent to ∂_z by the Legendrian condition (i.e. there are no vertical tangencies).

Two sheets of the front projection that have the same slopes (i.e. p-coordinates) above some given point in the base, project to a double-point inside T^*M. There is a bijection between double points of this projection and Reeb chords, where a Reeb chord is an integral curve of ∂_z with both endpoints on the Legendrian. The difference of z-coordinate of the endpoint and starting point of a Reeb chord c is called its **length** and is denoted by $\ell(c) \geq 0$.

Double-points of the Legendrian immersion itself correspond to self-tangencies of the front projection. This is not a stable phenomenon, and double-points of Legendrians generically arise only in one-parameter families. These double-points can be considered as Reeb chords of length zero.

Two Legendrian knots inside $J^1\mathbf{R}$ or J^1S^1 with generic fronts are Legendrian isotopic if and only if their front projections can be related by a sequence of *Legendrian Reidemeister moves* [26] together with an ambient isotopy of the front inside $S^1 \times \mathbf{R}_z$; see [15] for an introduction to Legendrian knots.

For convenience we will also introduce a composite move that we will make repeated use of; this is the one shown in Fig. 5, which involves taking two cusp edges with different slopes, and making them cross each other (it is important that the cusps have different slopes).

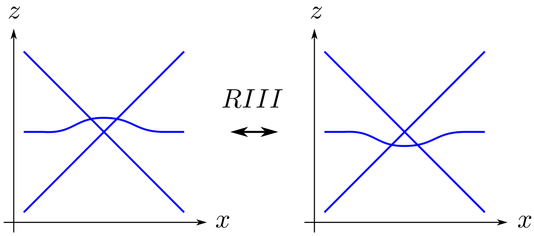


FIGURE 4. RIII: the third Legendrian Reidemeister move in the front projection

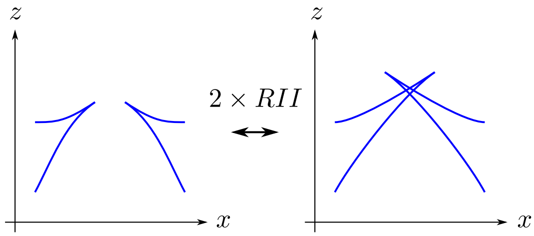


FIGURE 5. A composite move: the front to the right is obtained by performing two consecutive RII-moves on the front to the left together with an isotopy

2.1.2. The punctured torus. Here we construct an example of a two-dimensional non-planar Liouville domain: the two torus minus an open ball, which we denote by $(\Sigma_{1,1}, d\lambda)$.

First, consider the primitive

$$\lambda_0 = -\frac{1}{2}(p\,dq - q\,dp)$$

of the standard linear symplectic form $dq \wedge dp$ on $\mathbf{R}^2$. We have the identities

$$\lambda_0 + d\left(\frac{pq}{2}\right) = q\,dp,$$
$$\lambda_0 - d\left(\frac{pq}{2}\right) = -p\,dq.$$

Take a smooth function $\sigma\colon \mathbf{R}^2 \to \mathbf{R}$ which in the standard coordinates labelled by $(p,q) \in \mathbf{R}^2$ is given by

- $\sigma(p,q) = pq/2$ on $\{|q| \leq 1, |p| > 2\}$, while it is of the form $g(p)q/2$ for some smooth function g that satisfies $g(p), g'(p) \geq 0$ on $\{|q| \leq 1, |p| \geq 1\}$;
- $\sigma(p,q) = -pq/2$ on $\{|q| > 2, |p| \leq 1\}$, while it is of the form $-g(q)p/2$ for some smooth function g that satisfies $g(q), g'(q) \geq 0$ on $\{|q| \geq 1, |p| \leq 1\}$;
- $\sigma(p,q) = 0$ on $\{|q| < 1, |p| < 1\}$; and

Consider the exact symplectic manifold $(X, d\lambda)$ which is obtained by taking the cross-shaped domain

$$\{p \in [-2,2], q \in [-1,1]\} \cup \{q \in [-2,2], p \in [-1,1]\} \subset \mathbf{R}^2$$

and identifying $\{p = 2\}$ with $\{p = -2\}$, and $\{q = -2\}$ with $\{q = 2\}$ in the obvious manner. Topologically the result is a punctured torus. The Liouville form $\lambda_0 + d\sigma$ on $\mathbf{R}^2$ extends to a Liouville form λ on this punctured torus. The punctured torus has a skeleton $Sk \subset X$ which is the image of the cross $\{pq = 0\}$ under the quotient; in other words, $Sk \subset X$ is the union of two smooth Lagrangian circles that intersect transversely in a single point. Note that

$$Sk = \bigcap_{T=1}^{\infty} \phi^{-T}(X).$$

We claim that the sought Liouville domain $(\Sigma_{1,1}, d\lambda)$ can be realised as a suitable subset of this exact symplectic manifold, simply by smoothing its corners; see Fig. 13.

Since $(\Sigma_{1,1}, \lambda)$ is a surface with non-empty boundary, it admits a symplectic trivialisation of its tangent bundle. This implies that the all Lagrangian submanifolds of $\Sigma_{1,1}$ have a well-defined Maslov class; see Sect. 2.5 for more details. We will make heavy use of the fact that the Maslov class depends on the choice of a symplectic trivialisation; in this case, symplectic trivialisations up to homotopy can be identified with homotopy classes of maps

$$\Sigma_{1,1} \to S^1$$

i.e. cohomology classes $H^1(\Sigma_{1,1}; \mathbf{Z})$.

2.2. Barcode of a filtered complex and notions from spectral invariants

A **(strict) filtered complex** over some field $\mathbf{k}$ is a chain complex $(C, \partial, \mathfrak{a})$ in which each element is endowed with an action $\mathfrak{a}(c) \in \mathbf{R} \sqcup \{-\infty\}$ and such that the following properties are satisfied:

- $\mathfrak{a}(c) = -\infty$ if and only if $c = 0$,
- $\mathfrak{a}(r \cdot c) = \mathfrak{a}(c)$ for any $r \in \mathbf{k}^*$,
- $\mathfrak{a}(a + b) \le \max\{\mathfrak{a}(a), \mathfrak{a}(b)\}$, and
- $\mathfrak{a}(\partial(a)) < \mathfrak{a}(a)$ for any $a \neq 0$.

The subset

$$C^{<a} = \mathfrak{a}^{-1}(\{-\infty\} \cup (-\infty, a))$$

is a $\mathbf{k}$-subspace by the first three bullet points; this subspace is a subcomplex by the last bullet point.

We say that a basis $\{e_i\}$ is **compatible** with the filtration, if the action of a general element $c \in C$ is given by

$$\mathfrak{a}(r_1 e_1 + \cdots + r_n e_n) = \max\{\mathfrak{a}(e_i);\ r_i \neq 0\}, \quad r_i \in \mathbf{k}, \tag{2.1}$$

i.e. the action function is determined by its values on elements in the basis.

Remark 2.4. The non-trivial condition in the definition is the equality "=" in Formula (2.1); for a general basis the above equality gets replaced with an *inequality* "$\leq$".

The existence of a compatible basis for any filtered complex was proven by Barannikov [1]; see [30] for a more general version (in that article they are called "orthogonal bases"), as well as [24]. For proof adapted to the notation used here, see [10, Lemma 2.2].

Given a basis with a specified action on each basis element, one can also use the above formula to *construct* a filtration on the entire complex, under the assumption that the differential decreases action. The Floer complexes described below get endowed with filtrations in precisely this manner, i.e. by specifying an action for each canonical and geometrically induced basis element.

For every filtered complex there is a notion of a **barcode**; we refer to [10, Section 2] for the details of the presentation that we rely on here. The barcode is a set of intervals of the form $[a, b)$ and $[a, +\infty)$, where $a, b \in \mathbf{R}$, and we allow multiplicities. Instead of giving the usual definition of the barcode, we give it the following alternative characterisation.

Lemma 2.5. (Lemma 2.6 in [10]) *The barcode can be recovered from the following data:*

(1) For any basis which is compatible with the action filtration, there is a bijection between the set of actions of basis elements and the union of start and endpoints of bars (counted with multiplicities).

(2) For any two numbers $a < b$, the number of bars of C_* whose endpoints e satisfy $e \in (b, +\infty]$ and starting points s satisfy $s \in [a, b)$ is equal to $\dim H(C^{<b}/C^{<a})$.

Corollary 2.6. *Assume that the barcode contains a finite bar $[a, b)$. Then, for any compatible basis $\{e_i\}$, we can deduce the existence of basis elements e_i and e_j with $\mathfrak{a}(e_i) = b$, $\mathfrak{a}(e_j) = a$, such that $\langle \partial e_i, e_j \rangle \neq 0$.*

Conversely, if there exists a compatible basis $\{e_i\}$ for which $\partial e_i = r e_j$ for some coefficient $r \neq 0$, then the barcode contains the finite bar $[\mathfrak{a}(e_j), \mathfrak{a}(e_i))$.

Note that the barcode considered here is independent of the grading. An efficient way to deduce properties of the barcode is thus to find (possibly several different) gradings for the complex, for which the differential remains an operation of degree -1. The existence of such gradings imposes restrictions on the differential, which in view of the previous corollary imposes restrictions on the barcode. This technique will be used in the proofs given in Sects. 3.1 and 3.3.

For a filtered complex as above we can associate the following important notions.

Definition 2.7.

(1) The *spectral range* $\rho(C, \partial, \mathfrak{a}) \in \{-\infty\} \cup [0, +\infty]$ is the supremum of the distances between starting points of the semi-infinite bars in the barcode.

(2) The *boundary depth* $\beta(C, \partial, \mathfrak{a}) \in \{-\infty\} \cup [0, +\infty]$ is supremum of the lengths of the finite bars in the barcode.

Note that the above quantities automatically are equal to $-\infty$ in the case when the supremum is taken over the empty set (i.e. when there are no semi-infinite and finite bars, respectively).

An important feature of the barcode is that remains invariant under simple bifurcations of the complex, i.e. action preserving handle-slides and birth/deaths. Legendrian isotopies induce one-parameter families of the Floer complex considered here, which undergoes bifurcations of precisely this type; hence the corresponding barcode undergoes continuous deformations under Legendrian isotopies. Since this property will not be needed, we do not give more details here, but instead direct the interested reader to [10].

2.3. Outline of Floer homology and generating family homology for Legendrians

Floer homology for pairs (L_0, L_1) of closed exact Lagrangian submanifolds of cotangent bundles was originally defined by Floer [16]. For any such pair one obtains the Floer chain complex $CF(L_0, L_1)$ with a basis given by the intersections $L_0 \cap L_1$, which here are assumed to be transverse. Floer also showed that the homology of the complex—the so-called Floer homology $HF(L_0, L_1)$—is invariant under Hamiltonian isotopy of either Lagrangian L_i. Moreover, in the case when L_1 is a C^1-small Hamiltonian perturbation of L_0 the Floer complex $CF(L_0, L_1) = C^{\mathrm{Morse}}(f)$ is the Morse complex for a C^1-small Morse function $f\colon L_0 \to \mathbf{R}$ and suitable auxiliary data; see Floer's original computation [17]. (This property might not hold for the Floer homology of a Legendrian, due to additional generators corresponding to Reeb chords; see Sect. 1.1.)

Nowadays there are several different techniques available for constructing Floer homology. Here we will consider the setting of Legendrian submanifolds of contactisations $(\overline{W} \times \mathbf{R}, \alpha_{\mathrm{st}})$ of a Liouville manifold $(\overline{W}, d\lambda)$, in which Floer homology associates a chain complex $CF(\Lambda_0, \Lambda_1)$ to a pair of Legendrian submanifolds equipped with additional data. In this case, the homology of the complex is invariant under Legendrian isotopy of either Legendrian Λ_i. This is the version that we will use also in the case of exact Lagrangian embeddings in $(\overline{W}, d\lambda)$. To that end, recall that exact Lagrangians admit lifts to Legendrians by Lemma 2.2, and that a Hamiltonian isotopy of the Lagrangian induces a Legendrian isotopy of the Legendrian lift by Lemma 2.1.

In the case when $\overline{W} = T^*M$, and thus $\overline{W} \times \mathbf{R} = J^1M$, in [33] Zapolsky relied on generating family homology defined for generating families due to Chekanov [7] to define spectral invariants. Generating family homology is a Hamiltonian isotopy invariant obtained by Morse functions on finite-dimensional spaces, which behaves very similarly to Floer homology. In certain cases these two invariants have even been shown to be equivalent. Since we will work with contactisations that are more general than jet-spaces, we instead follow the techniques from [13] by Ekholm–Etnyre–Sabloff, where the Floer chain complex is constructed as the linearised Legendrian contact-homology complex associated to the Chekanov–Eliashberg algebra [6], [12].

First we outline the general set-up Floer homology in the setting of Legendrians, which applies equally well to either the version used here or the version defined using generating families (when applicable). Given a pair of Legendrians $\Lambda_0, \Lambda_1 \subset \overline{W} \times \mathbf{R}$, equipped with additional data denoted by ε_i to be specified below (in the version defined using generating families, these additional data are simply the choice of a generating family), one obtains a graded (grading is in $\mathbf{Z}$ or $\mathbf{Z}/\mu\mathbf{Z}$ depending on the Maslov class as described in Sect. 2.5) filtered chain complex

$$(CF_*((\Lambda_0, \varepsilon_0), (\Lambda_1, \varepsilon_1)), \partial, \mathfrak{a})$$

with a canonical compatible basis as a $\mathbf{k}$-vector space given by the

- Reeb chords c from Λ_0 to Λ_1 of action $\mathfrak{a}(c) = \ell(c)$ equal to the Reeb chord length; together with the
- Reeb chords c from Λ_1 to Λ_0 of action $\mathfrak{a}(c) = -\ell(c)$ equal to minus the Reeb chord length.

We assume that all Reeb chords are transversely cut out, and hence that they form a discrete subset, which thus is finite whenever the Legendrians are closed. With our conventions the differential is *strictly action decreasing and of degree* -1. In the case of generating family homology, the differential is the Morse homology differential for a Morse function on a finite-dimensional manifold that is constructed using the generating family. Below we give more details of the Floer complex defined via the Chekanov–Eliashberg algebra, for which the differential counts pseudoholomorphic strips in $\overline{W}$ with boundary on the Lagrangian projections $\Pi_{\overline{W}}(\Lambda_i) \subset \overline{W}$ (these are exact Lagrangian immersions with transverse self-intersections). In this case the strips are moreover allowed to have corners that map to the double points of the Lagrangian projections; the strips are then counted with weights given by the value of the augmentations on the corresponding pure Reeb chords. More details are given in Sect. 2.4 below.

The Floer complex satisfies the following important properties; see [13] for details.

- A Legendrian isotopy of the Legendrian Λ_i induces a canonical continuation of the additional data ε_i, and the resulting one-parameter family of Floer complexes undergoes only simple bifurcations, i.e. handle-slides and births/deaths. In particular, the homology of the complex is not changed under such a deformation.
- In the case when $\Lambda \subset \overline{W} \times \mathbf{R}$ has no Reeb chords (i.e. it is the lift of an exact Lagrangian embedding), and when Λ' is a C^1-small Legendrian perturbation, then the induced Floer complex

  $$(CF((\Lambda, \varepsilon), (\Lambda', \varepsilon')), \partial, \mathfrak{a}) = C^{\mathrm{Morse}}(f; \mathbf{k})$$

  is the Morse homology complex of some C^1-small Morse function $f \colon \Lambda \to \mathbf{R}$.

Again we refer to Sect. 1.1 for a description of the complex under the presence of pure Reeb chords; in this case the Morse complex is only realised as a quotient complex of a subcomplex.

2.4. Floer complex as the linearised Chekanov–Eliashberg algebra

Here we present the relevant technical details for the particular construction of Floer homology used in this paper, i.e. relying on the Chekanov–Eliashberg algebra for Legendrians in contactisation from [12]. Using the Chekanov–Eliashberg algebra to define Floer homology for Legendrian submanifolds is not new, it goes back to work [13] by Ekholm–Etnyre–Sabloff; also see [22] by Lanzat–Zapolsky for a nice application of this theory together with a systematic treatment.

Assume that $\Lambda_0, \Lambda_1 \subset \overline{W} \times \mathbf{R}$ are two Legendrian submanifolds. Further, assume that the Chekanov–Eliashberg algebras of Λ_i admit augmentations

$$\varepsilon_i \colon (\mathcal{A}(\Lambda_i), \partial) \to \mathbf{k};$$

recall that the Chekanov–Eliashberg algebra is a unital DGA generated by the Reeb chords of the Legendrian, and that an augmentation is a unital DGA morphism to the ground field. In particular, when the Legendrian Λ_i has no Reeb chords, the Chekanov–Eliashberg algebra takes the simple form $\mathcal{A}(\Lambda_i) = \mathbf{k}$, and there is a canonical augmentation. An important property of augmentations is that they can be pushed forward under a Legendrian isotopy; see, e.g. [4] and [6].

Typically one wants more additional data than just an augmentation. For instance, to use coefficients in a field of characteristic different from two, one also needs to fix the choice of a spin structure on both Legendrians Λ_i. To endow the Floer complex a $\mathbf{Z}$-grading, we need to specify a Maslov potential; we refer to Sect. 2.5 for more details concerning the grading, which will play an important role for us.

The Floer complex

$$CF((\Lambda_0, \varepsilon_0), (\Lambda_1, \varepsilon_1))$$

is generated by the chords that have one endpoint on Λ_0 and one endpoint on Λ_1 (either being a starting point). These Reeb chords on $\Lambda_0 \cup \Lambda_1$ are called the **mixed** Reeb chords. To define the differential, we will identify the above vector space with the underlying vector space linearised Legendrian contact homology complex of the link $\Lambda_0 \cup \phi^T_{\partial_z}(\Lambda_1)$, where the latter is the $\mathbf{k}$-vector space is generated by all Reeb chords that start on Λ_0 and end on the translation $\phi^T_{\partial_z}(\Lambda_1)$ of Λ_1 in the positive z-direction. Note that the mixed chords on $\Lambda_0 \cup \Lambda_1$ are in bijective correspondence with the mixed chords on $\Lambda_0 \cup \phi^T_{\partial_z}(\Lambda_1)$ for any choice of $T \in \mathbf{R}$. In the following we take $T \gg 0$ to be sufficiently large, so that no chord starts on $\phi^T_{\partial_z}(\Lambda_1)$ and ends on Λ_0. Of course, the length of a mixed chord c above depends on the parameter T and will not be equal to the action $\mathfrak{a}(c)$ defined above; the relation between action and length is given by

$$\ell(c) = \mathfrak{a}(c) + T.$$

The remaining Reeb chords on the link $\Lambda_0 \cup \phi^T_{\partial_z}(\Lambda_1)$ have both endpoints either on Λ_0 or $\phi^T_{\partial_z}(\Lambda_1)$, and are called **pure**. Note that the Reeb chords on $\phi^T_{\partial_z}(\Lambda_1)$ are in bijective correspondence with those of Λ_1. In fact, their Chekanov–Eliashberg algebras are even canonically isomorphic.

The differential is the Linearised Legendrian contact homology differential induced by a choice of almost complex structure, together with the augmentations ε_i for the Chekanov–Eliashberg algebras $\mathcal{A}(\Lambda_i)$ generated by the pure chords. This version of a Floer complex defined via the Chekanov–Eliashberg algebra was originally considered in [13]; also see [4] for a more recent realisation. We now give a sketch of the definition of the differential. It is roughly speaking defined by counts of rigid pseudoholomorphic discs in $(\overline{W}, d\lambda)$, for some choice of compatible almost complex structure, where the disc has

- boundary on the exact Lagrangian immersion $\Pi_{\overline{W}}(\Lambda_0 \cup \phi^T_{\partial_z}(\Lambda_1)) \subset (\overline{W}, \lambda)$;
- precisely one positive puncture at a double point which corresponds to a mixed chord—this is the input;
- precisely one negative puncture at a double point which corresponds to a mixed chord—this is the output; and
- several additional negative punctures at double points which correspond to pure chords.

By positive (resp. negative) boundary puncture, one means a point where the boundary of the pseudoholomorphic disc makes a jump that increases (resp. decreases) the z-value of the Legendrian $\Lambda_0 \cup \phi^T_{\partial_z}(\Lambda_1) \subset \overline{W} \times \mathbf{R}_z$ when following the boundary according to the orientation of the disc induced by the almost complex structure. When counting the strip, one weighs the count by the value of the augmentation ε_i on the pure chords from the last point. This is a part of the so-called linearised differential induced by the augmentation, as defined in [6]; also see the notion of the bilinearised Legendrian contact homology as defined by Bourgeois–Chantraine in [3].

From positivity of symplectic area of such pseudoholomorphic discs together with Stokes' theorem one obtains that the Reeb chord length of the input chord must be larger than the Reeb chord of the output. In other words, the complex is strictly filtered in the sense defined in Sect. 2.2, and the Reeb chords constitute a compatible basis.

From the index formula for the expected dimension of the moduli space of pseudoholomorphic discs, it follows that the degree of the input is one greater than the degree of the output; i.e. the differential is of degree -1.

2.5. Maslov potential and grading

The Maslov potential is a useful framework for introducing gradings in Lagrangian Floer homology which originally is due to Seidel [27]. The choice of a Maslov potential gives a well-defined grading in $\mathbf{Z}$. In general the potential is only well-defined modulo the Maslov number $\mu \in \mathbf{Z}$ (the positive generator of the subgroup of $\mathbf{Z}$ which is the image of the Maslov class); in that case the grading is only defined in $\mathbf{Z}/\mu\mathbf{Z}$. Here we describe a grading for which the differential of the Floer complex considered above becomes a map of degree -1, i.e. it decreases the degree.

Assume that $\overline{W}$ has vanishing first Chern class; this is, e.g. the case when $\overline{W}$ has a symplectic trivialisation, which is automatic when $\dim_{\mathbf{R}} \overline{W} = 2$. The $\mathbf{Z}$–grading of the generators is defined as follows.

Consider the determinant bundle

$$\mathbf{C}^* \to \det_{\mathbf{C}} T\overline{W} \to \overline{W}$$

induced by some choice of a compatible almost complex structure. The quotient

$$\mathbf{C}^*/\mathbf{R}^* = (\mathbf{R}^2\backslash\{0\})/\mathbf{R}^* = \mathbf{R}P^1 = \mathbf{R}/\pi\mathbf{Z}$$

gives rise to an induced $\mathbf{R}P^1$-bundle that we denote by

$$\mathcal{L} = (\det_{\mathbf{C}} T\overline{W})/\mathbf{R}^* \to \overline{W}.$$

Note that the bundle $\mathcal{L}$ is trivial when $\overline{W}$ has vanishing first Chern class (actually, the first Chern class being two-torsion is sufficient). In this case there might be several choices of homotopy classes of trivialisations.

First, we make the choice of a trivialisation of the above determinant bundle. This choice gives rise to a trivialisation $\mathcal{L} = \mathbf{R}P^1 \times \overline{W} \to \overline{W}$ of the $\mathbf{R}P^1$-bundle as well. Then, taking the fibre-wise universal cover of this trivial $\mathbf{R}P^1$-bundle, we obtain the affine $\mathbf{R}$-bundle $\tilde{\mathcal{L}} = \mathbf{R} \times \overline{W} \to \overline{W}$. The fibre of this bundle is thus the choice of an $\mathbf{R}$-lift of the angle in $\mathbf{R}P^1 = \mathbf{R}/\pi\mathbf{Z}$ of an unoriented real line.

Second, one makes the choice of a **Maslov potential** for each of the Legendrians Λ_i. This is the lift of the canonically defined section

$$(\det_{\mathbf{R}} T\Pi_{\overline{W}}(\Lambda_i))/\mathbf{R}^* \subset (\det_{\mathbf{C}} T\overline{W})/\mathbf{R}^*$$

along Λ_i of the above $\mathbf{R}P^1$-bundle $\mathcal{L}$ to the associated $\mathbf{R}$-bundle $\tilde{\mathcal{L}}$. Recall that a non-zero Maslov class is the obstruction to the existence of such a lift. When a Maslov potential exists and the Legendrian is connected, there is a natural free and transitive $\mathbf{Z}$-action on its Maslov potentials.

Given choices of Maslov potentials, the grading of a generator $c \in CF_*((\Lambda_0, \varepsilon_0), (\Lambda_1, \varepsilon_1))$ is finally obtained in the following manner. Denote by $\tilde{\varphi}_i \in \tilde{\mathcal{L}}_c$ the $\mathbf{R}$-lift of the angle of the real determinant line

$$(\det_{\mathbf{R}} T_c\Pi_{\overline{W}}(\Lambda_i))/\mathbf{R}^* \subset (\det_{\mathbf{C}} T_c\overline{W})/\mathbf{R}^*$$

specified by the choices of Maslov potentials. Consider a compatible almost complex structure J on $T_c\overline{W}$ for which $J \cdot T_c\Pi_{\overline{W}}(\Lambda_0) = T_c\Pi_{\overline{W}}(\Lambda_1)$ together with the induced family of Lagrangian planes $e^{it}T_c\Pi_{\overline{W}}(\Lambda_0) \in T_c\overline{W}$, $t \in [0, \pi/2]$, that joins $T_c\Pi_{\overline{W}}(\Lambda_0)$ to

$$e^{i\pi/2}T_c\Pi_{\overline{W}}(\Lambda_0) = J \cdot T_c\Pi_{\overline{W}}(\Lambda_0) = T_c\Pi_{\overline{W}}(\Lambda_1).$$

There is a continuous path of real determinant lines $\varphi_0^t \in \mathcal{L}_c$; denote by $\tilde{\varphi}_0^t \in \tilde{\mathcal{L}}_c$ the continuous lift to the fibre-wise universal cover $\mathbf{R} \to \mathbb{R}P^1$, where $\tilde{\varphi}_0^0 = \tilde{\varphi}_0$. In particular, $\tilde{\varphi}_0^{\pi/2}$ is a lift of the determinant line φ_1. The **degree of the generator** c is finally defined by

$$|c| = (\tilde{\varphi}_0^{\pi/2} - \tilde{\varphi}_1)/\pi \in \mathbf{Z}.$$

In the below examples we provide some useful techniques for specifying Maslov potentials and computing degrees in the cases that we are interested in here.

Example 2.8.

(1) In the case of $\overline{W} = T^*M$ there is a trivialisation of $\det_{\mathbf{C}} T(T^*M)$ in which the tangent planes to the zero section all coincide with the real part $\mathbf{R}^* \subset \mathbf{C}^*$ of the fibres. The zero-section $\Lambda_0 = j^1 0$ can be induced with the Maslov potential $\tilde{\varphi}_0$ which is zero in each $\mathbf{R}$-fibre of $\tilde{\mathcal{L}}$.

(2) A choice of Maslov potential for a general Legendrian $\Lambda_1 \subset J^1M$ for the trivialisation from Part (1) above (if it exists) can be described by comparing it to the canonical Maslov potential for the zero section $j^1 0$. More precisely, the difference between the Maslov potential at a point $x \in \Lambda_1$ for which $T_x \Pi_{\overline{W}}(\Lambda_1)$ is transverse to the Lagrangian fibre of T^*M and the canonical Maslov potential for $j^1 0 = \Lambda_0$ at the point $p(x) \in \Lambda_0$, where $p\colon J^1M \to M$ is the bundle projection, can be described by an integer $m(x) \in \mathbf{Z}$ in the following manner.
The fibre-wise rescaling of J^1M induces an isotopy of Legendrian tangent planes that isotopes any tangent plane $T_x\Lambda_1$ which is transverse to the fibre to the tangent plane $T_{p(x)} j^1 0$ of the zero section. There is an induced continuous path of determinant lines $\varphi_1^t \in \mathcal{L}$ where

$$\varphi_1^0 = (\det_{\mathbf{R}} T_x \Pi_{\overline{W}}(\Lambda_1))/\mathbf{R}^* \subset (\det_{\mathbf{C}} T_x(T^*M))/\mathbf{R}^*,$$

and φ_1^1 is the determinant line of the zero-section at the point $p(x)$. Consider the continuous choice of lifts $\tilde{\varphi_1^t} \in \tilde{\mathcal{L}}$ that extend the choice of Maslov potential for Λ_1. The difference

$$m(x) = (\tilde{\varphi}_1^1 - \tilde{\varphi}_0)/\pi \in \mathbf{Z}$$

is an integer that uniquely recovers the choice of Maslov potential at $x \in \Lambda_1$ (for x where the Lagrangian projection is transverse to the fibre).
The integer $m(x)$ is locally constant in the open subsets of Λ_1 for which the Lagrangian projection is transverse to the fibres, and changes by $+1$ as one traverses a cusp-edge in the direction of decreasing z-value. This is illustrated in Fig. 6.

(3) In the case when $\det_{\mathbf{C}} T\overline{W}$ is trivial, the homotopy classes of trivialisations of $\det_{\mathbf{C}} T\overline{W}$ are in bijection with homotopy classes of maps $\overline{W} \to \mathbf{C}^*$, which is the same as classes in $H^1(\overline{W}; \mathbf{Z})$. In the particular case $\overline{W} = T^*S^1$, the description of the Maslov potential given in Part (2) is only valid above a simply connected subset, e.g. $T^*(-\pi, \pi) \subset T^*S^1$. The Maslov potential for a general Legendrian in this setting can be described by the choice of numbers $m(x)$ as above, that, however, satisfy the additional property that they make a jump by a fixed value $l \in 2\mathbf{Z}$ when traversing the hypersurface $\{\theta = \pi\}$ in the direction of increasing θ-value. (The case $l = 0$ corresponds to the canonical trivialisation for which the zero-section admits a Maslov potential.) This is illustrated in the top of Fig. 7.

(4) Let $\Lambda_i \subset J^1M$, $i = 0, 1$, be two Legendrians with choices of Maslov potentials that have the form $j^1 f_i$ over some subset in M (i.e. the Lagrangian projections are transverse to the fibre there), where the Maslov potentials are determined by integers $m_i \in \mathbf{Z}$ in the manner described above. If $f_1 - f_0$ has a non-degenerate critical point at $p \in M$ (i.e. there is a transverse Reeb chord c there), then the above degree formula becomes

$$|c| = \operatorname{index}_p^{\text{Morse}}(f_1 - f_0) + m_0 - m_1$$

where the first term on the right-hand side is the Morse index of the critical point. We refer to [11, Lemma 3.4] for the computation.

Lemma 2.9.

(1) Let $\phi^1 : \overline{W} \times \mathbf{R} \to \overline{W} \times \mathbf{R}$ be the time-one map of a compactly supported contact isotopy. For any choice of Maslov potential on the Legendrian Λ there an induced Maslov potential on its image $\phi^1(\Lambda) \subset \overline{W} \times \mathbf{R}$ uniquely defined by the property that the Maslov potentials extend over the exact Lagrangian cobordism from Λ to $\phi^1(\Lambda)$ induced by the isotopy.

(2) If ϕ^1 is a generic C^1-small contact isotopy, then the small chords of $\Lambda \cup \phi^1(\Lambda)$ are in bijective correspondence with the critical points of a C^1-small Morse function $f : \Lambda \to \mathbf{R}$, and the above grading coincides with the Morse index, if $\phi^1(\Lambda)$ is endowed with the Maslov potential induced from Λ via the isotopy ϕ^t as in Part (1).

Proof. (1) The trace of the Legendrian isotopy can be made into a Lagrangian cylinder inside the symplectisation

$$(\mathbf{R}_t \times \overline{W} \times \mathbf{R}_z, d(e^t \alpha_{\text{st}}))$$

with cylindrical ends over the initial and final Legendrian; see work [5] by Chantraine. The Maslov potential of Λ induces a Maslov potential on the negative end of this cobordism. This Maslov potential can be extended to the entire cobordism by elementary topology (it is a Lagrangian cylinder). The induced Maslov potential on the positive end is the sought Maslov potential on $\phi^1(\Lambda)$.

(2) This computation is standard, and can be performed in a small neighbourhood of Λ. In particular, for a small perturbation of the zero-section $j^1 0 \subset J^1 M$ by a section $j^1 f$, this is an immediate consequence of Part (4) of Example 1. In general, recall that any Legendrian Λ has a standard neighbourhood which is contactomorphic to a neighbourhood of the zero section $j^1 0 \subset J^1\Lambda$, under which Λ, moreover, is identified with $j^1 0$; see [18]. The perturbation can be assumed to be given by the one-jet $j^1 f$ of some C^1-small smooth function $f : \Lambda \to \mathbf{R}$ in the same neighbourhood. □

3. Examples that exhibit unbounded spectral norms

The following basic auxiliary results facilitate our computations, and will be invoked repeatedly.

Lemma 3.1.

(1) *Let* $\phi^t\colon \Lambda_0 \hookrightarrow \overline{W} \times \mathbf{R}$ *be a Legendrian isotopy of a closed Legendrian* Λ_0 *that admits a Maslov potential, and endow* $\phi^1(\Lambda_0)$ *with the Maslov potential induced from* Λ_0 *via the isotopy, as described in Part (1) of Lemma 2.9. Further assume that* Λ_0 *has no Reeb chords. If the complex* $CF(\Lambda_0, \phi^1(\Lambda_0))$ *in degrees* 0 *and* $\dim \Lambda_0$ *consists of unique Reeb chord generators* c *and* d*, then the spectral range satisfies*

$$\rho(CF(\Lambda_0, \phi^1(\Lambda_0))) \geq |\ell(c) - \ell(d)|.$$

(In fact, it is even true that the spectral range is equal *to* $\ell(c) - \ell(d)$*, where this quantity, moreover, is positive, but we will not show this.)*

(2) *Consider a Floer complex* $CF(\Lambda_0, \Lambda_1)$ *which is* $\mathbf{Z}$*-graded and acyclic. Furthermore, assume that there is a choice of symplectic trivialisation and Maslov potential for which there are no generators in degrees* $i+1$ *or* $i-2$*, while there are unique Reeb chords* c, d *in the degrees* $|c| = i$ *and* $|d| = i-1$*. Then the boundary depth satisfies the bound*

$$\beta(CF(\Lambda_0, \Lambda_1)) \geq \ell(c) - \ell(d).$$

Proof. (1): This follows from invariance properties of the Floer homology. Note that the homology of $CF(\Lambda_0, \Lambda_0)$ has unique generators in degrees 0 and $\dim \Lambda$ which represent the point class and fundamental class in Morse homology. It follows by degree reasons that the Reeb chord generators c and d must both be cycles which are not boundaries. The two corresponding semi-infinite bars in the barcode have endpoints that are separated by precisely $|\ell(c) - \ell(d)|$ as sought.

(2): Acyclicity together with the degree assumptions implies that $\partial c = d$. The statement then follows by the second part of Corollary 2.6 since the Reeb chords form a compatible basis. □

3.1. Legendrian isotopy of the unknot (Proof of Part (2) of Theorem A)

Consider the contact manifold $J^1\mathbf{R} = \mathbf{R}_q \times \mathbf{R}_p \times \mathbf{R}_z$ with coordinates q, p, z and contact form $dz - p\,dq$. Under the quotient $\mathbf{R}_q \to \mathbf{R}/2\pi\mathbf{Z} = S^1$ we obtain the angular coordinate θ induced by $\theta \equiv q \mod 2\pi$. In other words, the aforementioned contact manifold $J^1\mathbf{R}$ is the universal cover of the contact manifold $J^1S^1 = S^1 \times \mathbf{R}_p \times \mathbf{R}_z$ equipped with the standard contact form $dz - p\,d\theta$.

First consider the standard Legendrian unknot $\Lambda_{\mathrm{st}} \subset J^1S^1$ with front projection as shown in Fig. 6, which thus is contained inside the subset $J^1(-\pi, \pi) \subset J^1S^1$. The p-coordinate of this particular representative can be seen to be estimated in terms of the ratio of a and b, which yields

$$\Lambda_{\mathrm{st}} \subset \{|p| \leq 2a/b\}.$$

Recall the well-known fact that Λ_{st} has vanishing Maslov class and hence admits a Maslov potential; see Fig. 6. Further, this Legendrian has a unique transverse Reeb chord and its Chekanov–Eliashberg algebra is equal to the polynomial algebra in one variable of degree 1 with no differential (either for $\mathbf{k} = \mathbf{Z}_2$ or for arbitrary $\mathbf{k}$ and the choice of bounding spin structure);

see [14]. In particular, its Chekanov–Eliashberg algebra admits the trivial augmentation.

We also fix a Legendrian fibre

$$F = F_{(\pi/4,0)} = \{\pi/2\} \times \mathbf{R}_p \times \{0\} \subset J^1(-\pi, \pi) \subset J^1S^1.$$

Note that the Reeb chords between any Legendrian Λ and F are in bijective correspondence with the intersection points of Λ and the hypersurface $\{\theta = \pi/4\}$. Note that the image of F under the front projection is given by the point $\{(\pi/4, 0)\}$; Reeb chords correspond to lines contained inside $\{\theta = \pi/4\}$ in the front projection that have one endpoint on $\{(\pi/4, 0\})$ and one endpoint on the projection of Λ. These chords are depicted in Fig. 6.

Since F that has no Reeb chords, its Chekanov–Eliashberg algebra trivially admits an augmentation. We can thus define the Floer homology complex $CF(\Lambda_{\mathrm{st}}, F)$ which is generated by two Reeb chords c and d, where $0 > \ell(c) > \ell(d)$ and $|c| = |d| + 1$. Note that $CF(\Lambda_{\mathrm{st}}, F)$ is an acyclic complex by invariance under Legendrian isotopy; after shrinking the unknot sufficiently, all mixed chords disappear.

The goal is to construct a Legendrian isotopy $\Lambda^t_{\mathrm{st}} \subset J^1S^1$ of the unknot confined to the subset

$$\{|p| \leq 2a/b\} \subset J^1S^1$$

for which the boundary depth of $CF(\Lambda^T_{\mathrm{st}}, F)$ becomes arbitrarily large as $t \to +\infty$. This isotopy will be constructed as the projection of an isotopy $\tilde{\Lambda}^t_{\mathrm{st}} \subset J^1\mathbf{R}$ of the unknot inside the universal cover $J^1\mathbf{R} \to J^1S^1$. In fact, the Legendrian isotopy $\tilde{\Lambda}^t_{\mathrm{st}}$ is very simple; it is the rescaling of

$$\tilde{\Lambda}_{\mathrm{st}} = \Lambda_{\mathrm{st}} \subset J^1(-\pi, \pi) \subset J^1\mathbf{R}$$

under the contact isotopy $(q, p, z) \mapsto (e^t \cdot q, p, e^t \cdot z)$ defined on the universal cover; note that this contact isotopy simply rescales the front projection.

It is easy to check that $CF(\tilde{\Lambda}^t_{\mathrm{st}}, F)$ satisfies the property that the boundary depth goes to $+\infty$ as $t \to +\infty$. Indeed, these complexes are generated by the two unique transversely cut out Reeb chords c_t and d_t between $\tilde{\Lambda}^t_{\mathrm{st}}$ and F for all values $t > 0$. These chords, moreover, satisfy the property that $\ell(c_t) - \ell(d_t)$ becomes arbitrarily large as $t \to +\infty$; c.f. Part (2) of Lemma 3.1.

What remains to prove is the following two claims for the projection $\Lambda^t_{\mathrm{st}} \subset J^1S^1$ of the Legendrian rescaling $\tilde{\Lambda}_t \subset J^1\mathbf{R}$. First, we claim that Λ^t_{st} indeed is a Legendrian isotopy. Second, we show that the boundary depth of $CF(\Lambda^t_{\mathrm{st}}, F)$ goes to $+\infty$ as $t \to +\infty$

The fact that Λ^t_{st} is a Legendrian isotopy can be seen by considering the sequence of front projections; see Figs. 7 and 8. Except for an isotopy of the front, the front also undergoes a sequence *RIII*-moves together with the composite move shown in Fig. 5. The Lagrangian projection of $\tilde{\Lambda}_2$ is shown in Fig. 9.

Then we need to estimate the boundary depth of the sequence of Floer complexes $CF(\Lambda^t_{\mathrm{st}}, F)$. In addition to Reeb chords c_t and d_t, which correspond to the Reeb mixed Reeb chords on the lift and have exactly the same

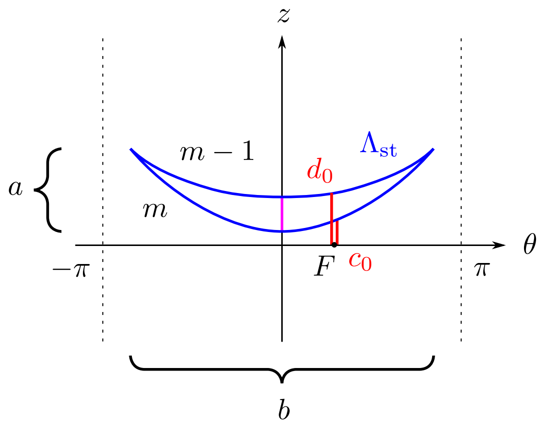


FIGURE 6. The standard Legendrian unknot Λ_{st} and the Legendrian fibre F. Note that there are precisely two transverse Reeb chords c_0, d_0 between F and Λ_{st}. The choice of $m \in \mathbf{Z}$ determines a Maslov potential on Λ_{st} as described in Part (2) of Example 1

actions, there are additional Reeb chords between Λ^t_{st} and F that appear as $t \to +\infty$. Nevertheless, we claim that the boundary depth of $CF(\Lambda^t_{\mathrm{st}}, F)$ still is bounded from below by the boundary depth $\beta(CF(\tilde{\Lambda}^t_{\mathrm{st}}, F))$.

To see the last claim, we will consider different gradings of the complexes $CF(\Lambda^t_{\mathrm{st}}, F)$, obtained by changing the symplectic trivialisation of T^*S^1. Note that Λ_{st} is null-homotopic inside J^1S^1 and thus has a vanishing Maslov class independently of the choice of symplectic trivialisation. Moreover, the chords c_t and d_t always satisfy $|c_t| - |d_t| = 1$ regardless of the choice of Maslov potential and symplectic trivialisation; see the top of Fig. 7.

We claim that, after changing the symplectic trivialisation of T^*S^1 by introducing a sufficiently large number $l/2 \gg 0$ of full rotations of the standard symplectic frame as one traverses the hypersurface $\{\theta = \pi\}$ in the direction of increasing θ-coordinate, all generators c' in the complex except different from c_t and d_t acquire degrees that satisfy

$$|c'| - |c_t| \notin [-10, 10].$$

To see this, we note that the Maslov potential of these sheets acquire an additional term kl where $k \in \mathbf{Z} \backslash \{0\}$; see Parts (3) and (4) of Example 1.

Since these degree properties can be achieved, the statement now follows directly by Part (2) of Lemma 3.1. □

3.2. Legendrian isotopy of the zero-section (Proof of Part (1) of Theorem A)

We use the same coordinates as in the above Sect. 3.1. In fact, the sought Legendrian isotopy is also constructed in a manner similar to the construction of Λ_t given there, by performing a rescaling of a part of the front inside the universal cover $J^1\mathbf{R}$ (and then projecting back to J^1S^1). The isotopy is shown in Figs. 10 and 11. One starts by considering a Legendrian perturbation

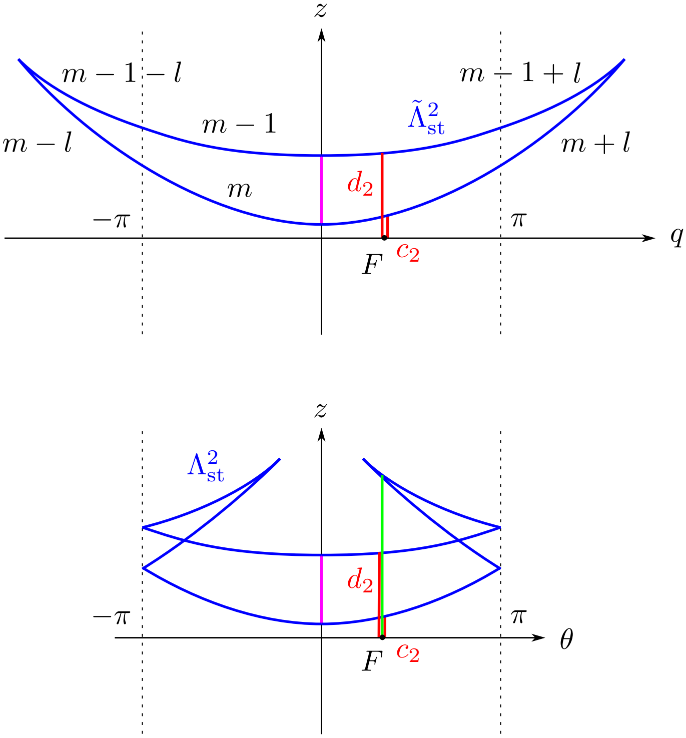


FIGURE 7. Above: $\tilde{\Lambda}^2_{\mathrm{st}}$ has a front which is a linear rescaling of the front of Λ_{st} inside $J^1\mathbf{R}$. The number m defines a choice of Maslov potential for Λ^2_{st}, where $l \in 2\mathbf{Z}$ depends on the homotopy class of the trivialisation of $\det_{\mathbf{C}}(T(T^*S^1))$. Below: Λ^2_{st} is the projection of $\tilde{\Lambda}^2_{\mathrm{st}}$ inside J^1S^1. Except for the mixed chords c_t and d_t that exist for the lift, there are now additional mixed chords

j^1f of j^10 which has precisely two chords. Then one performs a RII-move. Rescaling the front of the Legendrian introduced by the RII-move in the universal cover $\mathbf{R}^2$ and then projecting back to $S^1 \times \mathbf{R}$ is again a Legendrian isotopy. In Fig. 11 one sees that there are exactly two chords between j^10 and the produced Legendrians, while the difference in action between these two generators grows indefinitely as $t \to +\infty$. □

3.3. Hamiltonian isotopy on the punctured torus (Proof of Theorem D)

Here we consider the exact Lagrangian embedding $L \subset (\Sigma_{1,1}, d\lambda)$ of S^1 which is given as the image of $\{p = 0\} \subset \mathbf{R}^2$ under the quotient construction in Sect. 2.1.2; see Fig. 13. We perform a Hamiltonian perturbation L' that intersects the original Lagrangian transversely in precisely two points c and d. The spectral norm is thus $\gamma(CF(L, L')) = \ell(c) - \ell(d)$.

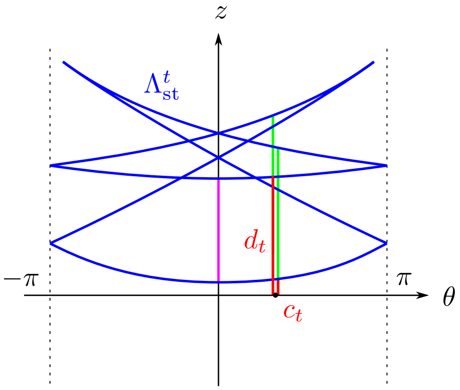


FIGURE 8. This shows the projection Λ^t_{st} of the rescaling $\tilde{\Lambda}^t_{\mathrm{st}}$ under the universal cover $J^1\mathbf{R} \to J^1S^1$

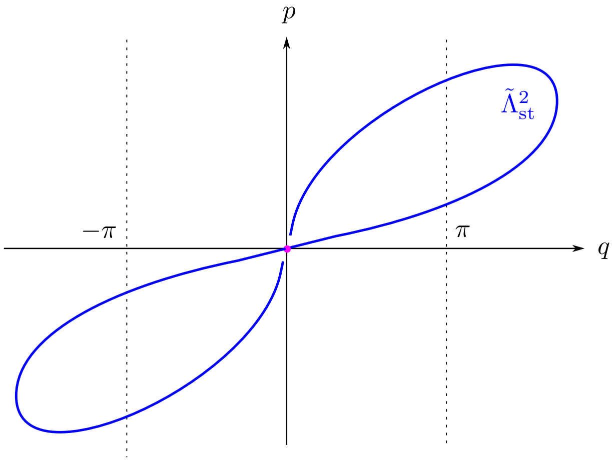


FIGURE 9. The figure depicts the Lagrangian projection of $\tilde{\Lambda}^2_{\mathrm{st}}$ to $T^*\mathbf{R}$. The Lagrangian projection of Λ^2_{st} to T^*S^1 is induced by the quotient projection $\mathbf{R} \to S^1$. The Lagrangian projection of $\tilde{\Lambda}^t_{\mathrm{st}}$ is obtained by rescaling the q-coordinate of $T^*\mathbf{R}$ followed by the canonical projection to T^*S^1

Then consider the autonomous Hamiltonian

$$\rho\colon \Sigma_{1,1} \to \mathbf{R}_{\leq 0}$$

with support inside $\{q \in [-\delta, \delta]\}$ for some small $\delta > 0$, and which is equal to the smooth bump-function $\rho(q) \leq 0$ in one variable of the form

- $\rho(q) \equiv -1$ in a neighbourhood of $q = 0$;
- $\rho(q) = \rho(-q)$;
- and $\rho'(q) \leq 0$ for $q < 0$.

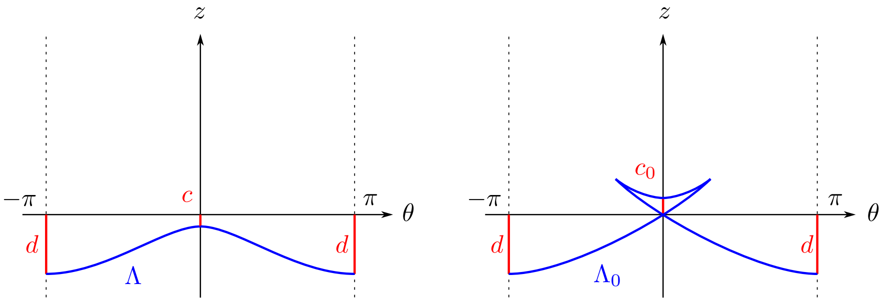


FIGURE 10. Left: a Legendrian perturbation of the zero section. The vertical chords denote the two Reeb chords between the zero-section $j^1 0$ and the perturbation. Right: the perturbed version of the zero-section after a suitable Legendrian RI-move

The Hamiltonian isotopy ϕ_ρ^t wraps the region $q \in (-\delta, 0)$ in the negative p-direction, while it wraps the region $q \in (0, \delta)$ in the positive p-direction.

We claim that $CF(L, \phi_\rho^t(L'))$ has a spectral norm which becomes arbitrarily large as $t \to +\infty$. What is clear is that $\ell(c) - \ell(d) \to +\infty$ as $t \to +\infty$. (Use, e.g. Lemma 2.1.) Again there are additional generators that appear as $t \to +\infty$, so knowing that $\ell(c) - \ell(d) \to +\infty$ is not sufficient.

As in Sect. 3.1 a change of symplectic trivialisation can again give us what we need. First consider the canonical symplectic trivialisation, induced by the trivialisation of $\mathbf{R}^2$ and the quotient projection. Then deform this trivialisation by adding a number $l/2 \gg 0$ of full rotations of the standard symplectic frame (relative the constant one) as one traverses the $\{p = 1\}$. Note that the Lagrangian corresponding to $\{p = 0\}$ still has a Maslov potential after this change of trivialisation. Similarly to the computation in Sect. 3.1, it is now readily seen that all generators c' different from c and d satisfy the property that

$$|c'| - |c| \notin [-10, 10],$$

after we have chosen $l \gg 0$ sufficiently large. In the meantime, $|c| - |d| = 1$ is always satisfied.

The spectral norm can now finally be computed by invoking Part (1) of Lemma 3.1.

4. Proof of Theorem C

By definition, our two Floer complexes are the linearised Legendrian contact homology complexes generated as a $\mathbf{k}$-vector space by the mixed Reeb chords on the Legendrian link

$$\Lambda_\pm \cup \phi_{\partial_z}^T(\Lambda).$$

Here $T \gg 0$ is fixed but sufficiently large.

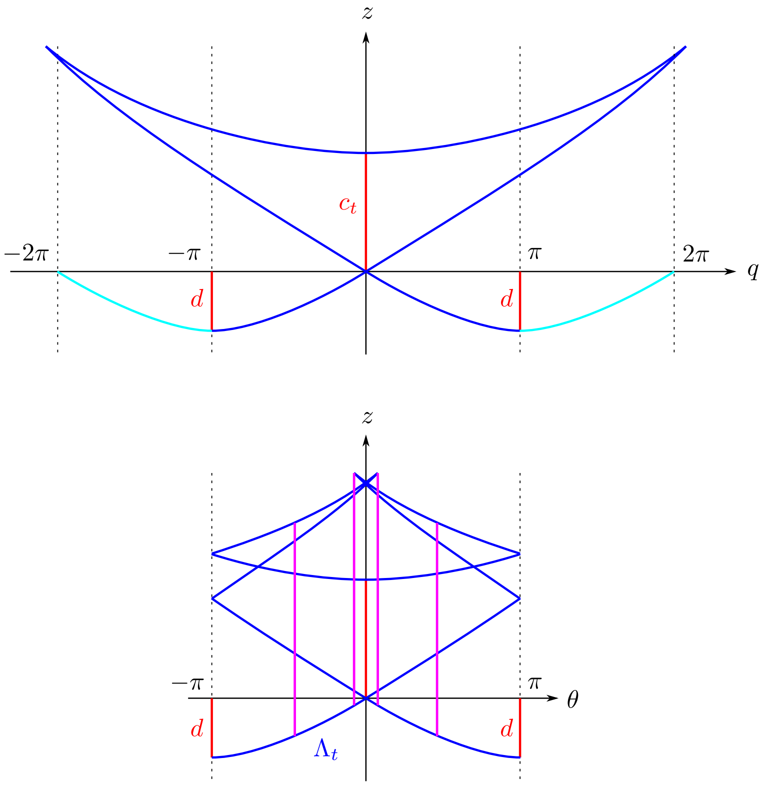


FIGURE 11. Λ_t is obtained from Λ_0 by a linear rescaling of the front inside $\{z \geq 0\}$ in the universal cover $J^1\mathbf{R}^2$ followed by the canonical projection $J^1\mathbf{R} \to J^1 S^1$. The front of Λ_t undergoes the composite move shown in Fig. 5 consisting of two consecutive *RII*-moves along with *RIII*-moves

The cusp-connected sum performed on $\Lambda_- \cup \phi^T_{\partial_z}(\Lambda)$ produces $\Lambda_+ \cup \phi^T_{\partial_z}(\Lambda)$ (of course, only the first component is affected). There is an associated exact standard Lagrangian handle-attachment cobordism

$$\mathcal{L} \subset (\mathbf{R}_t \times \overline{W} \times \mathbf{R}_z, d(e^t \alpha_{\mathrm{st}}))$$

inside the symplectisation as constructed in [9]. This is a cobordism with cylindrical ends from

$$\Lambda_- \cup \phi^T_{\partial_z}(\Lambda) \ \text{ to } \ \Lambda_+ \cup \phi^T_{\partial_z}(\Lambda),$$

i.e. from the Legendrian link before surgery (at the concave end) to the link after surgery (at the convex end). One component of this cobordism is simply the trivial cylinder $\mathbf{R} \times \phi^T_{\partial_z}(\Lambda)$. This Lagrangian cobordism induces a unital DGA-morphism

$$\Phi_{\mathcal{L}} \colon \mathcal{A}(\Lambda_+ \cup \phi^T_{\partial_z}(\Lambda)) \to \mathcal{A}(\Lambda_- \cup \phi^T_{\partial_z}(\Lambda))$$

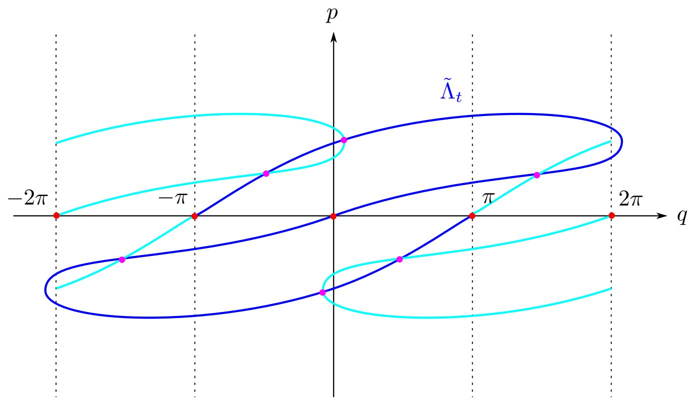


FIGURE 12. The Lagrangian projection in $T^*\mathbf{R}$ of the universal cover $\tilde{\Lambda} \subset J^1\mathbf{R}$ of $\Lambda_t \subset J^1S^1$ shown in Fig. 11, where $\tilde{\Lambda}_t \cong \mathbf{R}$. The interval shown in dark blue is a fundamental domain for $\tilde{\Lambda}_t$

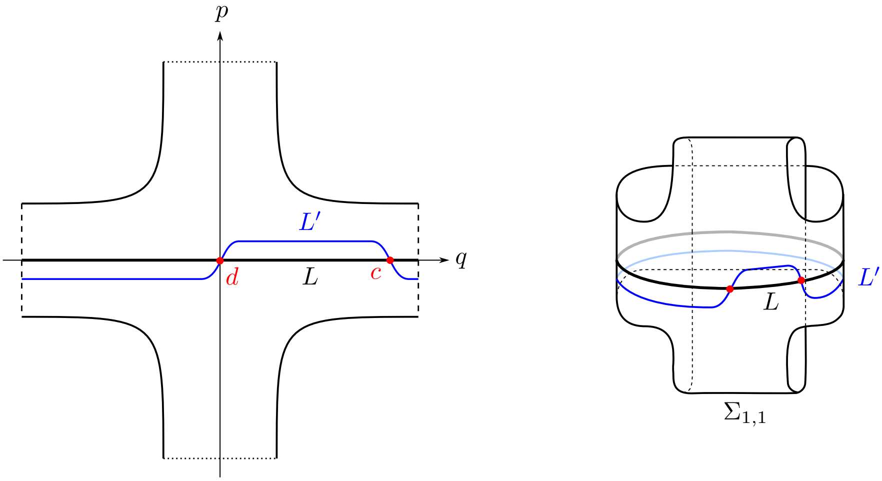


FIGURE 13. The left depicts a domain in $\mathbf{R}^2$ with piecewise smooth boundary. After identifying the two horizontal pieces of the boundary, as well as the two vertical pieces, one obtains the Liouville domain shown on the right, with Liouville form described in Sect. 2.1.2. The closed exact Lagrangian L is the image of $\{p = 0\}$ and L' is a small Hamiltonian perturbation of L

of the Chekanov–Eliashberg algebras. In particular, the choice of augmentation ε_- of the Chekanov–Eliashberg algebra of Λ_- pulls back to an augmentation $\varepsilon_+ = \varepsilon_- \circ \Phi_{\mathcal{L}}$ of the Chekanov–Eliashberg algebra of Λ_+.

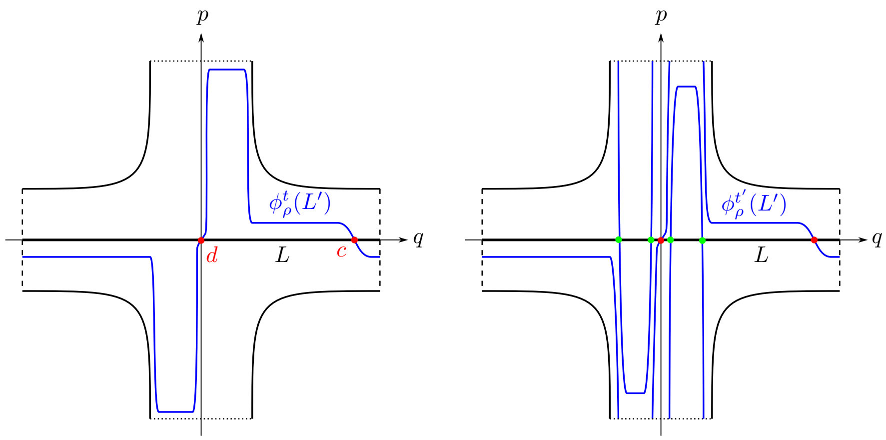


FIGURE 14. A Hamiltonian isotopy that wraps the Lagrangian $\{p = 0\}$ around the one-handle with core $\{q = 0\}$, while fixing a neighbourhood of the latter core. Note that the Hamiltonian function is positive but constant near $q = 0$. Here $t' > t$

The above DGA-morphism $\Phi_{\mathcal{L}}$ of the Chekanov–Eliashberg algebras after and before the surgery was computed in [9, Theorem 1.1] under the assumption that the handle-attachment is sufficiently small. This computations in particular shows that the mixed chords c on $\Lambda_+ \cup \phi^T_{\partial_z}(\Lambda)$ are mapped to

$$\Phi_{\mathcal{L}}(c) = c + \sum_i r_i \mathbf{d}_i, \;\; r_i \in \mathbf{k},$$

where $\mathbf{d}_i$ are words of Reeb chords that each contain an *odd* number of mixed chords of $\Lambda_- \cup \phi^T_{\partial z}(\Lambda)$, and in which every mixed chord, moreover, is of length strictly less than $\ell(c)$. It now follows by pure algebraic considerations that the map

$$CF_*((\Lambda_+, \varepsilon_+), (\Lambda, \varepsilon)) \to CF_*((\Lambda_-, \varepsilon_-), (\Lambda, \varepsilon))$$

induced by linearising the DGA-morphism $\Phi_{\mathcal{L}}$ using the augmentations ε, ε_+, and ε_- (see [3] and [4]) is an action-preserving isomorphism of the Floer complexes as claimed. □

Acknowledgements

The author is grateful to Rémi Leclercq for teaching him the basics about spectral invariants and Viterbo's conjecture, in a café somewhere on the Rive Gauche. The author would also like to thank the anonymous referee for carefully reading the manuscript.

Funding Open access funding provided by Uppsala University.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Barannikov, S.A.: The framed Morse complex and its invariants. In: Singularities and Bifurcations, Volume 21 of Advances in Soviet Mathematics, pp. 93–115. American Mathematical Society, Providence (1994)

[2] Biran, P., Cornea, O.: Bounds on the Lagrangian spectral metric in cotangent bundles. https://arxiv.org/abs/2008.04756 [math.SG] (2020)

[3] Bourgeois, F., Chantraine, B.: Bilinearized Legendrian contact homology and the augmentation category. J. Symplectic Geom. **12**(3), 553–583 (2014)

[4] Chantraine, B., Dimitroglou Rizell, G., Ghiggini, P., Golovko, R.: Floer theory for Lagrangian cobordisms. J. Differ. Geom. **114**(3), 393–465 (2020)

[5] Chantraine, B.: Lagrangian concordance of Legendrian knots. Algebr. Geom. Topol. **10**(1), 63–85 (2010)

[6] Chekanov, Y.: Differential algebra of Legendrian links. Invent. Math. **150**(3), 441–483 (2002)

[7] Chekanov, Yu.V.: Critical points of quasifunctions, and generating families of Legendrian manifolds. Funktsional. Anal. i Prilozhen. **30**(2), 56–69 (1996). (**96**)

[8] Carlsson, G., Zomorodian, A., Collins, A., Guibas, L.: Persistence barcodes for shapes. In: Scopigno, R., Zorin, D. (eds.) Symposium on Geometry Processing. The Eurographics Association (2004)

[9] Dimitroglou Rizell, G.: Legendrian ambient surgery and Legendrian contact homology. J. Symplectic Geom. **14**(3), 811–901 (2016)

[10] Dimitroglou Rizell, G., Sullivan, M.G.: The persistence of the Chekanov–Eliashberg algebra. Selecta Math. (N.S.) **26**(5), 69 (2020)

[11] Ekholm, T., Etnyre, J., Sullivan, M.: Non-isotopic Legendrian submanifolds in R^{2n+1}. J. Differ. Geom. **71**(1), 85–128 (2005)

[12] Ekholm, T., Etnyre, J., Sullivan, M.: Legendrian contact homology in $P \times R$. Trans. Am. Math. Soc. **359**(7), 3301–3335 (2007). ((**electronic**))

[13] Ekholm, T., Etnyre, J.B., Sabloff, J.M.: A duality exact sequence for Legendrian contact homology. Duke Math. J. **150**(1), 1–75 (2009)

[14] Etnyre, J.B., Ng, L.L.: Legendrian contact homology in $\mathbb{R}^3$. https://arxiv.org/abs/1811.10966 [math.SG] (2019)

[15] Etnyre, J.B.: Legendrian and transversal knots. In: Menasco W, Thistlethwaite M (eds) Handbook of Knot Theory, pp. 105–185. Elsevier B. V., Amsterdam (2005)

[16] Floer, A.: Morse theory for Lagrangian intersections. J. Differ. Geom. **28**(3), 513–547 (1988)

[17] Floer, A.: Witten's complex and infinite-dimensional Morse theory. J. Differ. Geom. **30**(1), 207–221 (1989)

[18] Geiges, H.: An introduction to contact topology. In: Bollobás B, Fulton W, Katok A, Kirwan F, Sarnak P, Simon B, Totaro B (eds) Cambridge Studies

in Advanced Mathematics, vol. 109. Cambridge University Press, Cambridge (2008)

[19] Kislev, A., Shelukhin, E.: Bounds on spectral norms and barcodes. Geom. Topol. **25**(7), 3257–3350 (2021)

[20] Leclercq, R.: Spectral invariants in Lagrangian Floer theory. J. Mod. Dyn. **2**(2), 249–286 (2008)

[21] Legout, N.: A-infinity category of Lagrangian cobordisms in the symplectization of PxR. https://arxiv.org/abs/2012.08245 [math.SG] (2020)

[22] Lanzat, S., Zapolsky, F.: On the contact mapping class group of the contactization of the A_m-Milnor fiber. Ann. Math. Qué. **42**(1), 79–94 (2018)

[23] Oh, Y.-G.: Spectral Invariants: Applications, Volume 2 of New Mathematical Monographs, pp. 348–407. Cambridge University Press, Cambridge (2015)

[24] Polterovich, L., Rosen, D., Samvelyan, K., Zhang, J.: Topological Persistence in Geometry and Analysis, Volume 74 of University Lecture Series. American Mathematical Society, Providence (2020)

[25] Polterovich, L., Shelukhin, E.: Autonomous Hamiltonian flows, Hofer's geometry and persistence modules. Selecta Math. (N.S.) **22**(1), 227–296 (2016)

[26] Świątkowski, J.: On the isotopy of Legendrian knots. Ann. Glob. Anal. Geom. **10**(3), 195–207 (1992)

[27] Seidel, P.: Graded Lagrangian submanifolds. Bull. Soc. Math. France **128**(1), 103–149 (2000)

[28] Shelukhin, E.: Viterbo conjecture for Zoll symmetric spaces. https://arxiv.org/abs/1811.05552 [math.SG] (2018)

[29] Shelukhin, E.: Symplectic cohomology and a conjecture of Viterbo. https://arxiv.org/abs/1904.06798 [math.SG] (2019)

[30] Usher, M., Zhang, J.: Persistent homology and Floer-Novikov theory. Geom. Topol. **20**(6), 3333–3430 (2016)

[31] Viterbo, C.: Symplectic topology as the geometry of generating functions. Math. Ann. **292**(4), 685–710 (1992)

[32] Viterbo, C.: Symplectic homogenization. https://arxiv.org/abs/0801.0206 [math.SG] (2014)

[33] Zapolsky, F.: Geometry of contactomorphism groups, contact rigidity, and contact dynamics in jet spaces. Int. Math. Res. Not. IMRN **20**, 4687–4711 (2013)

[34] Zapolsky, F.: On the Hofer geometry for weakly exact Lagrangian submanifolds. J. Symplectic Geom. **11**(3), 475–488 (2013)

Georgios Dimitroglou Rizell
Department of Mathematics
Uppsala University
Box 480751 06 Uppsala
Sweden
e-mail: georgios.dimitroglou@math.uu.se

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:30
https://doi.org/10.1007/s11784-022-00944-x
Published online April 8, 2022

Journal of Fixed Point Theory and Applications

Legendrian persistence modules and dynamics

Michael Entov and Leonid Polterovich

To Claude Viterbo on the occasion of his 60th birthday.

Abstract. We relate the machinery of persistence modules to the Legendrian contact homology theory and to Poisson bracket invariants, and use it to show the existence of connecting trajectories of contact and symplectic Hamiltonian flows.

Contents

Michael Entov was partially supported by the Israel Science Foundation grant 1715/18. Leonid Polterovich was partially supported by the Israel Science Foundation grant 1102/20.
This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

1. Introduction and main results

1.1. Interlinking

In the present paper, we discuss a new facet of a method, introduced in [8,28], of finding orbits of Hamiltonian systems connecting a pair of disjoint subsets (X_0, X_1) in the phases space. The method manifests a dynamical phenomenon called *interlinking*, which involves a quadruple of subsets (X_0, X_1, Y_0, Y_1) in a symplectic manifold (M, ω) satisfying $X_0 \cap X_1 = Y_0 \cap Y_1 = \emptyset$. A Hamiltonian $H : M \times \mathbb{S}^1 \to \mathbb{R}$ *separates* Y_0, Y_1 if

$$\Delta(H, Y_0, Y_1) := \inf_{Y_1 \times \mathbb{S}^1} H - \sup_{Y_0 \times \mathbb{S}^1} H > 0. \tag{1}$$

Let $\mu > 0$. According to the definition in [28], (Y_0, Y_1) *μ-interlinks* (respectively, *autonomously μ-interlinks*) (X_0, X_1), if for every Hamiltonian (respectively, every autonomous Hamiltonian) H separating Y_0, Y_1 and generating a flow $\{\phi_t\}_{t\in\mathbb{R}}$ on M, there exist $t_0 \in \mathbb{R}$, $x \in M$ and a positive $T \leq \mu/\Delta$, so that $\phi_{t_0} x \in X_0$ and $\phi_{t_0+T} x \in X_1$. The piece of the trajectory $\{\phi_t x\}$, $t \in [t_0, t_0 + T]$ is called a *chord of time-length* T connecting X_0 and X_1.

The pair (Y_0, Y_1) *interlinks* (respectively, *autonomously interlinks*) the pair (X_0, X_1) if it μ-interlinks (respectively, autonomously μ-interlinks) it for some $\mu > 0$.

Important remark: In this paper, we will consider only autonomous interlinking and for brevity will omit the word "autonomous". **Thus, "interlinking" further on in this paper is the same as "autonomous interlinking" in** [28].

We will focus on quadruples of a special form lying in the symplectization $S\Sigma$ of a contact manifold (Σ, ξ), and its fillings. Let λ be a contact form on Σ. Recall that $S\Sigma$ is $\Sigma \times \mathbb{R}_+(s)$ equipped with the symplectic form $\omega = d(s\lambda)$. We set $Y_0 = \{s = s_-\}$ and $Y_1 = \{s = s_+\}$ for some $0 < s_- \leq s_+$, and take X_0 and X_1 to be disjoint Lagrangian cobordisms whose boundaries project to Legendrian submanifolds in Σ. We mainly concentrate on the case of cylindrical cobordisms

$$X_i = \Lambda_i \times [s_-, s_+] \tag{2}$$

for some disjoint Legendrians Λ_0, Λ_1. The main result of the present paper provides a sufficient condition for interlinking in terms of contact topology of this pair of Legendrians.

A very rough sketch of the method of [8,28] is as follows. Using Poisson bracket invariants coming from function theory on symplectic manifolds, one deduces the desired interlinking from obstructions to some special deformations of the symplectic form ω constant near $Z := X_0 \cup X_1 \cup Y_0 \cup Y_1$. The constraints on such deformations are often provided by pseudo-holomorphic curves with the boundary on Z. The main new idea of the present paper is to extract such curves from filtered Legendrian contact homology of Λ_0 and Λ_1. Thus our main point is to deduce "dynamical interlinking" from "contact-topological interlinking".

The realization of this idea requires tools from the relative symplectic field theory (RSFT) [27] combined with the theory of persistence modules which originated in topological data analysis [20,45] and brought to symplectic topology in [41]. RSFT associates with a (non-degenerate) pair formed by a Legendrian submanifold and a contact form on the manifold an algebraic object, called the Legendrian contact homology, and to an exact Lagrangian cobordism between Legendrian submanifolds a morphism between the corresponding Legendrian contact homologies. In the simplest setting (say, where a contact manifold is the contactization of an exact symplectic manifold), the enhancement of the RSFT using the action filtration gives rise to the *filtered relative symplectic field theory (FRSFT)*. Namely, the Legendrian contact homology can be viewed as a persistence module formed by vector spaces (over $\mathbb{Z}_2$), while an exact Lagrangian cobordism between Legendrian submanifolds defines a morphism between the corresponding persistence modules; see papers [16,18] by Dimitroglou Rizell and Sullivan. In the present paper, we rely on some special instances of this theory only leaving more general results for a forthcoming manuscript [30].

In fact, theory of persistence modules enables us, in certain situations, to detect chords of contact flows even in the absence of interlinking, provided that the contact Hamiltonians have a sufficiently small oscillation in the uniform norm. This *robustness* of the existence mechanism for contact chords with respect to C^0-small perturbations of contact Hamiltonians is a feature of our approach.

Let us mention also that in a special case when the cobordisms are cylindrical as in (2) and the Hamiltonian $H : S\Sigma \to \mathbb{R}$ is $\mathbb{R}_+$-homogeneous and positive, interlinking reduces to existence of Reeb chords of *a modified* contact form λ/H connecting Legendrian submanifolds Λ_0 and Λ_1. Existence of such a chord with a good upper bound on its time-length would follow from a Legendrian contact homology theory valid for arbitrary contact forms. At the moment, such a theory is still under construction, though it sounds likely that Pardon's work [40] on foundations of absolute contact homologies will be eventually extended to the relative (i.e., Legendrian) case. Meanwhile, our results yield existence of Reeb chords for arbitrary forms on certain contact manifolds. Applications of these results include, in particular, contact-topological methods in non-equilibrium thermodynamics [29].

1.2. The pool of contact manifolds

In the present paper, we focus on contact manifolds of the form $\Sigma = P \times \mathbb{R}(z)$, where $(P^{2n}, d\vartheta)$, $n \in \mathbb{Z}_{>0}$, is an exact symplectic manifold with a symplectic form $d\vartheta$ and bounded geometry at infinity (see [3, 25] for the definition of this notion) and the contact structure on Σ is defined by the contact form $\lambda = dz + \vartheta$. We will call such contact manifolds *nice*: to the best of our knowledge, this is the largest class of contact manifolds for which the details of the Legendrian contact homology theory have been worked out rigorously in the published literature. Note that the Reeb vector field R of λ is $\partial/\partial z$ and its flow has no periodic orbits.

A specific example of a nice contact manifold is the 1-jet space $J^1Q = T^*Q(p,q) \times \mathbb{R}(z)$ of a closed manifold Q equipped with the contact form $dz \pm pdq$. The forms corresponding to different choices of the sign are related by the involution $p \to -p$. We freely use both forms depending on the context, hoping that this will not cause a confusion. Note that a neighborhood of the zero section in J^1Q provides the universal model for a neighborhood of any Legendrian copy of Q in an arbitrary contact manifold [36, Example 2.5.11].

Write $ST^*\mathbb{R}^n$ for the space of co-oriented contact elements of $\mathbb{R}^n$, identified with the unit cotangent sphere bundle of $\mathbb{R}^n$ with respect to the Euclidean metric. There exists a contactomorphism

$$(J^1\mathbb{S}^{n-1}, dz - pdq) \to (ST^*\mathbb{R}^n, pdq) \tag{3}$$

(known as "the hodograph map", see, e.g., [2, pp. 48–49]) identifying the standard contact forms on both spaces: it sends a point $(p, q, z) \in J^1\mathbb{S}^{n-1}$ to the unit cotangent vector q in the cotangent space of $zq + p \in \mathbb{R}^n$. (Here, p, q, z are local Darboux vector-coordinates on $J^1\mathbb{S}^{n-1}$.)

1.3. Sample applications: chords of symplectic Hamiltonians

Before discussing our results in a general setting, let us give a few basic definitions and present a sample of dynamical applications. In Remark 1.3, we discuss the relation between these applications and the dynamical results in our previous work.

A (time-dependent) Hamiltonian on a symplectic manifold is called *complete* if its Hamiltonian flow is defined for all times. Similarly, a (time-dependent) contact Hamiltonian (with respect to a contact form) on a contact manifold is called *complete* if its contact flow is defined for all times.

We extend the definition of a chord of a symplectic Hamiltonian (see Sect. 1.1) to the contact setting as follows: given a (time-dependent) contact Hamiltonian h (with respect to a contact form λ) on a contact manifold Σ and two-disjoint subsets Z_0, Z_1 of Σ, a *chord of* h *from* Z_0 *to* Z_1 *of time-length* $T > 0$ (*with respect to* λ) is a trajectory of the contact flow of h (with respect to λ) that passes through Z_0 at a time t_0 and through Z_1 at the time $t_0 + T$. If $h > 0$, such a chord is a Reeb chord of the contact form λ/h of the same time-length. Given a chord a, we write $|a|$ for its time-length.

Consider $\mathbb{R}^{2n}(p, q) = \mathbb{R}^n(p) \times \mathbb{R}^n(q)$ with the standard symplectic form $dp \wedge dq$. We view $\mathbb{R}^{2n}$ as the symplectization of the contact manifold $ST^*\mathbb{R}^n$ filled by the zero section $\{p = 0\}$. Let $|\cdot|$ denote the Euclidean norm on $\mathbb{R}^n(q)$.

Let $0 < s_- < s_+$. Let $x_0, x_1 \in \mathbb{R}^n(q)$, $x_0 \neq x_1$. If $n = 1$, assume that $x_0 < x_1$.

Define $X_0, X_1, Y_0, Y_1 \subset \mathbb{R}^{2n}$ as follows.

If $n > 1$, set

$$\begin{aligned} X_0 &:= \{(p, x_0) \in \mathbb{R}^{2n} \mid s_- \leq |p| \leq s_+\}, \\ X_1 &:= \{(p, x_1) \in \mathbb{R}^{2n} \mid s_- \leq |p| \leq s_+\}, \\ Y_0 &:= \{|p| = s_-\}, \\ Y_1 &:= \{|p| = s_+\}. \end{aligned}$$

If $n = 1$, set

$$\begin{aligned} X_0 &:= \{(p, x_0) \in \mathbb{R}^{2n} \mid s_- \leq p \leq s_+\}, \\ X_1 &:= \{(p, x_1) \in \mathbb{R}^{2n} \mid s_- \leq p \leq s_+\}, \\ Y_0 &:= \{(s_-, q) \in \mathbb{R}^{2n} \mid x_0 \leq q \leq x_1\}, \\ Y_1 &:= \{(s_+, q) \in \mathbb{R}^{2n} \mid x_0 \leq q \leq x_1\}. \end{aligned}$$

Let $H : \mathbb{R}^{2n} \times \mathbb{S}^1 \to \mathbb{R}$ be a complete Hamiltonian.

Set

$$c_{\min} := \min_{X_0 \times \mathbb{S}^1} H, \; c_{\max} := \max_{X_0 \times \mathbb{S}^1} H.$$

Theorem 1.1. *A. Assume $\Delta(H; X_1, X_0) > 0$ and the following conditions are satisfied:*

$$H \text{ is time-independent}, \tag{4}$$

$$H|_{X_0} \geq 0, \tag{5}$$

$$\mathrm{supp}\, H \cap \{s_- \leq |p| \leq s_+\} \text{ is compact}. \tag{6}$$

Then, there exists a chord of H from Y_0 to Y_1 of time-length bounded from above by $\dfrac{|x_0 - x_1|(s_+ - s_-)}{\Delta(H; X_1, X_0)}$.

In the case $n = 1$, the claim holds even without assuming (4), (5), and (6).

B. Assume $\Delta(H; Y_0, Y_1) =: \Delta > 0$, and for some $0 < e < 1/2$, the following condition is satisfied:

$$\sup_{c_{\min} - e\Delta \leq H \leq c_{\max} + e\Delta} \left| \frac{\partial H}{\partial t} \right| < \frac{(1 - 2e)e\Delta^2}{(s_+ - s_-)|x_0 - x_1|}. \tag{7}$$

Then, there exists a chord of H from X_0 to X_1 of time-length bounded from above by $\dfrac{|x_0 - x_1|(s_+ - s_-)}{(1 - 2e)\Delta(H; Y_0, Y_1)}$.

In particular, if H is time-independent (and thus, (7) holds for all $0 < e < 1/2$), then the time-length of the chord is $\leq \dfrac{|x_0 - x_1|(s_+ - s_-)}{\Delta(H; Y_0, Y_1)}$.

In the case $n = 1$, the claim holds even without assuming (7).

For the proof, see Sect. 7.

Remark 1.2. In the case of autonomous Hamiltonians, part B of Theorem 1.1 implies that (Y_0, Y_1) μ-interlinks (X_0, X_1), where $\mu = |x_0 - x_1|(s_+ - s_-)$. Part A of Theorem 1.1 does not imply that (Y_0, Y_1) μ-interlinks (X_0, X_1), but is a somewhat weaker claim.

In the proofs of the two parts, we use two slightly different Poisson bracket invariants—the tools used for detecting interlinking in Sect. 2.2. The appearance of the same constant $\mu = |x_0 - x_1|(s_+ - s_-)$ in both parts of Theorem 1.1 is due to the fact that the proofs of both claims are based on the same lower bound on both Poisson bracket invariants, coming from areas of certain pseudo-holomorphic curves used in the theory of Legendrian contact homology. In fact, one can think of μ as the area of a (pseudo-holomorphic) quadrilateral whose edges lie in X_0, X_1, Y_0, Y_1.

Remark 1.3. In the case $n = 1$, the claim of Theorem 1.1 follows rather directly from the results in [28] (cf. [8, Thm. 1.20])—see the proof of Theorem 1.1 in Sect. 7.

In the case $n > 1$, we do not know any way to obtain Theorem 1.1 from the results in [28]. (Basically, the obstacle is the non-existence of a positive Legendrian isotopy between unit cotangent spheres at different points in the unit cotangent bundle of $\mathbb{R}^n$—see [12,13].)

Let us note that the existence of chords as in Theorem 1.1 for similarly defined sets in the cotangent bundle of the torus and compactly supported Hamiltonians can be proved along the lines of the proof of [8, Thm. 1.13] using symplectic quasi-states, with the upper bound on the time-length of the chord depending on the size of the support of the Hamiltonian. The methods of this paper do not allow us to treat this case because the foundations of the Legendrian contact homology have not been worked out rigorously yet for the relevant setting.

Remark 1.4. For (time-independent) mechanical Hamiltonians H, the existence of Hamiltonian chords of H from X_0 to X_1, as in part B of Theorem 1.1, can be obtained by the classical Maupertius's least action principle (see, e.g., [1, p.247]).

Namely, assume that H is a complete mechanical Hamiltonian of the form $H(p, q) = |p|^2/2 + U(q)$, where $0 \leq \sup_{\mathbb{R}^n} |U| < +\infty$. Let $x_0, x_1 \in \mathbb{R}^n$, $0 < s_- < s_+$, and assume that for some $C > \sup_{\mathbb{R}^n} |U|$, the level set $\{H = C\}$ intersects the sets X_0, X_1.

Consider the Riemannian metric $\widetilde{g}$ on $\mathbb{R}^n$ of the form $\widetilde{g} = \sqrt{C - U(q)}g$, where g is the Euclidean metric (the metric $\widetilde{g}$ is called the Jacobi metric). It is not hard to verify that the metric $\widetilde{g}$ is complete, and therefore, by the Hopf-Rinow theorem, there exists a minimal geodesic of $\widetilde{g}$ from x_0 to x_1. By Maupertius's least action principle, the lift of the geodesic to the level set $\{H = C\}$ of H in $T^*\mathbb{R}^n = \mathbb{R}^{2n}$ is a Hamiltonian chord of H from X_0 to X_1.

1.4. Sample applications: contact dynamics

Let us present sample applications to contact dynamics.

1.4.1. Contact interlinking. Let (Σ, ξ) be a contact manifold equipped with a contact form λ. An ordered pair (Λ_0, Λ_1) of disjoint Legendrian submanifolds $\Lambda_0, \Lambda_1 \subset \Sigma$ is called *μ-interlinked* if there exists a constant $\mu = \mu(\Lambda_0, \Lambda_1, \lambda) > 0$, such that every bounded strictly positive contact Hamiltonian h on Σ with $h \geq c > 0$ possesses an orbit of time-length $\leq \mu/c$ starting at Λ_0 and arriving at Λ_1. The pair (Λ_0, Λ_1) is called *interlinked* if it is μ-interlinked for some $\mu > 0$. The pair (Λ_0, Λ_1) is called *robustly interlinked*, if every pair (Λ_0', Λ_1') of Legendrians obtained from (Λ_0, Λ_1) by a sufficiently C^1-small Legendrian isotopy is interlinked.

Write R_t for the Reeb flow of λ. Given two Legendrian submanifolds $\Lambda_0, \Lambda_1 \subset \Sigma$, a Reeb chord $R_t x$, $t \in [0, \tau]$ with $x \in \Lambda_0$ and $y := R_\tau x \in \Lambda_1$ is called *non-degenerate* if

$$D_x R_\tau (T_x \Lambda_0) \oplus T_y \Lambda_1 = \xi_y \ , \tag{8}$$

where ξ_y stands for the contact hyperplane at y.

Let Σ be the jet space $J^1 Q = T^* Q(p, q) \times \mathbb{R}(z)$ of a closed manifold Q equipped with the contact form $dz - pdq$. Let $R = \partial/\partial z$ be the Reeb vector field of λ. Let $\Lambda_0 = \{p = 0, z = 0\}$ be the zero section.

Theorem 1.5. (i) *Let ψ be a positive function on Q, and let $\Lambda_1 := \{z = \psi(q), p = \psi'(q)\}$ be the graph of its 1-jet. Then, the pair (Λ_0, Λ_1) is robustly interlinked.*

(ii) *Assume that $\Lambda_1 \subset \Sigma = J^1 Q$ is a Legendrian submanifold Legendrian isotopic to Λ_0, with the following property: there is a unique chord of the Reeb flow R_t starting on Λ_0 and ending on Λ_1, and this chord is non-degenerate. Then, the pair (Λ_0, Λ_1) is interlinked.*

The proof is given in Sect. 5.

Remark 1.6. The assumption in part (ii) of Theorem 1.5 that Λ_1 is Legendrian isotopic to Λ_0 can be considerably weakened. The actual assumption that we need for the proof is that the Legendrian contact homology of Λ_1 is not zero—see Sect. 6.

1.4.2. Beyond interlinking. Let ψ be a positive function on Q, and let $\Lambda := \{z = \psi(q), p = \psi'(q)\}$ be the graph of its 1-jet in $T^*Q \times \mathbb{R}$. Let us note that for every critical point q of ψ, there is a Reeb chord of the time-length $\psi(q)$ from the zero section Λ_0 to Λ. As we have seen in Theorem 1.5(i), the pair (Λ_0, Λ) is interlinked. Note that the order matters: the pair (Λ, Λ_0) is not interlinked—indeed, there is no Reeb chord from Λ to Λ_0. Assume now that K is Legendrian isotopic to Λ outside Λ_0, and that there exist exactly two Reeb chords A, a starting on K and ending on Λ_0. Assume further that both chords are non-degenerate, and their time-lengths $|A|, |a|$ satisfy

$$0 < |A| - |a| < |b|, \tag{9}$$

for every Reeb chord b starting and ending on $\Lambda_0 \sqcup K$. The next result states that for contact Hamiltonians with a *sufficiently small oscillation* (i.e., for small perturbations of the *constant* contact Hamiltonian 1), one can establish existence of a chord even in the absence of interlinking.

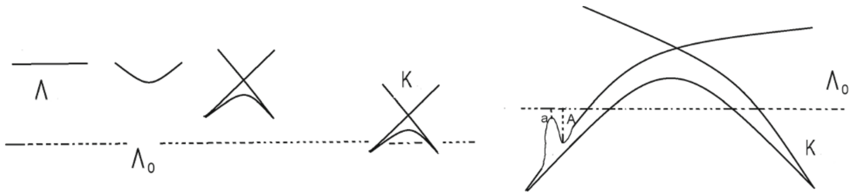


FIGURE 1. Isotopy from Λ to K (left); K zoomed in (right)

Theorem 1.7. *Let h be any positive bounded contact Hamiltonian on Σ with $c := \inf_\Sigma h \le h \le \sup_\Sigma h =: C$. If*

$$\frac{C}{c} < \frac{|A|}{|a|}\,, \tag{10}$$

then there is a chord of h starting on K and ending on Λ_0 of time-length $\le |a|(|A| - |a|)/(|A|c - |a|C)$.

The proof is given in Sect. 6. For an example of a Legendrian submanifold $K \subset T^*S^1 \times \mathbb{R}$ satisfying the assumption of Theorem 1.7, we refer to Fig. 1 describing the front projection of K. As we shall see later on, the proof of this result involves the machinery of persistence modules.

1.4.3. Chords of contact Hamiltonians. Next, we relax the setting of contact interlinking and work with contact Hamiltonians which may be time-dependent, unbounded, and may change sign.

Let $l \in \mathbb{R}$, $l > 0$. Let Λ_0, Λ_1 be the following Legendrian submanifolds of (Σ, ξ): if $\Sigma = ST^*\mathbb{R}^n$, then Λ_0 is the unit cotangent sphere at some $x_0 \in \mathbb{R}^n$ while

(i) either Λ_1 is the image of Λ_0 under the time-l Reeb flow;
(ii) or Λ_1 is the unit cotangent sphere at some $x_1 \in \mathbb{R}^n$, $|x_0 - x_1| = l$, in which case there exists a unique non-degenerate Reeb chord starting at Λ_0 and ending on Λ_1.

Note that these pairs (Λ_0, Λ_1) are interlinked by Theorem 1.5(i) and (ii), respectively. We shall also allow $\Sigma = J^1Q$, where Λ_0 is the zero section and Λ_1 is its image under the time-l Reeb flow, as in Theorem 1.5(i). We shall discuss applications concerning the existence of chords of contact Hamiltonians from Λ_0 to Λ_1.

Assume $h : \Sigma \times \mathbb{S}^1 \to \mathbb{R}$ is a complete (time-periodic) contact Hamiltonian (with respect to λ). Write $h_t := h(\cdot, t)$, $t \in \mathbb{S}^1$. Denote by v_t, $t \in \mathbb{S}^1$, the (time-periodic) contact Hamiltonian vector field of h with respect to the contact form λ. If v_t is time-independent, we write just v. Let $\{\varphi_t\}$ be the flow of v_t—that is, the contact Hamiltonian flow of h.

Definition 1.8. Let us say that $h : \Sigma \times \mathbb{S}^1 \to \mathbb{R}$ is *C-cooperative with* Λ_0, Λ_1 for $C > 0$ if either of the following conditions holds:

(a) $h < C$ on $\Lambda_1 \times \mathbb{S}^1$ and $dh_t(R) \ge 0$ on $\{h_t \ge C\}$ for all $t \in \mathbb{S}^1$.
(b) $h < C$ on $\Lambda_0 \times \mathbb{S}^1$ and $dh_t(R) \le 0$ on $\{h_t \ge C\}$ for all $t \in \mathbb{S}^1$.

We will say that h is *cooperative with* Λ_0, Λ_1 if it is C-cooperative with Λ_0, Λ_1 for some $C > 0$.

Note that conditions (a) and (b) hold, in particular, for a sufficiently large C if $\sup_{\Sigma\times\mathbb{S}^1} h < +\infty$. Conditions (a) and (b) guarantee that a chord of h from Λ_0 to Λ_1, if it exists, does not leave the set $\{h \leq C\}$ at any time.

Theorem 1.9 (cf. Rem. 1.14 in [28]). *Assume that h is cooperative with Λ_0, Λ_1 and that $\inf_{\Sigma\times\mathbb{S}^1} h > 0$. Assume also that for some $0 < e < 1/2$*

$$\sup_{\Sigma\times\mathbb{S}^1} |\partial h/\partial t| < \frac{(1-2e)e\big(\inf_{\Sigma\times\mathbb{S}^1} h\big)^3}{\big(\max_{\Lambda_0\times\mathbb{S}^1} h + e\inf_{\Sigma\times\mathbb{S}^1} h\big)l}.$$

Then, there exists a chord of h from Λ_0 to Λ_1 of time-length bounded from above by $\dfrac{l}{(1-2e)\inf_{\Sigma\times\mathbb{S}^1} h}$.

In particular, if h is time-independent, then the time-length can be bounded from above by $\dfrac{l}{\inf_{\Sigma\times\mathbb{S}^1} h}$.

For the proof of Theorem 1.9, see Sect. 7.

For other results on the existence of Reeb chords between different Legendrian submanifolds (or equivalently, chords of positive contact Hamiltonians), see the papers of Dimitroglou Rizell and Sullivan [17,18] (for a comparison of their results with the results in [28] and here, see [18, Sec. 1.3]).

Theorem 1.9, together with a basic dynamical assumption, allows to obtain the following results concerning the chords of contact Hamiltonians *that are not everywhere positive*.

Corollary 1.10. *Assume that h is cooperative with Λ_0, Λ_1 and there exists a (possibly non-compact or disconnected) closed codimension-0 submanifold $\Xi \subset \Sigma$ with a (possibly non-compact or disconnected) boundary $\partial\Xi$, so that*

1. $\inf_{\Xi\times\mathbb{S}^1} h > 0$ *(but h may be negative outside $\Xi\times\mathbb{S}^1$).*
2. $\sup_{\partial\Xi\times\mathbb{S}^1} h < +\infty$.
3. *For each $t \in \mathbb{S}^1$, the vector field v_t is transverse to $\partial\Xi$ (in particular, $\partial\Xi$ is a convex surface in the sense of contact topology—see [37]) and either points inside Ξ everywhere on $\partial\Xi$ or points outside Ξ everywhere on $\partial\Xi$.*
4. *Both Λ_0 and Λ_1 lie in Ξ.*

Assume also that for some $0 < e < 1/2$

$$\sup_{\Xi\times\mathbb{S}^1} |\partial h/\partial t| < \frac{(1-2e)e\big(\inf_{\Xi\times\mathbb{S}^1} h\big)^3}{\big(\max_{\Lambda_0\times\mathbb{S}^1} h + e\inf_{\Xi\times\mathbb{S}^1} h\big)l}.$$

Then, there exists a chord of h from Λ_0 to Λ_1 whose time-length is bounded from above by $\dfrac{l}{(1-2e)\inf_{\Xi\times\mathbb{S}^1} h}$.

If h is time-independent, then the time-length of the chord is bounded from above by $l/\inf_{\Xi} h$.

For the proof of Corollary 1.10, see Sect. 7.

Remark 1.11. Assume h is time-independent and $\Xi := \{h \geq c\}$ for some $c > 0$. Then, the conditions (1) and (2) are satisfied automatically, while condition (3) is equivalent to $\partial\Xi$ being transverse to the Reeb vector field R, because $dh(v) = hdh(R)$.

Example 1.12. An example satisfying the assumptions of Corollary 1.10 can be constructed as follows. Write R_t for the Reeb flow on J^1Q. Let Λ_0 and Λ_1 be the images of the zero section under R_c and R_{c+l}, respectively, with $c, l > 0$. Note that $\Lambda_0, \Lambda_1 \subset \Xi := \{z \geq c\}$. Thus, condition (4) in Corollary 1.10 is satisfied.

Consider a time-independent contact Hamiltonian $h = az + g$ on $\Sigma = J^1Q = T^*Q \times \mathbb{R}$, where $a > 0$, z is the coordinate along the $\mathbb{R}$-factor and g is a smooth bounded function on T^*Q. Assume that $\inf_\Xi h > 0$—this can be achieved if a is sufficiently large compared to $||g||_{L^\infty}$. Then, condition (1) in Corollary 1.10 is satisfied. Condition (2) is satisfied, since g is bounded. Finally, condition (3) is satisfied, since the Reeb vector field R of the standard contact form on J^1Q is $\partial/\partial z$ and therefore $dh(v) = hdh(R) = ah > 0$ on $\partial\Xi = \{z = c\}$. It is also easy to verify that h is C-cooperative with Λ_0, Λ_1 for a sufficiently large C.

We have verified that the objects above—and accordingly their preimages under ψ—satisfy the assumptions of Corollary 1.10. Consequently, Corollary 1.10 yields the existence of a chord of h from Λ_0 to Λ_1 of time-length $\leq l/\inf_\Xi h$.

1.4.4. Contact flows with large conformal factor. Our next result illustrates that contact Hamiltonians separating (in a suitable sense) certain pairs of Legendrian submanifolds generate contact flows with an arbitrarily large conformal factor. Here, Λ_0 and Λ_1 are as in the beginning of Sect. 1.4.3.

Theorem 1.13. *Assume that h is time-independent, compactly supported, and*

$$h|_{\Lambda_0}\geq 0,\;\; h|_{\Lambda_1}<0.$$

Then, the conformal factor of φ_t takes arbitrarily large values as t varies between 0 and $+\infty$:

$$\inf_{t\in(0,+\infty),y\in\Sigma} \frac{\left(\varphi_t^{-1}\right)^* \lambda\left(\varphi_t\left(y\right)\right)}{\lambda\left(\varphi_t\left(y\right)\right)} = +\infty.$$

Recall that the conformal factor is an important dynamical characteristic playing the role of the contact Lyapunov exponent. For the proof, see Sect. 7.

1.5. Scheme of the proof: homologically bonded pairs

Let us outline a key property of Σ, λ, Λ_0, Λ_1 above that allows to prove the results in Sects. 1.3, 1.4 and outline the general scheme of the proofs.

Let $\Lambda_0, \Lambda_1 \subset \Sigma$ be disjoint Legendrian (not necessarily connected) compact submanifolds without boundary of a nice contact manifold (see Sect. 1.2). Let $\Lambda := \Lambda_0 \sqcup \Lambda_1$. Assume that the pair (Λ, λ) is *non-degenerate*—that is, there are only finitely many Reeb chords of Λ and they are all non-degenerate (this can be always achieved by a C^∞-small Legendrian perturbation of either of the two Legendrian submanifolds). Then, one can associate with the pair $(\Lambda = \Lambda_0 \sqcup \Lambda_1, \lambda)$ its (full) Chekanov–Eliashberg algebra—a free non-commutative unital algebra over $\mathbb{Z}_2$ generated by all the Reeb chords of Λ. The algebra is filtered by the action (the action of a Reeb chord is its time-length; the action of a monomial, or a product of the Reeb chords, is the sum of the actions of its factors).

We consider a vector subspace of the Chekanov–Eliashberg algebra, which we will call the 01-*subspace*—it is generated by the monomials $a_1 \cdot \ldots \cdot a_k$, $k \in \mathbb{Z}_{>0}$, where a_1 starts at Λ_0, a_k ends at Λ_1, and for each $m = 1, \ldots, k-1$, the end of a_m lies in the same component of Λ as the origin of a_{m+1}.

Recall that the differential ∂_J on the Chekanov–Eliashberg algebra depends on an almost complex structure J on the symplectization of $(\Sigma, \ker \lambda)$ and is defined as follows: the differential of a generator (that is, a Reeb chord) is defined using the count of J-holomorphic disks in the symplectization with one positive and possibly several negative punctures on the boundary, whose boundary lies in Λ and that converge near the punctures to cylinders over Reeb chords of Λ; the differential is then extended to the whole algebra using the Leibniz rule and the condition $\partial_J(1) := 0$ (see Sect. 4.2).

The differential preserves the 01-subspace and lowers the filtration. This allows to view the resulting homology of the 01-subspace—*the filtered Legendrian contact homology of* $(\Lambda := \Lambda_0 \sqcup \Lambda_1, \lambda)$—as a persistence module defined over $(-\infty, +\infty)$ and apply the theory of persistence modules to its study. In particular, one can associate with it its barcode—a collection of intervals, called *bars*, lying in $(0, +\infty)$.

For $s \in (1, +\infty)$, let $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda)$ denote the smallest left end of a bar of multiplicative length greater than s in the barcode. (The multiplicative length of a bar in $(0, +\infty)$ is the ratio of its right and left ends; note that it may be infinite.) Let $l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda)$ denote the smallest left end of an infinite bar. If there are no such bars, set $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda) := +\infty$ or, respectively, $l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) := +\infty$.

If the pair $(\Lambda = \Lambda_0 \sqcup \Lambda_1, \lambda)$ is degenerate, then $\Lambda = \Lambda_0 \sqcup \Lambda_1$ can be approximated by Legendrian submanifolds $\Lambda' = \Lambda'_0 \sqcup \Lambda'_1$ obtained from Λ by a C^∞-small Legendrian isotopy, so that the pair (Λ', λ) is non-degenerate. Extend the definition of $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda)$, $s \in (1, +\infty]$, to all pairs $(\Lambda = \Lambda_0 \sqcup \Lambda_1, \lambda)$ as follows:

$$l_{\min,s}(\Lambda_0, \Lambda_1, \lambda) := \liminf l_{\min,s}(\Lambda'_0, \Lambda'_1, \lambda),$$

where the lim inf is taken over all such $\Lambda' = \Lambda'_0 \sqcup \Lambda'_1$ converging to Λ (in the C^∞-topology). One can show that for non-degenerate pairs $(\Lambda = \Lambda_0 \sqcup \Lambda_1, \lambda)$, this definition and the original one yield the same $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda)$—cf. Remark 4.8.

If $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda) < +\infty$ for all $s \in (1, +\infty)$, we will say that the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is *weakly homologically bonded*. If $l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) < +\infty$, we will say that the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is homologically bonded. If the pair is non-degenerate, these conditions mean that the corresponding barcode contains bars of arbitrarily large multiplicative length, or, respectively, an infinite bar.

The key property of the setting in Sects. 1.3, 1.4 is that the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is *homologically bonded*.

Consider the *stabilization of* Σ which is the manifold $\widehat{\Sigma} := \Sigma \times \mathbb{R}(r) \times \mathbb{S}^1(\tau)$ equipped with the contact form $\widehat{\lambda} := \lambda + r d\tau$. This contact manifold is also nice. For a Legendrian submanifold $\Delta \subset \Sigma$, define a Legendrian submanifold $\widehat{\Delta} \subset \widehat{\Sigma}$ by

$$\widehat{\Delta} := \Delta \times \{r = 0\}.$$

For each $s \in (1, +\infty]$, define

$$\widehat{l}_{\min,s}(\Lambda_0, \Lambda_1, \lambda) := l_{\min,s}(\widehat{\Lambda}_0, \widehat{\Lambda}_1, \widehat{\lambda}).$$

The pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is said to be *stably homologically bonded* if $\widehat{l}_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) < +\infty$.

Remark 1.14. It is likely that homological bondedness implies stable homological bondedness. If Σ is the standard contact $\mathbb{R}^3$, this follows from a result of Ekholm-Kálmán [24, Thm. 1.1], but, to the best of our knowledge, the case of a general (nice) contact manifold has not been worked out so far.

For the Legendrian submanifolds $\Lambda = \Lambda_0 \sqcup \Lambda_1$ in Sect. 1.4, the pair $(\Lambda = \Lambda_0 \sqcup \Lambda_1, \lambda)$ is homologically bonded and stably homologically bonded—see Sects. 6, 7.

Let us now explain how (stable) homological bondedness is used to prove the results in Sects. 1.3, 1.4.

For a pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ in a general nice contact manifold (Σ, λ), consider the trivial exact Lagrangian cobordism $(\Lambda_0 \sqcup \Lambda_1) \times [s_-, s_+]$ in the trivial exact symplectic cobordism $(\Sigma \times [s_-, s_+], d(s\lambda))$. For instance, in the setting of Theorem 1.1, the latter exact symplectic cobordism can be identified with the manifold $\{(p, q) \in \mathbb{R}^{2n} \mid s_0 \leq |p| \leq s_1\}$ whose boundary components—the sets Y_0, Y_1—are identified, respectively, with $\Sigma \times s_-$ and $\Sigma \times s_-$. The parts $\Lambda_0 \times [s_-, s_+]$, $\Lambda_1 \times [s_-, s_+]$ of the trivial exact Lagrangian cobordism $(\Lambda_0 \sqcup \Lambda_1) \times [s_-, s_+]$ are then identified, respectively, with the sets X_0, X_1.

If the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is non-degenerate, the exact Lagrangian cobordism defines a cobordism map (in the category of the persistence modules) from the persistence module associated to $(\Lambda = \Lambda_0 \sqcup \Lambda_1, s_+\lambda)$ to the one associated with $(\Lambda = \Lambda_0 \sqcup \Lambda_1, s_-\lambda)$. These persistence modules are multiplicative shifts of each other and the cobordism map is the multiplicative shift by s_+/s_-.

If the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is weakly homologically bonded (and, in particular, if it is homologically bonded), then the cobordism map is not the zero morphism between persistence modules and the pseudo-holomorphic curves used to define the map can be also used to prove that a version of the Poisson bracket invariant of quadruples of sets is positive for the following quadruple of sets: $\Lambda_0 \times [s_-, s_+]$, $\Lambda_1 \times [s_-, s_+]$, $\Sigma \times s_-$, $\Sigma \times s_+$. This is the key result in the paper—see Sect. 2 for the precise definition of the Poisson bracket invariant (it is a version of the invariant defined previously in [8,28]) and Theorem 4.9 for the precise statement of the result. If the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is degenerate but still weakly homologically bonded, the same result is obtained using a semi-continuity property of the Poisson bracket invariant.

The existence of chords for *time-independent* symplectic Hamiltonians as in Theorem 1.1 follows then from the positivity of the Poisson bracket invariant by Fathi's dynamical Urysohn lemma (see Theorem 2.1 for its statement).

If the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is stably homologically bonded, then a similar argument yields the existence of chords as in part B of Theorem 1.1 for *time-dependent* Hamiltonians.

The existence of a chord from Λ_0 to Λ_1 for a contact Hamiltonian cooperative with Λ_0, Λ_1, as in Sect. 1.4.3, is deduced then from the existence of the chords of the corresponding homogeneous symplectic Hamiltonian on the symplectization of Σ.

Remark 1.15. The claims of Theorem 1.1 (and its analogues for other homologically bonded pairs) on the existence of Hamiltonian chords remain true if X_0, X_1 are perturbed as exact Lagrangian cobordisms (cylindrical near the boundaries) in the trivial exact symplectic cobordism $\{(p, q) \in \mathbb{R}^{2n} \mid s_0 \leq |p| \leq s_1\}$, so that the Legendrian isotopies, induced on the boundaries by the exact Lagrangian isotopies, are sufficiently small—e.g., sufficiently C^∞-small. (Note that away from the boundary, the perturbations may be arbitrarily. large, as long as the perturbed X_0, X_1 are disjoint!). The upper bound on the time-length of the Hamiltonian chords between the perturbed X_0, X_1 is then only slightly larger than the one for the original X_0, X_1.

Similarly, the claims of the results in Sect. 1.4 remain true if the Legendrian submanifolds Λ_0, Λ_1 are perturbed by sufficiently C^∞-small Legendrian isotopies into Legendrian submanifolds Λ_0', Λ_1'. The upper bound on the time-length of the chord between Λ_0', Λ_1' is then only slightly larger than the one for the original Λ_0, Λ_1 and tends to it as the sizes of the Legendrian isotopies tend to zero.

Let us also remark that the scheme of the proof can be extended to a more general setting and, in particular, to non-trivial exact Lagrangian cobordisms; one can also use the linearized Legendrian contact homology instead of the full one [30].

For more details and a reference to the proofs, see more general Remarks 4.8, 4.13, 5.8.

1.6. Plan of the paper

Let us outline the plan of the paper.

In Sect. 2, we define a Poisson bracket invariant of a quadruple of sets (a modified version of the invariant defined previously in [8,28]) and state a recent theorem of Fathi (a generalization of the result in [31]), which allows to deduce the existence of a Hamiltonian chord from the positivity of the invariant.

In Sect. 3, we recall basic facts about persistence modules.

In Sect. 4, we describe the Legendrian contact homology setting that we need and explain how to associate a persistence module, and a corresponding barcode, to a (non-degenerate) pair formed by a Legendrian submanifold and a contact form. Then, we show how the existence of bars of sufficiently large multiplicative length in the barcode implies the positivity of the Poisson bracket invariant for appropriate quadruples of sets, which in turn yields the existence of the wanted chords.

In Sect. 5, we discuss applications of the result proved in Sect. 4 to contact dynamics.

In Sects. 6, 7, we explain how the results of Sect. 4 can be applied to the Legendrian submanifolds in J^1Q and $ST^*\mathbb{R}^n$, which yields the results of Sects. 1.3, 1.4.

2. A modified pb^+-invariant and Hamiltonian chords

In this section, we discuss a Poisson bracket invariant of quadruples of sets in a symplectic manifold. The proof of the result relating the Poisson bracket invariant to the existence of Hamiltonian chords is based on the theorems of Fathi in general smooth/topological dynamics [31,32] and is similar to the proof of the relevant results in [28] (with the only difference that we can now use the results from [32] that were unavailable when [28] was written). Let us recall these results of Fathi.

2.1. Chords of smooth vector fields

In this section, let M be any smooth manifold (without boundary), v a complete smooth time-independent vector field on M (meaning that its flow is defined for all times) and $X_0, X_1 \subset M$ disjoint closed subsets of M. Denote by $T(X_0, X_1; v)$ the infimum of the time-lengths of the chords (that is, integral trajectories) of v from X_0 to X_1. If there is no such chord, set $T(X_0, X_1; v) := +\infty$.

The following theorem (a "dynamical Urysohn lemma") was proved by Fathi in [31] in the case when X_0, X_1 are compact and in [32] in the case when X_0, X_1 are arbitrary closed sets.

Theorem 2.1 (Fathi, [31,32]). *Assume $T > T(X_0, X_1; v)$.*

Then, there exists a smooth function $F : M \to \mathbb{R}$, such that $F|_{X_0} \leq 0$, $F|_{X_1} > 1$, and $L_v F < 1/T$.

If X_0, X_1 are compact, then F can be chosen to be compactly supported.

Define

$$\mathcal{S}(X_0, X_1) := \{F \in C^\infty(M) \mid F|_{X_0} \leq 0,\ F|_{X_1} \geq 1\}, \tag{11}$$

$$\begin{aligned}&\mathcal{S}'(X_0, X_1)\\ &\quad := \{F \in C^\infty(M) \mid \mathrm{Im}\,(F) \subset [0,1],\ F|_{Op(X_0)} = 0,\ F|_{Op(X_1)} = 1\},\end{aligned} \tag{12}$$

where $Op(\cdot)$ denotes some open neighborhood of a set. Define

$$\mathcal{K}(X_0, X_1) := \mathcal{S}(X_0, X_1) \cap C_c^\infty(M),\ \mathcal{K}'(X_0, X_1) := \mathcal{S}'(X_0, X_1) \cap C_c^\infty(M),$$

where $C_c^\infty(M)$ is the space of compactly supported smooth functions on M.

Define

$$L(X_0, X_1; v) := \inf_{F \in \mathcal{S}(X_0, X_1)} \sup_M L_v F,$$

and

$$L_c(X_0, X_1; v) := \inf_{F \in \mathcal{K}(X_0,X_1)} \sup_M L_v F.$$

If X_1 is non-compact, the set $K(X_0, X_1)$ is empty. Our convention is that $\inf \emptyset = +\infty$. Clearly

$$L(X_0, X_1; v) \leq L_c(X_0, X_1; v).$$

Proposition 2.2. *In the definition of $L(X_0, X_1; v)$ and $L_c(X_0, X_1; v)$, one can replace $\mathcal{S}(X_0, X_1)$ and $\mathcal{K}(X_0, X_1)$, respectively, by $\mathcal{S}'(X_0, X_1)$ and $\mathcal{K}'(X_0, X_1)$*

$$L(X_0, X_1; v) = \inf_{F \in \mathcal{S}'(X_0,X_1)} \sup_M L_v F,$$

and if X_0 and X_1 are compact

$$L_c(X_0, X_1; v) = \inf_{F \in \mathcal{K}'(X_0,X_1)} \sup_M L_v F.$$

Proof of Proposition 2.2. Let us prove the claim for $L(X_0, X_1; v)$—the case of $L_c(X_0, X_1; v)$ is similar.

Clearly, $L(X_0, X_1; v) \leq \inf_{F \in \mathcal{S}'(X_0,X_1)} \sup_M L_v F$, and thus, it suffices to prove that

$$\inf_{F \in \mathcal{S}'(X_0,X_1)} \sup_M L_v F \leq L(X_0, X_1; v). \tag{13}$$

Let $\delta > 0$. Pick a non-decreasing smooth function $\chi : \mathbb{R} \to \mathbb{R}$ so that $\sup_{t \in \mathbb{R}} \chi'(t) \leq 1 + \delta$, and for some $\epsilon > 0$, we have $\chi(t) = 0$ on $(-\infty, \epsilon]$ and $\chi(t) = 1$ on $[1 - \epsilon, +\infty)$.

Then, $\chi \circ F \in \mathcal{S}'(X_0, X_1)$ for any $F \in \mathcal{S}(X_0, X_1)$ and

$$L_v(\chi \circ F) = (\chi' \circ F) L_v F \leq (1 + \delta) L_v F.$$

Taking the infimum over $F \in \mathcal{S}(X_0, X_1)$ in both sides, we get

$$\inf_{F \in \mathcal{S}'(X_0,X_1)} \sup_M L_v F \leq (1 + \delta) L(X_0, X_1; v).$$

Since this is true for any $\delta > 0$, we obtain (13) and this finishes the proof of the proposition. $\square$

The following corollary follows readily from Theorem 2.1.

Corollary 2.3. *$T(X_0, X_1; v) = 1/L(X_0, X_1; v)$.*

Consequently, if $L(X_0, X_1; v) > 0$, then for any $\epsilon > 0$, there exists a chord of v from X_0 to X_1 of time-length $\leq 1/L(X_0, X_1; v) + \epsilon$ (if $\operatorname{supp} v$ is compact, then one can drop ϵ from the bound).

If X_0, X_1 are compact, then $L(X_0, X_1; v) = L_c(X_0, X_1; v)$ and $T(X_0, X_1; v) = 1/L_c(X_0, X_1; v)$.

Consequently, if X_0 and X_1 are compact and $L_c(X_0, X_1; v) > 0$, then there exists a chord of v from X_0 to X_1 of time-length $\leq 1/L_c(X_0, X_1; v)$.

2.2. Poisson bracket invariant

Let (M, ω) be a (not necessarily compact) connected symplectic manifold, possibly with boundary.

We use the following sign conventions in the definitions of a Hamiltonian vector field and the Poisson bracket on M: the Hamiltonian vector field *sgrad* H of a Hamiltonian H is defined by

$$i_{sgrad\, H}\omega = -dH,$$

and the Poisson bracket of two Hamiltonians F, G is given by

$$\begin{aligned}\{F, G\} &:= \omega(sgrad\, G, sgrad\, F) = dF(sgrad\, G) = -dG(sgrad\, F) = \\ &= L_{sgrad\, G}F = -L_{sgrad\, F}G.\end{aligned}$$

Assume X_0, X_1, Y_0, Y_1 are closed subsets of M, such that

$$X_0 \cap X_1 = Y_0 \cap Y_1 = \emptyset.$$

Such a collection of sets X_0, X_1, Y_0, Y_1 will be called *an admissible quadruple.*

Consider the following conditions on pairs $(F, G) \in C^\infty(M) \times C^\infty(M)$:

(1) $\sup_M \{F, G\} < +\infty$ and the vector field $F\, sgrad\, G$ on M is complete.
(1′) *supp* F is compact and *supp* $(F\, sgrad\, G)$ lies in the interior of M.
(2) $F|_{X_0} \leq 0,\ F|_{X_1} \geq 1,\ G|_{Y_0} \leq 0,\ G|_{Y_1} \geq 1$.
(2′) $F|_{X_0} \leq 0,\ F|_{X_1} \geq 1,\ G|_{Op(Y_0)} \equiv 0,\ G|_{Op(Y_1)} \equiv 1$. (Here, $Op(\cdot)$ denotes some open neighborhood of a set.)
(2″) $F|_{Op(X_0)} \equiv 0,\ F|_{Op(X_1)} \equiv 1,\ G|_{Op(Y_0)} \equiv 0,\ G|_{Op(Y_1)} \equiv 1$.
(3) $\mathrm{Im}\, F \subset [0, 1]$.

Note that (1′) ⇒ (1), since $supp\, \{F, G\} \subset supp\, (F dG) = supp\, (F\, sgrad\, G)$.

Define

$$\mathcal{F}_M(X_0, X_1, Y_0, Y_1) := \{(F, G) \in C^\infty(M) \times C^\infty(M) \mid (F, G) \text{ satisfies (1) and (2)}\},$$
$$\mathcal{F}'_M(X_0, X_1, Y_0, Y_1) := \{(F, G) \in C^\infty(M) \times C^\infty(M) | (F, G) \text{ satisfies (1),(2) and (3)}\},$$
$$\mathcal{F}''_M(X_0, X_1, Y_0, Y_1) := \{(F, G) \in C^\infty(M) \times C^\infty(M) | (F, G) \text{ satisfies (1),(2”) and (3)}\},$$
$$\mathcal{G}_M(X_0, X_1, Y_0, Y_1) := \{(F, G) \in C^\infty(M) \times C^\infty(M) | (F, G) \text{ satisfies } (1') \text{ and } (2)\},$$
$$\mathcal{G}'_M(X_0, X_1, Y_0, Y_1) := \{(F, G) \in C^\infty(M) \times C^\infty(M) \mid (F, G) \text{ satisfies } (1'), (2')\},$$
$$\mathcal{G}''_M(X_0, X_1, Y_0, Y_1) := \{(F, G) \in C^\infty(M) \times C^\infty(M) | (F, G) \text{ satisfies } (1'), (2'') \text{ and } (3)\}.$$

For brevity, we will omit the sets X_0, X_1, Y_0, Y_1 from this notation, when needed.

Clearly

$$\begin{aligned}&\mathcal{G}''_M \subset \mathcal{G}'_M \subset \mathcal{G}_M \subset \mathcal{F}_M,\ \mathcal{F}''_M \subset \mathcal{F}'_M \subset \mathcal{F}_M,\\ &\mathcal{G}'_M \subset \mathcal{F}'_M,\ \mathcal{G}''_M \subset \mathcal{F}''_M.\end{aligned}$$

It is also easy to see that the sets $\mathcal{G}_M$, $\mathcal{G}'_M$ and $\mathcal{G}''_M$ are non-empty if only if X_1 is compact.

Set

$$pb^+_M(X_0, X_1, Y_0, Y_1) := \inf_{\mathcal{F}_M} \sup_M \{F, G\},$$
$$pb^+_{M,\mathrm{comp}}(X_0, X_1, Y_0, Y_1) := \inf_{\mathcal{G}_M} \max_M \{F, G\}.$$

If the set over which the infimum is taken is empty, we set the infimum to be $+\infty$.

Clearly

$$pb^+_M(X_0, X_1, Y_0, Y_1) \leq pb^+_{M,\mathrm{comp}}(X_0, X_1, Y_0, Y_1).$$

Remark 2.4. The quantities pb^+_M, $pb^+_{M,\mathrm{comp}}$ are versions of the pb_4 invariant of quadruples of sets defined originally in [8] (where the C^0-norm of $\{F, G\}$ was used instead of $\max_M\{F, G\}$) and of the pb^+_4 invariant defined in [28] (where the sets X_0, X_1, Y_0, Y_1 were assumed to be compact, and both F and G were assumed to be compactly supported). See also [35] for a result that allows to define $pb^+_4(X_0, X_1, Y_0, Y_1)$ in terms of the topology of the set $X_0 \cup X_1 \cup Y_0 \cup Y_1$. Note that, unlike pb_4 and pb^+_4, the invariants pb^+_M, $pb^+_{M,\mathrm{comp}}$ are *not* symmetric with respect to the permutation $(X_0, X_1, Y_0, Y_1) \mapsto (Y_0, Y_1, X_1, X_0)$.

Similarly to Proposition 2.2 (also see [8]), one can prove that the sets $\mathcal{F}_M$, $\mathcal{G}_M$ in the definitions of pb^+_M, $pb^+_{M,\mathrm{comp}}$ can be replaced, respectively, by $\mathcal{F}'_M$, $\mathcal{F}''_M$ and by $\mathcal{G}'_M$, $\mathcal{G}''_M$:

$$pb^+_M(X_0, X_1, Y_0, Y_1) = \inf_{\mathcal{F}'_M} \sup_M \{F, G\} = \inf_{\mathcal{F}''_M} \sup_M \{F, G\},$$
$$pb^+_{M,\mathrm{comp}}(X_0, X_1, Y_0, Y_1) = \inf_{\mathcal{G}'_M} \max_M \{F, G\} = \inf_{\mathcal{G}''_M} \max_M \{F, G\}.$$

We will need the following basic properties of pb^+_M, $pb^+_{M,\mathrm{comp}}$.

Monotonicity

Proposition 2.5 (cf. [8,28]). *Assume that M is a codimension-zero submanifold (with boundary) of a symplectic manifold N (without boundary), which is closed as a subset of N. Let $U \subset N$ be an open set. Assume that X_0, X_1, Y_0, Y_1 is an admissible quadruple lying in M, so that $X_0, X_1 \subset U \cap M$, $Y_0 \cap U, Y_1 \cap U \neq \emptyset$ and $\partial M \subset Y_0 \cup Y_1$.*

Then

$$pb^+_N(X_0, X_1, Y_0, Y_1) \geq pb^+_M(X_0, X_1, Y_0, Y_1), \tag{14}$$
$$pb^+_{N,\mathrm{comp}}(X_0, X_1, Y_0, Y_1) \geq pb^+_{M,\mathrm{comp}}(X_0, X_1, Y_0, Y_1), \tag{15}$$
$$pb^+_{M\cap U,\mathrm{comp}}(X_0, X_1, Y_0 \cap U, Y_1 \cap U) \geq pb^+_{M,\mathrm{comp}}(X_0, X_1, Y_0, Y_1). \tag{16}$$

Proof. If $(F, G) \in \mathcal{F}''_N(X_0, X_1, Y_0, Y_1)$, then it follows easily from the definitions that $(F|_M, G|_M) \in \mathcal{F}''_M(X_0, X_1, Y_0, Y_1)$. This yields (14). The inequality (15) follows similarly.

Let us prove (16). Assume $(F, G) \in \mathcal{G}''_{M\cap U}(X_0, X_1, Y_0 \cap U, Y_1 \cap U)$. In particular, this means that *supp* $F \subset M \cap U$ is compact and G is equal to 0 and 1 on some open neighborhoods (in U) of, respectively, $Y_0 \cap U$ and $Y_1 \cap U$. Extend F by zero outside $M \cap U$ to a smooth compactly supported function $\widetilde{F} : M \to [0, 1]$ and extend G to a smooth function $\widetilde{G} : M \to \mathbb{R}$, so that $\widetilde{G}$

is equal to 0 and 1 on some open neighborhoods (in M) of, respectively, Y_0 and Y_1. Then, $(\widetilde{F}, \widetilde{G}) \in \mathcal{G}''_M(X_0, X_1, Y_0, Y_1)$, while $\{\widetilde{F}, \widetilde{G}\} = \{F, G\}$ (because outside U both Poisson brackets vanish, while on U they coincide, since $\widetilde{F}|_U = F$, $\widetilde{G}|_U = G$). This immediately yields (16).

The following property follows from the definitions (cf. [8,28]).

Semi-continuity

Suppose that (X_0, X_1, Y_0, Y_1) is an admissible quadruple in (M, ω), X_0, X_1 are compact, and $\{X_0^{(j)}\}, \{X_1^{(j)}\}$, $j \in \mathbb{N}$, are sequences of compact subsets of M converging (in the sense of the Hausdorff distance between sets), respectively, to X_0, X_1, so that the quadruples $(X_0^{(j)}, X_1^{(j)}, Y_0, Y_1)$ are admissible for all $j \in \mathbb{N}$.

Then

$$\limsup_{j\to+\infty} pb_M^+(X_0^{(j)}, X_1^{(j)}, Y_0, Y_1) \le pb_M^+(X_0, X_1, Y_0, Y_1), \tag{17}$$

$$\limsup_{j\to+\infty} pb_{M,\mathrm{comp}}^+(X_0^{(j)}, X_1^{(j)}, Y_0, Y_1) \le pb_{M,\mathrm{comp}}^+(X_0, X_1, Y_0, Y_1). \tag{18}$$

The next proposition is proved as in [28] using Corollary 2.3.

Proposition 2.6. *Assume that M is a codimension-zero submanifold (with boundary) of a symplectic manifold N (without boundary), so that M is closed as a subset of N. Assume that X_0, X_1, Y_0, Y_1 is an admissible quadruple lying in M, so that $\partial M \subset Y_0 \cup Y_1$.*

Let $H : N \to \mathbb{R}$ be a complete time-independent Hamiltonian. Then, the following claims hold:

I. *Assume $\Delta(H; Y_0, Y_1) =: \Delta > 0$. If $pb_M^+(X_0, X_1, Y_0, Y_1) =: p > 0$, then for any $\epsilon > 0$, there exists a chord of H from X_0 to X_1 of time-length $\le \dfrac{1}{p\Delta} + \epsilon$. If X_0, X_1 are compact and $pb_{M,\mathrm{comp}}^+(X_0, X_1, Y_0, Y_1) =: p_{\mathrm{comp}} > 0$, then there exists a chord of H from X_0 to X_1 of time-length $\le \dfrac{1}{p_{\mathrm{comp}}\Delta}$.*

II. *Assume that X_0, X_1 are compact, supp $H \cap M$ is compact, $H|_{X_0} \ge 0$ and $H|_{X_1} \le -\Delta$ for some $\Delta > 0$. Assume also that $pb_{M,\mathrm{comp}}^+(X_0, X_1, Y_0, Y_1) =: p_{\mathrm{comp}} > 0$. Then, there exists a chord of H from Y_0 to Y_1 of time-length $\le \dfrac{1}{p_{\mathrm{comp}}\Delta}$.*

Proof of Proposition 2.6. Let us prove part I. We may assume without loss of generality that $H|_{Y_0} \le 0$, $H|_{Y_1} \ge 1$ (this can be always achieved by replacing H with $aH + b$ for some $a, b \in \mathbb{R}$, $a \ne 0$). For any $F \in \mathcal{S}(X_0, X_1)$ (see (11)) satisfying $\sup_N L_{sgrad\, H} F = \sup_N \{F, H\} < +\infty$, we have $(F, H) \in \mathcal{F}'_N$, and if *supp* F is compact, then $(F, H) \in \mathcal{G}_N$. Indeed, since the vector field *sgrad* H is complete, then so is the vector field $F\, sgrad\, H$, since $0 \le F \le 1$. Hence, by (14) and (15)

$$\begin{aligned}\sup_N L_{sgrad\, H} F = \sup_N \{F, H\} &\ge pb_N^+(X_0, X_1, Y_0, Y_1)\\ &\ge pb_M^+(X_0, X_1, Y_0, Y_1),\end{aligned}$$

or, if *supp* F is compact

$$\sup_N L_{sgrad\ H} F = \sup_N \{F, H\} \geq pb^+_{N,\mathrm{comp}}(X_0, X_1, Y_0, Y_1)$$
$$\geq pb^+_{M,\mathrm{comp}}(X_0, X_1, Y_0, Y_1).$$

Taking the infimum over all such F, we get

$$L(X_0, X_1; sgrad\ H) \geq pb^+_M(X_0, X_1, Y_0, Y_1),$$

and if X_0, X_1 are compact

$$L_c(X_0, X_1; sgrad\ H) \geq pb^+_{M,\mathrm{comp}}(X_0, X_1, Y_0, Y_1).$$

Now, the claims of part I follow from Corollary 2.3.

Let us prove part II. We may assume without loss of generality that $H|_{X_0} \geq 0$, $H|_{X_1} \leq -1$ (this can be always achieved by replacing H with H/Δ). For any $G \in \mathcal{S}'(Y_0, Y_1)$ (see (12)), we have $(-H|_M, G|_M) \in \mathcal{G}'_M$. Recall that here *supp* $H \cap M$ is assumed to be compact and G is constant near Y_0 and Y_1, since $G \in \mathcal{S}'(Y_0, Y_1)$. Hence

$$\sup_N L_{sgrad\ H} G \geq \sup_M L_{sgrad\ H} G = \sup_M \{-H, G\} \geq pb^+_{N,\mathrm{comp}}(X_0, X_1, Y_0, Y_1).$$

Taking the infimum over all $G \in \mathcal{S}'(Y_0, Y_1)$ and using Proposition 2.2, we get

$$L(Y_0, Y_1; sgrad\ H) \geq pb^+_{N,\mathrm{comp}}(X_0, X_1, Y_0, Y_1)$$
$$\geq pb^+_{M,\mathrm{comp}}(X_0, X_1, Y_0, Y_1) = p_{\mathrm{comp}}.$$

Now, the claim of part II follows from Corollary 2.3. □

Let us now discuss an implication of Proposition 2.6 for the existence of chords of time-dependent Hamiltonians.

Let $E > 0$. Let $r \in (-E, E)$ and $\tau \in \mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$ be the coordinates, respectively, on $(-E, E)$ and $\mathbb{S}^1$. Set

$$\widetilde{M_E} := M \times (-E, E) \times \mathbb{S}^1$$

and equip $\widetilde{M_E}$ with the product symplectic form $\omega \oplus dr \wedge d\tau$. Let $X_0, X_1, Y_0, Y_1 \subset M$ be an admissible quadruple, such that X_0, X_1 are compact. Set

$$\widetilde{X}_0 := X_0 \times \{r = 0\},\ \widetilde{X}_1 := X_1 \times \{r = 0\},$$
$$\widetilde{Y}_0(E) := Y_0 \times (-E, E) \times \mathbb{S}^1,\ \widetilde{Y}_1(E) := Y_1 \times (-E, E) \times \mathbb{S}^1.$$

Proposition 2.7. *Assume that* $\partial M = \emptyset$. *With an admissible quadruple* $X_0, X_1, Y_0, Y_1 \subset M$ *as above, let* $H : M \times \mathbb{S}^1 \to \mathbb{R}$ *be a complete Hamiltonian and* $\{\phi_t\}_{t\in\mathbb{R}}$ *its flow. Let* $E > 0$. *Assume that*

(a) $pb^+_{\widetilde{M}_E,\mathrm{comp}}(\widetilde{X}_0, \widetilde{X}_1, \widetilde{Y}_0(E), \widetilde{Y}_1(E)) =: \widetilde{p}_E > 0$.
(b) $\Delta(H; Y_0, Y_1) =: \Delta > 2E$.
(c) $\sup_{t_0\in\mathbb{R}} \big(\sup_{t\in[t_0,t_0+T]} H(\phi_t(x), t) - \inf_{t\in[t_0,t_0+T]} H(\phi_t(x), t)\big) < E$ *for any* $x \in X_0$, *where* $T := \dfrac{1}{\widetilde{p}_E(\Delta - 2E)}$.

Then, there exists a chord of H from X_0 to X_1 of time-length $\leq T = \dfrac{1}{\widetilde{p}_E(\Delta - 2E)}$.

Proof of Proposition 2.7. In view of (c), we can pick $0 < E_1 < E_2 < E$, so that

$$\sup_{t\in[t_0,t_0+T]} H(\phi_t(x),t) - \inf_{t\in[t_0,t_0+T]} H(\phi_t(x),t) < E_1 \text{ for all } t_0 \in \mathbb{R}. \tag{19}$$

Pick a smooth cut-off function $\chi : \mathbb{R} \to \mathbb{R}$, such that $\chi(x) = 0$ if $|x| \geq E_2$ and $\chi(x) = 1$ if $|x| \leq E_1$.

Define a *time-independent* Hamiltonian $\widetilde{H} : \widetilde{M}_E \to \mathbb{R}$ as $\widetilde{H}(x, r, \tau) := r + H(x, \tau)$. One easily verifies that the Hamiltonian $\chi\widetilde{H}$ is complete and in view of (b)

$$\Delta\left(\chi H; \widetilde{Y}_0(E), \widetilde{Y}_1(E)\right) \geq \Delta - 2E > 0.$$

Together with (a), this implies, by part I of Proposition 2.6, that there exists a chord γ of $\chi\widetilde{H}$ from $\widetilde{X}_0$ to $\widetilde{X}_1$ of time-length $\leq T = \dfrac{1}{\widetilde{p}_E(\Delta - 2E)}$.

We claim that γ is, in fact, a chord of $\widetilde{H}$ from $\widetilde{X}_0$ to $\widetilde{X}_1$—this would imply that the projection of γ to M is a chord of H from X_0 to X_1 of time-length $\leq T$.

Indeed, note that $\chi\widetilde{H} = \widetilde{H} = H$ on $M \times [-E_1, E_1] \times \mathbb{S}^1 \subset \widetilde{M}_E$ and the projection to M of each trajectory of the Hamiltonian flow of $\widetilde{H}$ on $\widetilde{M}_E$ is a trajectory of the Hamiltonian flow of H on M. Since the time-independent Hamiltonian $\widetilde{H}$ is preserved by its own Hamiltonian flow, (19) implies that for any $t_0 \in \mathbb{R}$, any time-$[t_0, t_0 + T]$ trajectory of the Hamiltonian flow of $\widetilde{H}$ passing at some moment $t \in [t_0, t_0+T]$ through $\widetilde{X}_0$ stays in $M\times[-E_1, E_1]\times\mathbb{S}^1$ for all $t \in [t_0, t_0+T]$. Therefore, for any $t_0 \in \mathbb{R}$ any time-$[t_0, t_0+T]$, trajectory of the Hamiltonian flow of $\chi\widetilde{H}$ passing at some moment $t \in [t_0, t_0+T]$ through $\widetilde{X}_0$ is, in fact, a trajectory of the Hamiltonian flow of $\widetilde{H}$ for all $t \in [t_0, t_0+T]$. In particular, γ is a chord of $\widetilde{H}$ from $\widetilde{X}_0$ to $\widetilde{X}_1$, which proves the claim and finishes the proof of the proposition. □

Remark 2.8. Proposition 2.7 admits an analogue for the case when X_0, X_1 are not necessarily compact—in that case, one should replace the quantity $pb^+_{\widetilde{M}_E,\mathrm{comp}}(\widetilde{X}_0, \widetilde{X}_1, \widetilde{Y}_0(E), \widetilde{Y}_1(E))$ in the claim by $pb^+_{\widetilde{M}_E}(\widetilde{X}_0, \widetilde{X}_1, \widetilde{Y}_0(E), \widetilde{Y}_1(E)) =: pb^+_{\widetilde{M}_E}$, and then, there would exist a chord of H from X_0 to X_1 of time-length $\leq T = \dfrac{1}{pb^+_{\widetilde{M}_E}(\Delta - 2E)}$. The proof of this claim virtually repeats the proof of Proposition 2.7.

3. Persistence modules

In this section, we recall basic facts about persistence modules. For a more detailed introduction to persistence modules, see, e.g., [19], [10], [39], or [42].

We work over the base field $\mathbb{Z}_2$.

Let $\mathbb{I} := (a, +\infty)$, $-\infty \leq a < +\infty$. (In fact, we will be concerned only with $\mathbb{I} = (0, +\infty)$ and $\mathbb{I} = (-\infty, \infty)$.)

Definition 3.1. A *persistence module over* $\mathbb{I}$ is given by a pair

$$(V = \{V_t\}_{t\in\mathbb{I}}, \pi = \{\pi_{s,t}\}_{s,t\in\mathbb{I}, s\leq t}),$$

where all V_t, $t \in \mathbb{I}$, are finite-dimensional $\mathbb{Z}_2$-vector spaces and $\pi_{s,t} : V_s \to V_t$ are linear maps, so that

(i) (Persistence) $\pi_{t,t} = Id$, $\pi_{s,r} = \pi_{t,r} \circ \pi_{s,t}$, for all $s, t, r \in \mathbb{I}$, $s \leq t \leq r$.

(ii) (Discrete spectrum and semi-continuity) There exists a (finite or countable) discrete closed set of points

$$spec(V) = \{l_{\min}(V) := t_0 < t_1 < t_2 < \ldots < +\infty\} \subset \mathbb{I},$$

called *the spectrum of* V, so that
- for any $r \in \mathbb{I} \setminus spec(V)$, there exists a neighborhood U of r in $\mathbb{I}$ such that $\pi_{s,t}$ is an isomorphism for all $s, t \in U$, $s \leq t$;
- for any $r \in spec(V)$, there exists $\epsilon > 0$, such that $\pi_{s,t}$ is an isomorphism of vector spaces for all $s, t \in (r - \epsilon, r] \cap \mathbb{I}$.

(iii) (Semi-bounded support) For the smallest point $l_{\min}(V) := t_0$ of $spec(V)$, one has $l_{\min}(V) > a$ and $V_t = 0$ for all $t \leq l_{\min}(V)$.

In [42], such persistence modules are called *persistence modules of finite type*.

The zero (or trivial) persistence module (over $\mathbb{I}$*)* is a persistence module formed by zero vector spaces and trivial maps between them.

The notions of a persistence submodule, the direct sum of persistence modules, and a morphism/isomorphism between persistence modules (over $\mathbb{I}$) are defined in a straightforward manner.

Recall that two morphisms $\Phi_i : V_i \to W_i$, $i = 1, 2$ between persistence modules are called *right-left equivalent* (or, for brevity, simply equivalent) if there exist isomorphisms $\Psi : V_1 \to V_2$ and $\Theta : W_1 \to W_2$, such that $\Theta\Phi_1 = \Phi_2\Psi$.

Example 3.2. A persistence module V over $(0, +\infty)$ can be extended to a persistence module over $(-\infty, +\infty)$ by setting $V_s = 0$ and $\pi_{s,t} = 0$ for all $s \in (-\infty, 0]$.

Example 3.3. Let $\mathbb{J} \subset \mathbb{I}$ be either of the form $(a_\mathbb{J}, b_\mathbb{J}]$ for some $0 < a_\mathbb{J} < b_\mathbb{J} < +\infty$, or of the form $(a_\mathbb{J}, +\infty)$. Define an *interval (persistence) module*

$$(Q(\mathbb{J}), \pi) := (\{Q(\mathbb{J})\}_{t\in\mathbb{I}}, \{\pi_{s,t}\}_{s,t\in\mathbb{I}, s\leq t}),$$

over $\mathbb{I}$ as follows: $Q(\mathbb{J})_t = \mathbb{Z}_2$ for $t \in \mathbb{J}$ and $Q(\mathbb{J})_t = 0$ for $t \in \mathbb{I} \setminus \mathbb{J}$, while the morphisms $\pi_{s,t}$ are the identity maps for $s, t \in \mathbb{J}$ and zeroes otherwise.

The following structure theorem for persistence modules can be found in [45], [10, Thm. 2.7,2.8], [14]. Its various versions appeared prior to invention of persistence modules; see [4–6, 34, 43, 44].

Theorem 3.4. *For every persistence module* (V, π) *over* $\mathbb{I}$*, there exists a unique (finite or countable) collection of intervals* $\mathbb{J}_j \subset \mathbb{I}$*—where the intervals may not be distinct, but each interval appears in the collection only finitely many times—so that* (V, π) *is isomorphic to* $\oplus_j Q(\mathbb{J}_j)$*:*

$$(V, \pi) = \oplus_j Q(\mathbb{J}_j).$$

The collection $\{\mathbb{J}_j\}$ of the intervals is called the *barcode* of V. The intervals $\mathbb{J}_j$ themselves are called the *bars of* V. Note that the same bar may appear in the barcode several (but finitely many) times (this number of times is called the *multiplicity* of the bar)—in other words, a barcode is a multiset of bars.

The barcode of the trivial persistence module is empty. Set

$$V_\infty := (\mathbb{Z}_2)^k,$$

where $k \in \mathbb{N} \cup \{0, +\infty\}$ is the number of the infinite bars in the barcode of V.

Example 3.5. Let $\mathbb{I} = (-\infty, +\infty)$ and let V be a persistence module over $\mathbb{I}$. Let $c \in \mathbb{R}$.

Define a new persistence module $V^{[+c]}$ over $\mathbb{I}$ by adding c to all indices of V_t and $\pi_{s,t}$—in particular

$$V_t^{[+c]} := V_{t+c}.$$

The barcode of $V^{[+c]}$ is the barcode of V shifted by c to the left.

If $c \geq 0$, one can also define a morphism

$$Sh_V[+c] : V \to V^{[+c]},$$

called *the additive shift of* V *by* c, as follows:

$$Sh_V[+c] := \left\{\pi_{t,t+c} : V_t \to V_t^{[+c]}\right\}_{t\in\mathbb{I}}.$$

Example 3.6. Let $\mathbb{I} = (0, +\infty)$ and let $V = (V, \pi)$ be a persistence module over $\mathbb{I}$. Let $c > 0$. Define a new persistence module $V^{[\times c]}$ over $\mathbb{I}$ by multiplying all the indices of V_t and $\pi_{s,t}$ by c—in particular

$$V_t^{[\times c]} := V_{ct}.$$

The barcode of $V^{[\times c]}$ is the barcode of V divided by c.

If $c \geq 1$, one can also define a morphism

$$Sh_V[\times c] : V \to V^{[\times c]},$$

called *the multiplicative shift of* V *by* c, as follows:

$$Sh_V[\times c] := \left\{\pi_{t,ct} : V_t \to V_t^{[\times c]}\right\}_{t\in\mathbb{I}}.$$

Remark 3.7. One easily sees that additive/multiplicative shifts of isomorphic persistence modules by the same constant are also isomorphic. Thus, one can speak about additive/multiplicative shifts of isomorphism classes of persistence modules.

4. Legendrian contact homology

Let $(P^{2n}, d\vartheta)$, $n \in \mathbb{Z}_{\geq 0}$, be an exact symplectic manifold with bounded geometry at infinity (see [3] for the definition of this class of manifolds; in particular, this class includes symplectic manifolds that are convex in the sense of [25]). Consider the contactization of $(P, d\vartheta)$: Let

$$\Sigma := P \times \mathbb{R}(z), \tag{20}$$

and let $\xi = \ker \lambda$ be the contact structure on Σ defined by the contact form

$$\lambda := dz + \vartheta.$$

Let $\{R_t\}$ be the Reeb flow of λ—each R_t is a shift by t in the coordinate z.

Further on, we will always assume that P, and consequently Σ, are connected.

Consider a compact (not necessarily connected) Legendrian submanifold $\Lambda \subset (\Sigma, \xi)$ without boundary.

If $\Lambda = \Lambda_0 \sqcup \Lambda_1$ is a disjoint union of compact (not necessarily connected) Legendrian submanifolds Λ_0, Λ_1 without boundary, we call Λ *a two-part Legendrian submanifold* and Λ_0, Λ_1 *the* (0- *and* 1-) *parts of* Λ.

Denote the set of all Reeb chords of (Λ, λ) by $\mathcal{R}(\Lambda, \lambda)$.

If $\Lambda = \Lambda_0 \sqcup \Lambda_1$ is a two-part Legendrian submanifold, we say that a Reeb chord of (Λ, λ) is an *ij-chord* for $i, j = 0, 1$ if it starts on Λ_i and ends on Λ_j. Denote the set of all ij-chords of (Λ, λ) by $\mathcal{R}_{ij}(\Lambda, \lambda)$. The ii-chords will be called *pure*, while ij-chords for $i \neq j$ will be called *mixed*.

We say that the pair (Λ, λ) is *non-degenerate* if the following conditions are satisfied:

- For each Reeb chord $a : [0, T] \to \Sigma$, $a(t) = R_t(a(0))$, $a(0), a(T) \in \Lambda$, of Λ with respect to λ, the tangent spaces of the Legendrian submanifolds $R_T(\Lambda)$ and Λ are transversal inside the contact hyperplane at the point $a(T) = R_T(a(0))$.
- Each trajectory of the Reeb flow $\{R_t(x)\}$, $-\infty < t < +\infty$, $x \in \Sigma$, intersects Λ at most in two points. (Equivalently, the images of distinct Reeb chords are disjoint.)

If (Λ, λ) is non-degenerate and Λ is compact, then the set $\mathcal{R}(\Lambda, \lambda)$ is finite.

4.1. Exact Lagrangian cobordisms

The symplectization of (Σ, ξ) can be identified with $\big(\Sigma \times \mathbb{R}_+(s), d(s\lambda)\big)$.

Let $0 < s_- < s_+$.

Consider the manifold with boundary $\Sigma \times [s_-, s_+] \subset \Sigma \times \mathbb{R}_+$ equipped with the symplectic form $\omega := d(s\lambda)$—it is a *trivial exact symplectic cobordism* whose *positive and negative boundaries* and the restrictions of $s\lambda$ to them are identified, respectively, with $(\Sigma, s_+\lambda)$ and $(\Sigma, s_-\lambda)$.

A differential 1-form θ on $\Sigma \times [s_-, s_+]$ will be called a *cobordism 1-form* if the following conditions are satisfied:

- $d\theta = \omega$;
- θ coincides with $s\lambda$ near the boundaries of $\Sigma \times [s_-, s_+]$;

In particular, $s\lambda$ itself is a cobordism 1-form.

Let $\Lambda^\pm \subset (\Sigma, \xi)$ be compact Legendrian submanifolds without boundary, viewed, respectively, as submanifolds of $\Sigma \times s_\pm$. A Lagrangian cobordism L in $(\Sigma \times [s_-, s_+], s\lambda)$ between $\Lambda^\pm$ is a smooth compact cobordism in $\Sigma \times [s_-, s_+]$ between $\Lambda^+ \subset \Sigma \times s_+$ and $\Lambda^- \subset \Sigma \times s_-$ which is a Lagrangian submanifold of $(\Sigma \times \mathbb{R}_+, \omega)$, so that there exist $\delta_\pm > 0$ for which

$$L \cap \Sigma \times [s_+ - \delta_+, s_+] = \Lambda^+ \times [s_+ - \delta_+, s_+],$$
$$L \cap \Sigma \times [s_-, s_- + \delta_-] = \Lambda^- \times [s_-, s_- + \delta_-].$$

The sets $L \cap \Sigma \times [s_+ - \delta_+, s_+]$ and $L \cap \Sigma \times [s_-, s_- + \delta_-]$ will be called the *positive and the negative collars* of L. The Legendrian submanifolds Λ^- and Λ^+ will be called, respectively, *the negative* and *the positive boundary of* L.

We say that L as above is a *two-part Lagrangian cobordism* if L is a disjoint union of two (not necessarily connected) Lagrangian cobordisms L_0 and L_1

$$L = L_0 \sqcup L_1,$$

where L_0 is a Lagrangian cobordism between the Legendrian submanifolds $\Lambda_0^+ := \Lambda^+ \cap L_0$ and $\Lambda_0^- := \Lambda^- \cap L_0$ and L_1 is a Lagrangian cobordism between the Legendrian submanifolds $\Lambda_1^+ := \Lambda^+ \cap L_1$ and $\Lambda_1^- := \Lambda^- \cap L_1$. In particular, $\Lambda^\pm = \Lambda_0^\pm \sqcup \Lambda_1^\pm$ are two-part Legendrian submanifolds.

Note that in our terminology, a two-part Lagrangian cobordism includes a numbering of its parts.

Let θ be a cobordism 1-form. We say that a two-part Lagrangian cobordism $L = L_0 \sqcup L_1$ is *θ-exact* if $\theta|_{L_i} = df_i$, $i = 0, 1$, for a smooth function $f_i : L_i \to \mathbb{R}$ which is zero on the negative collar of L and is identically equal to a constant C_i on the positive collar of L_i.

If a two-part Lagrangian cobordism L is θ-exact for some cobordism 1-form θ, we call it just *exact*.

The constant $C := C_1 - C_0$ will be called *the gap of L with respect to* θ. We will also say that L is *C-gapped* (with respect to θ).

We say that a two-part Lagrangian cobordism L is *belted* if there exists a null-homologous piecewise-smooth closed path γ in $\Sigma \times [s_-, s_+]$ tracing $\Sigma \times s_+$ from $\Lambda_0^+ = L_0 \cap (\Sigma \times s_+)$ to $\Lambda_1^+ = L_1 \cap (\Sigma \times s_+)$, then tracing L_1 from $\Lambda_1^+ = L_1 \cap (\Sigma \times s_+)$ to $\Lambda_1^- = L_1 \cap (\Sigma \times s_-)$, then following possibly several arcs in L_1 connecting points in Λ_1^-, then tracing $\Sigma \times s_-$ from $\Lambda_1^- = L_1 \cap (\Sigma \times s_-)$ to $\Lambda_0^- = L_0 \cap (\Sigma \times s_-)$, then following possibly several arcs in L_0 connecting points in Λ_0^-, and finally tracing L_0 from $\Lambda_0^- = L_0 \cap (\Sigma \times s_-)$ to $\Lambda_0^+ = L_0 \cap (\Sigma \times s_+)$. We will call such a γ a *belt path of* L.

We claim that if L is belted, then the gap of L with respect to a cobordism 1-form (with respect to which L is exact) does not depend on the form.

Indeed, assume that θ and θ' are cobordism 1-forms on $\Sigma \times [s_-, s_+]$, so that L is exact with respect to both θ and θ'. Then, $\theta - \theta'$ is a closed 1-form vanishing near the positive and the negative boundaries of $\Sigma \times [s_-, s_+]$. Since a belt path γ of L is null-homologous, the integral of $\theta - \theta'$ over γ vanishes. Since $\theta - \theta'$ vanishes near the positive and the negative boundaries of $\Sigma \times [s_-, s_+]$, the latter zero integral is the sum of the integrals of $\theta - \theta'$ over the parts of γ lying in L_0 and L_1, which readily implies the claim.

A trivial Lagrangian cobordism in $(\Sigma \times [s_-, s_+], s\lambda)$ is a cobordism $\Lambda \times [s_-, s_+]$ where Λ is a Legendrian submanifold of (Σ, ξ).

A trivial two-part Lagrangian cobordism

$$L = (\Lambda_0 \sqcup \Lambda_1) \times [s_-, s_+]$$

is belted: to construct a belt path γ for L, take a path Γ in Σ from $x \in \Lambda_0$ to $y \in \Lambda_1$ (it exists, since, by our assumption, Σ is connected). Now, define γ as the path tracing $\Gamma \times s_+$ from $x \times s_+ \in \Lambda_0 \times s_+$ to $y \times s_+ \in \Lambda_1 \times s_+$, then tracing $y \times [s_-, s_+]$ from $y \times s_+$ to $y \times s_-$, then tracing $\Gamma \times s_-$ from $y \times s_-$ to $x \times s_-$, and finally tracing $x \times [s_-, s_+]$ from $x \times s_-$ to $x \times s_+$. Note also that L is $(s\lambda)$-exact with the gap 0.

Given a (two-part exact) Lagrangian cobordism $L \subset \Sigma \times [s_-, s_+]$ between $\Lambda^\pm$, consider the Lagrangian submanifold $\overline{L} \subset \big(\Sigma \times \mathbb{R}_+, d(s\lambda)\big)$ defined as

$$\overline{L} := \big(\Lambda^- \times (0, s_-]\big) \cup L \cup \big(\Lambda^+ \times [s_+, +\infty)\big).$$

We call $\overline{L}$ *the completion of* L.

Let $L \subset (\Sigma \times [s_-, s_+], s\lambda)$ be an exact two-part Lagrangian cobordism between two-part Legendrian submanifolds $\Lambda^\pm \subset (\Sigma, \xi)$. By *an exact Lagrangian cobordism isotopy of* L, we mean a smooth family $\{L^\tau, \theta^\tau\}_{0 \le \tau \le T}$, where $\{L^\tau\}_{0 \le \tau \le T}$ is a Lagrangian isotopy of $L = L^0$ in $\Sigma \times [s_-, s_+]$ and $\{\theta^\tau\}_{0 \le \tau \le T}$ is a smooth family of cobordism 1-forms, so that

(1) Each L^τ is a two-part θ^τ-exact Lagrangian cobordism between its positive and negative boundaries that will be denoted by $\Lambda^\pm_\tau$. In particular, $\{\Lambda^\pm_\tau\}_{0 \le \tau \le T}$ are Legendrian isotopies in $(\Sigma^\pm, \xi^\pm)$. We will say that these are *the Legendrian isotopies induced by the exact Lagrangian cobordism isotopy* $\{L^\tau, \theta^\tau\}_{0 \le \tau \le T}$.

(2) There exist $\delta_\pm > 0$, such that for all $\tau \in [0, T]$

$$L^\tau \cap \Sigma \times [s_+ - \delta_+, s_+] = \Lambda^+_\tau \times [s_+ - \delta_+, s_+],$$
$$L^\tau \cap \Sigma \times [s_-, s_- + \delta_-] = \Lambda^-_\tau \times [s_-, s_- + \delta_-].$$

Note that the gaps $C(\tau)$ of L^τ, $0 \le \tau \le T$, with respect to θ^τ form a smooth function of τ.

Clearly, if L^0 is belted, then so are all L^τ, $0 \le \tau \le T$.

4.2. Chekanov–Eliashberg algebra and the corresponding persistence modules

Let us recall the definition of the Chekanov–Eliashberg algebra associated to a non-degenerate pair (Λ, λ), where $\Lambda \subset (\Sigma, \xi)$ is a compact Legendrian submanifold without boundary. See [22] (cf. [15,21,23]) for more details and [11,26,27] for the original ideas underlying the construction.

Denote by $\mathcal{A}(\Lambda, \lambda)$ a free non-commutative unital algebra over $\mathbb{Z}_2$ generated by the elements of $\mathbb{Z}_2$ and by $\mathcal{R}(\Lambda, \lambda)$.

The algebra $\mathcal{A}(\Lambda, \lambda)$ comes with a filtration defined by the action: the action of a Reeb chord is its time-length and the action of a monomial which is a product of Reeb chords is the sum of the actions of the factors. The action of the constant monomial $1 \in \mathbb{Z}_2$ is set to be zero and the action of 0 is defined as $-\infty$. For $r \in (-\infty, +\infty)$, define $\mathcal{A}_r(\Lambda, \lambda)$ as the vector subspace of $\mathcal{A}(\Lambda, \lambda)$ spanned over $\mathbb{Z}_2$ by the monomials whose action is smaller than r. It is easy to see that for a finite r, the vector space $\mathcal{A}_r(\Lambda, \lambda)$ is finite-dimensional. For any $r \leq r'$, there is a natural morphism $\mathcal{A}_r(\Lambda, \lambda) \to \mathcal{A}_{r'}(\Lambda, \lambda)$ induced by the inclusion of the generators.

For an appropriate (a so-called cylindrical) almost complex structure J on $\Sigma \times \mathbb{R}_+$, one can define a differential ∂_J on $\mathcal{A}(\Lambda, \lambda)$ using a count of J-holomorphic maps of a disk with one positive and several (possibly no) negative boundary punctures into $\Sigma \times \mathbb{R}_+$. Such a map should send the boundary of the disk to $\Lambda \times \mathbb{R}_+$ and converge near positive/negative puncture to a positive/negative cylinder over a Reeb chord in $\mathcal{R}(\Lambda, \lambda)$.

The set $\mathcal{J}(\Lambda)$ of J for which ∂_J is well-defined and $\partial_J^2 = 0$ is connected and dense in the space of all cylindrical almost complex structures on $\Sigma \times \mathbb{R}_+$ [22]; see also [15,21,23].

Let $J \in \mathcal{J}(\Lambda)$.

A computation using Stokes' theorem and similar to [7, Lemma 5.16] and [22, Lemma B.3] shows that for all r, the spaces $\mathcal{A}_r(\Lambda, \lambda)$ are invariant under ∂_J.

For each $r \in (-\infty, +\infty)$, define a vector space $V_r(\Lambda, \lambda)$ over $\mathbb{Z}_2$

$$V_r(\Lambda, \lambda, J) := \frac{\operatorname{Ker} \partial_J|_{\mathcal{A}_r(\Lambda,\lambda)}}{\operatorname{Im} \partial_J|_{\mathcal{A}_r(\Lambda,\lambda)}}.$$

The inclusion maps $\mathcal{A}_r(\Lambda, \lambda) \to \mathcal{A}_{r'}(\Lambda, \lambda)$, $r \leq r'$, induce morphisms $V_r(\Lambda, \lambda, J) \to V_{r'}(\Lambda, \lambda, J)$.

It is easy to see that the vector spaces $V_r(\Lambda, \lambda, J)$, $r \in (0, +\infty)$, together with the morphisms between them, form a persistence module $V(\Lambda, \lambda, J)$ over $(0, +\infty)$.

One can show (see [30]) that for fixed Λ, λ, a different choice of $J \in \mathcal{J}(\Lambda)$ does not change the isomorphism class of the persistence module $V(\Lambda, \lambda, J)$. The isomorphism class of this persistence module will be denoted by $V(\Lambda, \lambda)$. The vector space $V_\infty(\Lambda, \lambda)$ is then isomorphic to the (non-filtered) *Legendrian contact homology of* (Λ, λ)—that is, $\operatorname{Ker} \partial_J / \operatorname{Im} \partial_J$. The dimension of the vector space $V_\infty(\Lambda, \lambda)$ is invariant under Legendrian isotopies of Λ [22]; see also [15,21,23].

Assume now that $\Lambda = \Lambda_0 \sqcup \Lambda_1$ is a two-part Legendrian submanifold.

A sequence of Reeb chords $a_1, \ldots, a_k \in \mathcal{R}(\Lambda, \lambda)$ is called *ij-composable* for $i, j = 0, 1$, if a_1 starts at Λ_i, a_k ends at Λ_j, and for each $m = 1, \ldots, k-1$, the end of a_m lies in the same part of Λ as the origin of a_{m+1}. Note that an ij-composable sequence of Reeb chords must contain at least one chord from $\mathcal{R}_{ij}(\Lambda, \lambda)$. The corresponding monomial $a_1 \cdot \ldots \cdot a_k$ in $\mathcal{A}(\Lambda, \lambda)$ will be also called *ij-composable*.

Denote by $\mathcal{A}(\Lambda_0, \Lambda_1, \lambda)$ the vector subspace of $\mathcal{A}(\Lambda, \lambda)$ (over $\mathbb{Z}_2$) generated by all the 01-composable monomials $a_1 \cdot \ldots \cdot a_k$, $k \in \mathbb{Z}_{>0}$. (Note that the polynomials appearing in $\mathcal{A}(\Lambda_0, \Lambda_1, \lambda)$ have no constant terms!). We will call $\mathcal{A}(\Lambda_0, \Lambda_1, \lambda)$ the 01-*subspace of* $\mathcal{A}(\Lambda, \lambda)$.

Set

$$\mathcal{A}_r(\Lambda_0, \Lambda_1, \lambda) := \mathcal{A}(\Lambda_0, \Lambda_1, \lambda) \cap \mathcal{A}_r(\Lambda, \lambda).$$

Let $J \in \mathcal{J}(\Lambda = \Lambda_0 \sqcup \Lambda_1)$. Then, in particular, $J \in \mathcal{J}(\Lambda_0) \cap \mathcal{J}(\Lambda_1)$.

It is easy to see that $\mathcal{A}(\Lambda_0, \Lambda_1, \lambda)$ is invariant under ∂_J and hence so are the spaces $\mathcal{A}_r(\Lambda_0, \Lambda_1, \lambda)$ for all r. For each $r \in (-\infty, +\infty)$, define a vector space $V_r(\Lambda_0, \Lambda_1, \lambda)$ over $\mathbb{Z}_2$

$$V_r(\Lambda_0, \Lambda_1, \lambda, J) := \frac{\operatorname{Ker} \partial_J|_{\mathcal{A}_r(\Lambda_0,\Lambda_1,\lambda)}}{\operatorname{Im} \partial_J|_{\mathcal{A}_r(\Lambda_0,\Lambda_1,\lambda)}}.$$

The inclusion maps $\mathcal{A}_r(\Lambda_0, \Lambda_1, \lambda) \to \mathcal{A}_{r'}(\Lambda_0, \Lambda_1, \lambda)$, $r \leq r'$, induce morphisms $V_r(\Lambda_0, \Lambda_1, \lambda, J) \to V_{r'}(\Lambda_0, \Lambda_1, \lambda, J)$.

It is easy to see that the vector spaces $V_r(\Lambda_0, \Lambda_1, \lambda, J)$, $r \in (0, +\infty)$, together with the morphisms between them, form a persistence module over $(0, +\infty)$. We will denote this persistence module by $V(\Lambda_0, \Lambda_1, \lambda, J)$. Below, whenever needed, we will also view $V(\Lambda_0, \Lambda_1, \lambda, J)$ as a persistence module over $(-\infty, +\infty)$ using the trivial extension as in Example 3.2.

Similarly to the above, one can show (see [30]) that

– For fixed $\Lambda = \Lambda_0 \sqcup \Lambda_1, \lambda$ a different choice of $J \in \mathcal{J}(\Lambda)$ does not change the isomorphism class of the persistence module $V(\Lambda_0, \Lambda_1, \lambda, J)$. The isomorphism class of this persistence module will be denoted by $V(\Lambda_0, \Lambda_1, \lambda)$.

– The dimension of the vector space $V_\infty(\Lambda_0, \Lambda_1, \lambda)$ is invariant under Legendrian isotopies of Λ.

Abusing the terminology, we will call $V(\Lambda_0, \Lambda_1, \lambda)$ *the LCH persistence module associated to* (Λ, λ), where *LCH* stands for "Legendrian contact homology".

It is easy to see that if $c > 0$, then

$$V(\Lambda_0, \Lambda_1, c\lambda) = V^{[\times 1/c]}(\Lambda_0, \Lambda_1, \lambda),$$

where $V^{[\times 1/c]}(\Lambda_0, \Lambda_1, \lambda)$ is the isomorphism class of persistence modules obtained from $V(\Lambda_0, \Lambda_1, \lambda)$ by the multiplicative shift by $1/c$ (see Remark 3.7).

4.3. Morphisms of persistence modules defined by Lagrangian cobordisms

Let $L = L_0 \sqcup L_1 \subset \big(\Sigma \times [s_-, s_+], \omega = d(s\lambda)\big)$, $s_- < s_+$, be a two-part exact Lagrangian cobordism between two-part Legendrian submanifolds $\Lambda^\pm = \Lambda_0^\pm \sqcup \Lambda_1^\pm \subset (\Sigma, \xi)$.

Assume the pairs $(\Lambda^\pm, \lambda)$ are non-degenerate—in this case, we will say that L is *non-degenerate.*

For an appropriate—a so-called *adapted* (*to* L *and* ω)—almost complex structure I on $\Sigma \times \mathbb{R}_+$ define a unital algebra morphism

$$\Phi^{L,I} : \mathcal{A}(\Lambda^+, s_+\lambda, I^+) \to \mathcal{A}(\Lambda^-, s_-\lambda, I^-)$$

by prescribing its values on the generators

$$\Phi^{L,I}(1) := 1$$

and for any $a \in \mathcal{R}(\Lambda^+, s_+\lambda)$

$$\Phi^{L,I}(a) := \sum_{\dim \mathcal{M}_{L,I}(a;b_1,\ldots,b_m)=0} |\mathcal{M}_{L,I}(a; b_1, \ldots, b_m)| b_1 \cdot \ldots \cdot b_m,$$

where $\mathcal{M}_{L,I}(a; b_1, \ldots, b_m)$ is the moduli space of I-holomorphic maps of a disk with one positive and $m \geq 0$ negative boundary punctures into $\Sigma \times \mathbb{R}_+$ that send the boundary of the disk to $\overline{L}$ and converge near the positive puncture to a cylinder over a and the negative punctures to the cylinders over the chords $b_1, \ldots, b_m \in \mathcal{R}(\Lambda^-, s_-\lambda)$. Here, $I^\pm$ are cylindrical almost complex structures on $\Sigma \times \mathbb{R}_+$ induced by the restrictions of I at the ends of $\Sigma \times \mathbb{R}_+$. See [22], cf. [23], for more details.

The set $\mathcal{I}(L)$ of I for which $\Phi^{L,I}$ is a well-defined unital algebra morphism is dense in the space of all adapted almost complex structures on $\Sigma \times \mathbb{R}_+$ and the map

$$\mathcal{I}(L) \to \mathcal{J}(\Lambda^+) \times \mathcal{J}(\Lambda^-),\ I \mapsto (I^+, I^-)$$

is surjective [22], cf. [23].

The map $\Phi^{L,I}$ is called *the cobordism map associated to L, I.*

Remark 4.1. Assume that $I \in \mathcal{I}(L)$ and the restriction of $\Phi^{L,I}$ to the 01-subspace $\mathcal{A}(\Lambda_0^+, \Lambda_1^+, s_+\lambda)$ is not the zero map. Then, for some 01-chord a and some *non-empty* set of chords $b_1, \ldots, b_m$, the moduli space $\mathcal{M}_{L,I}(a; b_1, \ldots, b_m)$ is non-empty. Rescale an I-holomorphic map of a disk D', with boundary punctures, that defines an element of the moduli space and obtain a map of D' into $\Sigma \times [s_-, s_+]$. Concatenating the image of $\partial D'$ under the latter map with the chords $a, b_1, \ldots, b_m$, we get a belt path for the two-part Lagrangian cobordism L. In other words, we have obtained that, unless L is belted, $\Phi^{L,I}$ has to be the zero map.

The following two claims are proved in [30] using [23, Lemma 3.14] (a chain homotopy result for the cobordism maps, which is a version of [22, Lemma B.15])—see the proof of Proposition 4.5 for a similar use of the same result.

Proposition 4.2 [30]. *Assume that L is a non-degenerate belted two-part exact Lagrangian cobordism. Let $I \in \mathcal{I}(L)$ and assume that the restriction of the cobordism map $\Phi^{L,I}$ to $\mathcal{A}(\Lambda_0^+, \Lambda_1^+, s_+\lambda)$ is non-trivial. Let C be the gap of L (since L is belted, the gap is independent of the cobordism 1-form with respect to which L is exact).*

Then, $\Phi^{L,I}$ maps $\mathcal{A}_r(\Lambda_0, \Lambda_1, s_+\lambda, I^+)$ into $\mathcal{A}_{r-C}(\Lambda_0, \Lambda_1, s_-\lambda, I^-)$ for each $r \in (-\infty, +\infty)$ and, accordingly, defines a morphism of persistence modules over $(-\infty, +\infty)$:

$$\Phi_*^{L,I} : V(\Lambda_0, \Lambda_1, s_+\lambda, I^+) \to V^{[-C]}(\Lambda_0, \Lambda_1, s_-\lambda, I^-),$$

where $V^{[-C]}(\Lambda_0, \Lambda_1, s_-\lambda, I^-)$ is the isomorphism class of persistence modules over $(-\infty, +\infty)$ obtained from $V(\Lambda_0, \Lambda_1, s_-\lambda, I^-)$ by the additive shift by $-C$ (see Remark 3.7).

If the restriction of the cobordism map $\Phi^{L,I}$ to $\mathcal{A}(\Lambda_0^+, \Lambda_1^+, s_+\lambda)$ is the zero map, we set $\Phi_*^{L,I}$ to be the zero morphism into the trivial persistence module.

Proposition 4.3 [30]. *The morphism $\Phi_*^{L,I}$ is independent of $I \in \mathcal{I}(L)$ up to the right–left equivalence in the category of persistence modules over $(-\infty, +\infty)$.*

Further on, we will denote the equivalence class of the persistence module morphism $\Phi_*^{L,I}$ by Φ_*^L.

Abusing the terminology, we will not distinguish between an isomorphism class of a persistence module and a specific persistence module representing it, as well as between an equivalence class of morphisms of persistence modules and a specific morphism representing it. In particular, we will write

$$\Phi_*^L : V(\Lambda_0, \Lambda_1, s_+\lambda) \to V^{[-C]}(\Lambda_0, \Lambda_1, s_-\lambda).$$

Note that compositions of equivalence classes of morphisms of persistence modules are well-defined.

The following proposition is proved in [30] using the well-known results (see [22]) about the cobordism map associated to a trivial exact Lagrangian cobordism.

Proposition 4.4 [30]. *Assume $L = (\Lambda_0 \sqcup \Lambda_1) \times [s_-, s_+]$ is a trivial non-degenerate two-part exact Lagrangian cobordism. Set*

$$U := V(\Lambda_0, \Lambda_1, s_+\lambda).$$

Then

$$\Phi_*^L = Sh_U[\times s_+/s_-].$$

□

Having recalled the needed preparational statements, we present now the key result of this section.

Proposition 4.5. *With L being a non-degenerate belted two-part exact Lagrangian cobordism as above, suppose that $\{L^\tau = L_0^\tau \sqcup L_1^\tau, \theta_\tau\}_{0\le\tau\le T}$ is an exact Lagrangian cobordism isotopy of L with fixed boundary. Assume that Φ_*^L is non-trivial.*

Let $C(\tau)$ be the gap of L^τ, $0 \le \tau \le T$ (since L is belted, it is independent of the cobordism 1-form with respect to which L^τ is exact). Let

$$\begin{aligned} &U := V(\Lambda_0, \Lambda_1, s_+\lambda),\\ &W := V(\Lambda_0, \Lambda_1, s_-\lambda),\\ &C_{\min} := \min_{\tau\in[0,T]} C(\tau),\\ &C_0 := C(0) - C_{\min},\ C_T := C(T) - C_{\min}. \end{aligned}$$

Then

$$Sh_{W^{[-C(0)]}}[+C_0] \circ \Phi_*^{L^0} = Sh_{W^{[-C(T)]}}[+C_T] \circ \Phi_*^{L^T}. \tag{21}$$

Here, the equality is between equivalence classes of morphisms $U \to W^{[-C_{\min}]}$.

Proof of Proposition 4.5. Pick $I_0 \in \mathcal{I}(L^0)$, $I_T \in \mathcal{I}(L^T)$. It follows from [23, Lemma 3.14] (which is a version of [22, Lemma B.15]) that there exists an $\mathbb{Z}_2$-linear map:

$$K : \mathcal{A}(\Lambda^+, s_+\lambda) \to \mathcal{A}(\Lambda^-, s_-\lambda),$$

such that

$$K \circ \partial_{I_0^+} + \partial_{I_T^-} \circ K = \Phi^{L^T, I_T} - \Phi^{L^0, I_0}. \tag{22}$$

The map K is defined on each monomial $a_1 \cdot \ldots \cdot a_k$, $a_1, \ldots, a_k \in \mathcal{R}(\Lambda^+, s_+\lambda)$, as

$$K(a_1 \cdot \ldots \cdot a_k) := \sum_{j=1}^{k} \Phi^{L^T, I_T}(a_1 \cdot \ldots \cdot a_{j-1}) K(a_j) \Phi^{L^0, I_0}(a_{j+1} \cdot \ldots \cdot a_k), \quad (23)$$

so that for each $a \in \mathcal{R}(\Lambda^+, s_+\lambda)$

$$K(a) := \sum_{b_1, \ldots, b_m \in \mathcal{R}(\Lambda^-, s_-\lambda)} n_{\{L^T\}, \{I_T\}}(a; b_1, \ldots, b_m) b_1 \cdot \ldots \cdot b_m, \quad (24)$$

where $\{I_\tau\}_{0 \leq \tau \leq T}$ is a generic family of adapted almost complex structures on M connecting I_0 and I_T, and $n_{\{L^T\}, \{I_T\}}(a; b_1, \ldots, b_m)$ is the mod-2 number of elements (counted using an abstract perturbation scheme, see [22] for details) of a certain moduli space which is non-empty only if

$$\cup_{0 \leq \tau \leq T} \mathcal{M}_{L^\tau, I_\tau}(a; b_1, \ldots, b_m) \neq \emptyset.$$

Here, $\mathcal{M}_{L^\tau, I_\tau}(a; b_1, \ldots, b_m)$ is the moduli space of pseudo-holomorphic maps of the disk with boundary punctures into $\Sigma \times \mathbb{R}_+$ used to define Φ^{L^τ, I_τ}. This moduli space is non-empty only for some finite set of $\tau \in [0, T]$—the corresponding pseudo-holomorphic maps have index -1. See [23, Lemma 3.14] and [22, Lemma B.15] for the details.

Using the energy inequality and the Stokes theorem, one can show that if $\mathcal{M}_{L^\tau, I_\tau}(a; b_1, \ldots, b_m) \neq \emptyset$, then

$$\sum_{i=1}^{m} s_- l(b_i) < s_+ l(a) - C(\tau) \text{ for } a \in \mathcal{R}(\Lambda_0^+, \Lambda_1^+, s_+\lambda),$$
$$\sum_{i=1}^{m} s_- l(b_i) < s_+ l(a) + C(\tau) \text{ for } a \in \mathcal{R}(\Lambda_1^+, \Lambda_0^+, s_+\lambda),$$
$$\sum_{i=1}^{m} s_- l(b_i) < s_+ l(a) \text{ for } a \in \mathcal{R}(\Lambda_i^+, \Lambda_i^+, s_+\lambda), \; i = 0, 1,$$

where $l(\cdot)$ is the action (time-length) of a Reeb orbit. Using these inequalities together with (23), (24) it is not hard to see that K maps the subspace $\mathcal{A}_r(\Lambda_0^+, \Lambda_1^+, s_+\lambda)$ into $\mathcal{A}_{r - C_{\min}}(\Lambda_0^-, \Lambda_1^-, s_-\lambda)$ for any $r \in (-\infty, +\infty)$. Therefore, the chain homotopy formula (22) implies that the restrictions of Φ^{L^T, I_T} and Φ^{L^0, I_0} to $\mathcal{A}_r(\Lambda_0^+, \Lambda_1^+, s_+\lambda)$ induce the same map on homology—that is, a map

$$V_r(\Lambda_0^+, \Lambda_1^+, s_+\lambda, I_0^+) \to V_{r - C_{\min}}(\Lambda_0^-, \Lambda_1^-, s_-\lambda, I_T^-)$$

for any $r \in (-\infty, +\infty)$. This latter map equals, on one hand, the composition of

$$\Phi_{r,*}^{L^0} : V_r(\Lambda_0^+, \Lambda_1^+, s_+\lambda, I_0^+) \to V_{r - C(0)}(\Lambda_0^-, \Lambda_1^-, s_-\lambda, I_T^-)$$

and the shift

$$V_{r - C(0)}(\Lambda_0^-, \Lambda_1^-, s_-\lambda, I_T^-) \to V_{r - C_{\min}}(\Lambda_0^-, \Lambda_1^-, s_-\lambda, I_T^-)$$

and, on the other hand, the composition of

$$\Phi^{L^T}_{r,*} : V_r(\Lambda_0^+, \Lambda_1^+, s_+\lambda, I_0^+) \to V_{r-C(T)}(\Lambda_0^-, \Lambda_1^-, s_-\lambda, I_T^-)$$

and the shift

$$V_{r-C(T)}(\Lambda_0^-, \Lambda_1^-, s_-\lambda, I_T^-) \to V_{r-C_{\min}}(\Lambda_0^-, \Lambda_1^-, s_-\lambda, I_T^-).$$

The equality between the two compositions yields (21). This finishes the proof of the proposition. □

4.4. An invariant of two-part Legendrians via LCH persistence modules

With Σ and λ as above, assume $\Lambda = \Lambda_0 \sqcup \Lambda_1$ is a two-part Legendrian submanifold in $(\Sigma, \xi = \operatorname{Ker}\lambda)$.

Definition 4.6. If the pair (Λ, λ) is non-degenerate, then for each $s > 1$, define $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda)$ as the smallest left end of the bars of multiplicative length greater than s in the barcode of $V(\Lambda_0, \Lambda_1, \lambda)$. Denote by $l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda)$ the smallest left end of the infinite bars in the barcode of $V(\Lambda_0, \Lambda_1, \lambda)$. If there are no such bars, set $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda) := +\infty$ or, respectively, $l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) := +\infty$.

For a general, possibly degenerate, pair (Λ, λ) and $s > 1$, define

$$\begin{aligned} l_{\min,s}(\Lambda_0, \Lambda_1, \lambda) &:= \liminf l_{\min,s}(\Lambda_0', \Lambda_1', \lambda), \\ l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) &:= \liminf l_{\min,\infty}(\Lambda_0', \Lambda_1', \lambda), \end{aligned}$$

where each $\Lambda' = \Lambda_0' \sqcup \Lambda_1'$ is a two-part Legendrian submanifold obtained from Λ by a C^∞-small Legendrian isotopy and such that the pair (Λ', λ) is non-degenerate, and the liminf is taken over all such Λ' as the C^∞-size of the Legendrian isotopy converges to zero. (One can show—see [30]—that for a non-degenerate pair $(\Lambda = \Lambda_0 \sqcup \Lambda_1, \lambda)$, both definitions yield the same $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda)$ and $l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda)$).

For an example where $l_{\min,s}$ can be computed explicitly, see Proposition 6.5.

Note that $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda)$ is a non-decreasing function of s with values in $(0, +\infty]$. In particular

$$l_{\min,s}(\Lambda_0, \Lambda_1, \lambda) \leq l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) \text{ for all } s \in (1, +\infty),$$

and therefore, if the number $l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda)$ is finite, then so is $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda)$.

Recall from Sect. 1.5 that the *stabilization of* Σ is the manifold $\widehat{\Sigma} := \Sigma \times T^*\mathbb{S}^1(r, \tau)$, $\tau \in \mathbb{S}^1$, equipped with the contact form $\widehat{\lambda} := \lambda + r d\tau = dz + \vartheta + r d\tau$. Since $(\Sigma = P \times \mathbb{R}(z), \lambda = dz + \vartheta)$ is nice, so is $(\widehat{\Sigma}, \widehat{\lambda})$ (if $(P, d\vartheta)$ has bounded geometry at infinity, then so does $(P \times T^*\mathbb{S}^1, d\vartheta + dr \wedge dt)$). For a two-part Legendrian submanifold $\Lambda_0 \sqcup \Lambda_1 \subset \Sigma$ and $s > 1$, write

$$\begin{aligned} &\widehat{\Lambda}_i := \Lambda_i \times \{r = 0\} \subset \widehat{\Sigma}, \ i = 0, 1, \\ &\widehat{l}_{\min,s}(\Lambda_0, \Lambda_1, \lambda) := l_{\min,s}(\widehat{\Lambda}_0, \widehat{\Lambda}_1, \widehat{\lambda}), \\ &\widehat{l}_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) := l_{\min,\infty}(\widehat{\Lambda}_0, \widehat{\Lambda}_1, \widehat{\lambda}). \end{aligned}$$

Definition 4.7. We say that the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is *weakly homologically bonded*, if $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda) < +\infty$ for all $s > 1$—that is, there are bars of arbitrarily large multiplicative length in the barcode of $V(\Lambda_0, \Lambda_1, \lambda)$.

We say that the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is *homologically bonded*, if $l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) < +\infty$—that is, there are infinite bars in the barcode of $V(\Lambda_0, \Lambda_1, \lambda)$.

We say that the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is *stably homologically bonded*, if $\widehat{l}_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) < +\infty$.

Clearly, homological bondedness implies weak homological bondedness. We do not know whether homological bondedness implies stable homological bondedness—see Remark 1.14.

Remark 4.8. Assume that $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda) < +\infty$ for some $s \in (1, +\infty]$. Let $\Lambda' = \Lambda'_0 \sqcup \Lambda'_1$ be a two-part Legendrian submanifold in (Σ, ξ) obtained from $\Lambda = \Lambda_0 \sqcup \Lambda_1$ by a contact isotopy $\{\phi_t\}$ with the conformal factor satisfying $||(\phi_t^*)^{-1}\lambda/\lambda - 1|| < \delta$ for all t. If $\delta = \delta(s)$ is small enough, then $l_{\min,s}(\Lambda'_0, \Lambda'_1, \lambda)$ is also a finite number which tends to $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda)$ as $\delta \to 0$. The proof will appear in [30].

4.5. A lower bound on pb^+ via LCH persistence modules

With Σ as in (20) and $0 < s_- < s_+$, denote for brevity

$$M := \Sigma \times [s_-, s_+].$$

Recall that $\omega := d(s\lambda)$. Assume $\Lambda = \Lambda_0 \sqcup \Lambda_1$ is a two-part Legendrian submanifold in (Σ, ξ). Let $L = \Lambda \times [s_-, s_+]$ be the corresponding trivial two-part exact Lagrangian cobordism in $(M = \Sigma \times [s_-, s_+], s\lambda)$.

Define an admissible quadruple $X_0, X_1, Y_0, Y_1 \subset M$ as follows:

$$X_0 := \Lambda_0 \times [s_-, s_+],\ X_1 := \Lambda_1 \times [s_-, s_+], \tag{25}$$
$$Y_0 := \Sigma \times s_-,\ Y_1 := \Sigma \times s_+. \tag{26}$$

Theorem 4.9. *Assume that $l_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda) < +\infty$.*

Then

$$pb^+_M(X_0, X_1, Y_0, Y_1) \geq \frac{1}{(s_+ - s_-)l_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda)} > 0.$$

Proof of Theorem 4.9. Assume first that the pair (Λ, λ) is non-degenerate.

Set

$$V := V(\Lambda_0, \Lambda_1, \lambda),$$
$$U := V^{[\times 1/s_+]} = V(\Lambda_0, \Lambda_1, s_+\lambda),$$
$$W := V^{[\times 1/s_-]} = U^{[\times s_+/s_-]} = V(\Lambda_0, \Lambda_1, s_-\lambda).$$

Let $(F, G) \in \mathcal{F}''_M(X_0, X_1, Y_0, Y_1)$. We need to show that

$$\sup_M \{F, G\} \geq \frac{1}{(s_+ - s_-)l_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda)}. \tag{27}$$

Following [8], consider the deformation

$$\omega_\tau := \omega + \tau dF \wedge dG,\ \tau \in \mathbb{R},$$

of ω. A direct calculation shows that

$$dF \wedge dG \wedge \omega^{n-1} = -\frac{1}{n}\{F,G\}\omega^n,$$

and therefore

$$\omega_\tau^n = (1 - \tau\{F,G\})\omega^n.$$

Thus, ω_τ is symplectic for any $\tau \in \mathsf{I} := [0, 1/\sup_M\{F,G\})$. Note that L is Lagrangian with respect to ω_τ for all $\tau \in \mathsf{I}$.

Following the idea underlying Moser's method [38], define a (time-dependent) vector field v_τ, $\tau \in \mathsf{I}$, on M by

$$FdG = -i_{v_\tau}\omega_\tau.$$

One can check that

$$v_\tau = \frac{F}{1 - \tau\{F,G\}}\, sgrad\, G.$$

Since $(F,G) \in \mathcal{F}''_M(X_0, X_1, Y_0, Y_1)$, the vector field $F sgrad\, G$ is complete. Also, the function $1 - \tau\{F,G\}$ is bounded from below by the constant $1 - \tau \sup_M\{F,G\}$ which is positive since $\tau \in \mathsf{I}$. Thus, for each $T \in \mathsf{I}$, the time-dependent vector field v_τ, $0 \le \tau \le T$, equals to the product of $F sgrad\, G$ with a non-negative function bounded from above by a constant depending on F, G and T. Therefore, the time-$[0,T]$ flow $\sigma_{\tau_0} : M \to M$ of v_τ, $0 \le \tau \le T$, is well-defined. It is identity near the boundary of M, because v_τ vanishes there (since G is constant near the positive and negative boundaries of M).

For each $\tau \in \mathsf{I}$, define

$$L^\tau := (\sigma_\tau)^{-1}(L).$$

A direct check shows that $\sigma_\tau^*\omega_\tau = \omega$ for all $\tau \in \mathsf{I}$. Therefore, L^τ is Lagrangian with respect to ω for all $\tau \in \mathsf{I}$.

For each $\tau \in \mathsf{I}$, set

$$\theta^\tau := \sigma_\tau^*(s\lambda + \tau F dG).$$

Since $d(s\lambda + \tau F dG) = \omega_\tau$ and $\sigma_\tau^*\omega_\tau = \omega$ and since G is constant and σ_τ is identity near the positive and negative boundaries of M, one readily gets that each θ^τ is a cobordism 1-form.

Since the trivial cobordism L is $(s\lambda)$-exact and F is 0 near L_0 and 1 near L_1, we get that $(s\lambda + \tau F dG)|_{L_i} = df_i$, $i = 0, 1$, for a smooth function $f_i : L_i \to \mathbb{R}$ equal to 0 near the negative boundary of L_i and to some constant near the positive boundary of L_i. Consequently, for each $\tau \in \mathsf{I}$, the two-part Lagrangian cobordism L^τ is θ^τ-exact.

Therefore, for each $T \in \mathsf{I}$, the family $\{L^\tau, \theta^\tau\}_{0 \le \tau \le T}$ is an exact Lagrangian cobordism isotopy.

Since $L^0 = L$ is a trivial exact Lagrangian cobordism, Proposition 4.4 yields

$$\Phi_*^{L^0} = Sh_U[\times s_+/s_-] : U \to W.$$

Note that since L^τ is belted, the gap $C(\tau)$ of L^τ is independent of the cobordism 1-form with respect to which L^τ is exact. Using the assumptions on F and G, one easily sees that

$$C(\tau) = \tau.$$

Since $C_0 = C_{\min} = C(0) = 0$ and $C_T = C(T) = T$, by Proposition 4.5

$$Sh_U[\times s_+/s_-] = Sh_{W[-T]}[+T] \circ \Phi_*^{L^T}.$$

Therefore, for each $t > 0$ the morphism $U_t \to U_{ts_+/s_-}$ in the persistence module, U is a composition of a linear map $U_t \to U_{(t-T)s_+/s_-}$, given by $\Phi_*^{L^T}$, and a morphism $U_{(t-T)s_+/s_-} \to U_{ts_+/s_-}$ in the persistence module U.

Assume that for some $t > 0$, both t and ts_+/s_- lie in a bar $\mathbb{J}$ in the barcode of U. Denote by $\pi_{s,ts_+/s_-,\mathbb{J}} : U_t \cap Q(\mathbb{J}) = Q_t(\mathbb{J}) \to U_{ts_+/s_-} \cap Q(\mathbb{J}) = Q_{ts_+/s_-}(\mathbb{J})$ the restriction of the morphism $U_t \to U_{ts_+/s_-}$ restricted to the interval persistence submodule $Q(\mathbb{J})$ of U_t (see Example 3.3). The persistence morphism $\pi_{s,ts_+/s_-,\mathbb{J}}$ is an isomorphism. By the conclusion above, this isomorphism is the composition of a map $U_t \cap Q(\mathbb{J}) = Q_t(\mathbb{J}) \to U_{(t-T)s_+/s_-}$ and a morphism $U_{(t-T)s_+/s_-} \cap Q(\mathbb{J}) \to U_{ts_+/s_-} \cap Q(\mathbb{J})$ in the persistence module $U_t \cap Q(\mathbb{J})$, which also has to be an isomorphism. This is possible only if $(t-T)s_+/s_- \in \mathbb{J}$, which yields an upper bound on T in the following way. Namely, take $\mathbb{J}$ to be a bar of multiplicative length greater than s_+/s_- in the barcode of U whose left end is $s_+ l_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda)$ (this is the smallest left end of the bars of multiplicative length greater than s_+/s_- in the barcode of U) and let t be arbitrarily close from above to the left end of $\mathbb{J}$—that is, to $s_+ l_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda)$. Then, t and ts_+/s_- lie in the same bar $\mathbb{J}$ and, therefore, as we have shown above, $(t-T)s_+/s_-$ lies in $\mathbb{J}$ and, in particular, is greater or equal than the left end of $\mathbb{J}$:

$$(t-T)s_+/s_- \geq s_+ l_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda).$$

Hence

$$T \leq (s_+ - s_-) l_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda).$$

Since this is true for any $T \in \mathsf{I} = [0, 1/\max_M\{F, G\})$, we obtain (27) as required. This finishes the proof of the theorem for the case where the pair (Λ, λ) is non-degenerate.

The general case now follows from the non-degenerate one by the semi-continuity of the Poisson bracket invariant—see (17). $\square$

As in Sect. 4.4, let

$$\widehat{\Sigma} := \Sigma \times \mathbb{R}(r) \times \mathbb{S}^1(\tau) = \Sigma \times T^*\mathbb{S}^1(r, \tau)$$

be the stabilization of Σ equipped with the contact form

$$\widehat{\lambda} := \lambda + r d\tau = dz + \vartheta + r d\tau.$$

Set

$$\widetilde{M} := M \times T^*\mathbb{S}^1(r, \tau), \quad M = \Sigma \times [s_-, s_+],$$

and equip $\widetilde{M}$ with the symplectic form $\omega + dr \wedge d\tau = d(s\lambda) + dr \wedge d\tau$.

For each $E > 0$ define $\widetilde{M}_E \subset \widetilde{M}$ as

$$\widetilde{M}_E := M \times (-E, E) \times \mathbb{S}^1 \subset \widetilde{M}.$$

With the admissible quadruple $X_0, X_1, Y_0, Y_1 \subset M$ as above, set

$$\begin{aligned}
&\widetilde{X}_0 := X_0 \times \{r = 0\},\ \widetilde{X}_1 := X_1 \times \{r = 0\} \subset \widetilde{M}_E \subset \widetilde{M},\\
&\widetilde{Y}_0 := Y_0 \times T^*\mathbb{S}^1,\ \widetilde{Y}_1 := Y_1 \times T^*\mathbb{S}^1 \subset \widetilde{M},\\
&\widetilde{Y}_0(E) := Y_0 \times (-E, E) \times \mathbb{S}^1,\ \widetilde{Y}_1(E) := Y_1 \times (-E, E) \times \mathbb{S}^1 \subset \widetilde{M}_E.
\end{aligned}$$

Theorem 4.10. *Assume that* $\widehat{l}_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda) < +\infty$.
Then, for each $E > 0$

$$\begin{aligned}
&pb^+_{\widetilde{M}_E,\mathrm{comp}}\left(\widetilde{X}_0, \widetilde{X}_1, \widetilde{Y}_0(E), \widetilde{Y}_1(E)\right) \geq pb^+_{\widetilde{M},\mathrm{comp}}(\widetilde{X}_0, \widetilde{X}_1, \widetilde{Y}_0, \widetilde{Y}_1)\\
&\geq \frac{1}{(s_+ - s_-)\widehat{l}_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda)} > 0.
\end{aligned}$$

Proof of Theorem 4.10. Consider the exact symplectic cobordism

$$\widehat{M} := \left(\widehat{\Sigma} \times [s_-, s_+], d(s\widehat{\lambda})\right).$$

Define $\widehat{X}_0, \widehat{X}_1, \widehat{Y}_0, \widehat{Y}_1 \subset \widehat{M}$ by

$$\begin{aligned}
\widehat{X}_0 &:= \Lambda_0 \times \mathbb{S}^1 \times [s_-, s_+] = \widehat{\Lambda}_0 \times [s_-, s_+],\\
\widehat{X}_1 &:= \Lambda_1 \times \mathbb{S}^1 \times [s_-, s_+] = \widehat{\Lambda}_1 \times [s_-, s_+],\\
\widehat{Y}_0 &:= \Sigma \times T^*\mathbb{S}^1 \times s_-,\\
\widehat{Y}_1 &:= \Sigma \times T^*\mathbb{S}^1 \times s_+.
\end{aligned}$$

Since $\widehat{l}_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda) < +\infty$, Theorem 4.9, applied to $\widehat{M}$, $\widehat{X}_0$, $\widehat{X}_1$, $\widehat{Y}_0$, $\widehat{Y}_1$, together with the inequality $pb^+_{\widehat{M},\mathrm{comp}}(\widehat{X}_0, \widehat{X}_1, \widehat{Y}_0, \widehat{Y}_1) \geq pb^+_{\widehat{M}}(\widehat{X}_0, \widehat{X}_1, \widehat{Y}_0, \widehat{Y}_1)$, yields

$$pb^+_{\widehat{M},\mathrm{comp}}(\widehat{X}_0, \widehat{X}_1, \widehat{Y}_0, \widehat{Y}_1) \geq \frac{1}{(s_+ - s_-)\widehat{l}_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda)} > 0. \tag{28}$$

Consider a map

$$\begin{aligned}
\widehat{M} = \widehat{\Sigma} \times [s_-, s_+] &= \Sigma \times T^*\mathbb{S}^1(r, \tau) \times [s_-, s_+]\\
&\to \widetilde{M} = \Sigma \times [s_-, s_+] \times T^*\mathbb{S}^1(u, \tau)
\end{aligned}$$

that sends each $(x, r, \tau, s) \in \Sigma \times T^*\mathbb{S}^1(r, \tau) \times [s_-, s_+]$ to $(x, s, u = sr, \tau) \in \Sigma \times [s_-, s_+] \times T^*\mathbb{S}^1$. (We use two copies of $T^*\mathbb{S}^1$—one with the coordinates r, τ and one with the coordinates u, τ). It is a symplectomorphism—it identifies the symplectic form

$$d(s\widehat{\lambda}) = d\left(s(\lambda + rd\tau)\right) = d(s\lambda) + d(srd\tau) = \omega + d(srd\tau)$$

on $\widehat{M}$ with the symplectic form

$$\omega + du \wedge d\tau$$

on $\widetilde{M}$. This symplectomorphism maps the sets $\widehat{X}_0, \widehat{X}_1, \widehat{Y}_0, \widehat{Y}_1 \subset \widehat{M}$, respectively, to the sets $\widetilde{X}_0, \widetilde{X}_1, \widetilde{Y}_0, \widetilde{Y}_1 \subset \widetilde{M}$. Therefore, (28) yields

$$\begin{aligned} pb^+_{\widehat{M},\mathrm{comp}}(\widehat{X}_0, \widehat{X}_1, \widehat{Y}_0, \widehat{Y}_1) &= pb^+_{\widetilde{M},\mathrm{comp}}(\widetilde{X}_0, \widetilde{X}_1, \widetilde{Y}_0, \widetilde{Y}_1) \\ &\geq \frac{1}{(s_+ - s_-)\widehat{l}_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda)} > 0. \end{aligned}$$

Combining this with the inequality

$$pb^+_{\widetilde{M}_E,\mathrm{comp}}\left(\widetilde{X}_0, \widetilde{X}_1, \widetilde{Y}_0(E), \widetilde{Y}_1(E)\right) \geq pb^+_{\widetilde{M},\mathrm{comp}}(\widetilde{X}_0, \widetilde{X}_1, \widetilde{Y}_0, \widetilde{Y}_1)$$

(that follows from (16)) finishes the proof of the theorem. □

Assume that $(M, d(s\lambda))$ is a codimension-zero symplectic submanifold with boundary of a symplectic manifold (N, Ω) which is closed as a subset of N. Consequently, X_0, X_1, Y_0, Y_1 can be viewed as subsets of N. Let $H : N \times \mathbb{S}^1(t) \to \mathbb{R}$ be a complete Hamiltonian.

Corollary 4.11. *Assume that the Hamiltonian H is time-independent and $l_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda) < +\infty$.*

Then, the following claims hold:

A. *If $\Delta(H; Y_0, Y_1) > 0$, then there exists a chord of H from X_0 to X_1 of time-length $\leq \dfrac{(s_+ - s_-)l_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda)}{\Delta(H; Y_0, Y_1)}$.*

B. *If $\mathrm{supp}\, H \cap M$ is compact and $H|_{X_0} \geq 0$ and $\Delta(H; X_1, X_0) > 0$, then there exists a chord of H from Y_0 to Y_1 of time-length bounded from above by $\dfrac{(s_+ - s_-)l_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda)}{\Delta(H; X_1, X_0)}$.*

Proof of Corollary 4.11. Follows directly from (14), Proposition 2.6, and Theorem 4.9. □

Let us now consider the case where H is time-dependent. For each $t_0 \leq t$, denote by $\phi_{t_0,t} : N \to N$ the time-$[t_0, t]$ flow of H. Set

$$\begin{aligned} &\Delta := \Delta(H; Y_0, Y_1), \\ &\widehat{l}_{\min,s_+/s_-} := \widehat{l}_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda), \\ &c_{\min} := \min_{X_0 \times \mathbb{S}^1} H, \; c_{\max} := \max_{X_0 \times \mathbb{S}^1} H. \end{aligned}$$

Corollary 4.12. *Let $0 < e < 1/2$. Set*

$$\begin{aligned} &E := e\Delta, \\ &T := \frac{(s_+ - s_-)\widehat{l}_{\min,s_+/s_-}}{(1 - 2e)\Delta}. \end{aligned}$$

Assume that

(a) $\widehat{l}_{\min,s_+/s_-} < +\infty$,

(b) $\Delta > 0$,

(c) $\displaystyle\sup_{c_{\min} - E \leq H \leq c_{\max} + E} |\partial H/\partial t| < E/T = \frac{e(1 - 2e)\Delta^2}{(s_+ - s_-)\widehat{l}_{\min,s_+/s_-}}$.

Then, there exists a chord of H from X_0 to X_1 of time-length bounded from above by $T = \dfrac{\widehat{l}_{\min,s_+/s_-}(s_+ - s_-)}{(1-2e)\Delta}$.

Proof of Corollary 4.12. Define

$$\widetilde{N}_E := N \times (-E, E) \times \mathbb{S}^1,$$

and equip it with the symplectic form $\Omega + dr \wedge d\tau$, where r, τ are, respectively, the coordinates on $(-E, E)$ and $\mathbb{S}^1$. Then, $\widetilde{M}_E = M \times (-E, E) \times \mathbb{S}^1$ is a closed codimension-zero symplectic submanifold with boundary of $\widetilde{N}_E$. With $\widetilde{X}_0, \widetilde{X}_1, \widetilde{Y}_0(E), \widetilde{Y}_1(E) \subset \widetilde{M}_E \subset \widetilde{N}_E$ defined as above, we get, by (14),(15) and Theorem 4.10, that

$$\begin{aligned}\widetilde{p}_E &:= pb^+_{\widetilde{N}_E,\mathrm{comp}}\left(\widetilde{X}_0, \widetilde{X}_1, \widetilde{Y}_0(E), \widetilde{Y}_1(E)\right) \geq \\ &\geq pb^+_{\widetilde{M}_E,\mathrm{comp}}\left(\widetilde{X}_0, \widetilde{X}_1, \widetilde{Y}_0(E), \widetilde{Y}_1(E)\right) \geq \frac{1}{(s_+ - s_-)\widehat{l}_{\min,s_+/s_-}} > 0.\end{aligned}$$

Note that for any $x \in N$ and $t_0 \in \mathbb{R}$

$$\frac{\mathrm{d}}{\mathrm{d}t} H(\phi_{t_0,t}(x), t) = \frac{\partial H}{\partial t}(\phi_{t_0,t}(x), t).$$

Therefore, condition (c) implies that for any $x \in X_0$ and any $t_0 \in \mathbb{R}$, $t \in [t_0, t_0 + T]$,

$$\phi_{t_0,t}(x) \in \{c_{\min} - E \leq H \leq c_{\max} + E\}$$

and

$$|\partial H/\partial t(\phi_{t_0,t}(x), t)| < E/T.$$

In particular, this means that

$$\sup_{t_0 \in \mathbb{R}} \left(\sup_{t \in [t_0, t_0+T]} H(\phi_{t_0,t}(x), t) - \inf_{t \in [t_0, t_0+T]} H(\phi_{t_0,t}(x), t) \right) < \frac{E}{T} \cdot T = E.$$

This inequality, together with the positivity of $\widetilde{p}_E$ and of Δ, allows to apply Proposition 2.7 to H, N, and $\widetilde{N}_E$, which yields the existence of the chord of H from X_0 to X_1 of time-length $\leq \dfrac{1}{\widetilde{p}_E(\Delta - 2E)} \leq T$. □

Remark 4.13. The claim on the positivity of $pb^+_M(X_0, X_1, Y_0, Y_1)$ in Theorem 4.9 can be generalized to arbitrary two-part exact Lagrangian cobordisms, with $(s_+ - s_-)l_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda)$ being replaced by a certain invariant associated with the morphism of the persistence modules defined by the cobordism. In particular, the positivity of $pb^+_M(X_0, X_1, Y_0, Y_1)$ remains true if the sets X_0, X_1 as in Theorem 4.9 are perturbed as exact Lagrangian cobordisms in $(M, d(s\lambda))$ (cylindrical near the boundaries), so that the Legendrian isotopies, induced on the boundaries by the exact Lagrangian isotopies, are sufficiently C^1-small (note that away from the boundary the perturbations may be arbitrarily large, as long as the perturbed X_0, X_1 are disjoint!). The lower bound on $pb^+_M(X_0, X_1, Y_0, Y_1)$ for the perturbed X_0, X_1 is then only slightly larger than the one appearing in Theorem 4.9—the difference between the bounds tends to zero as the sizes of the Legendrian isotopies above tend to zero. In the particular case where a trivial Lagrangian cobordism is deformed among trivial Lagrangian cobordisms, the robustness follows from Theorem 4.9 and Remark 4.8.

For the proofs and details, see [30].

Consequently, the results of Theorem 4.10 and Corollaries 4.11 and 4.12 are also robust with respect to the above perturbations.

5. Applications to contact dynamics

With Σ and λ as above, assume that $\Lambda = \Lambda_0 \sqcup \Lambda_1$ is a two-part Legendrian submanifold in $(\Sigma, \xi = \ker \Lambda)$.

Let $h : \Sigma \times \mathbb{S}^1 \to \mathbb{R}$ be a complete (time-dependent) contact Hamiltonian (with respect to λ). Set $h_t := h(\cdot, t) : \Sigma \to \mathbb{R}$. Let v_t, $t \in \mathbb{S}^1$, denote the contact vector field of h_t. If h and v are time-independent, we write v instead of v_t. Denote by $\{\varphi_t\}$ the time-$[0,t]$ contact flow of h—that is, the time-$[0,t]$ flow of v_t. The flow $\{\varphi_t\}$ lifts to a Hamiltonian flow $\{\phi_t\}$ on $(\Sigma \times \mathbb{R}_+, d(s\lambda))$ equivariant with respect to the multiplicative $\mathbb{R}_+$-action on $\Sigma \times \mathbb{R}_+$ and generated by the Hamiltonian $H : \Sigma \times \mathbb{R}_+ \times \mathbb{S}^1 \to \mathbb{R}$, $H(y, s, t) := s \cdot h(y, t)$. The flow $\{\phi_t\}$ has the form

$$\phi_t(y, s) = \left(\varphi_t\left(y\right), s \frac{\left(\varphi_t^{-1}\right)^* \lambda\left(\varphi_t\left(y\right)\right)}{\lambda\left(\varphi_t\left(y\right)\right)} \right). \tag{29}$$

Since the contact flow $\{\varphi_t\}$ of h is defined for all times, so is the Hamiltonian flow $\{\phi_t\}$ of H, meaning that H is complete.

Let us recall Definition 1.8.

Definition 5.1. Let us say that $h : \Sigma \times \mathbb{S}^1 \to \mathbb{R}$ is *C-cooperative with* Λ_0, Λ_1 for $C > 0$ if either of the following conditions holds:

(a) $h < C$ on $\Lambda_1 \times \mathbb{S}^1$ and either the set $\{h \geq C\} = \bigcup_{t \in \mathbb{S}^1} \{h_t \geq C\}$ is empty or $dh_t(R) \geq 0$ on $\{h_t \geq C\}$ for all $t \in \mathbb{S}^1$.

(b) $h < C$ on $\Lambda_0 \times \mathbb{S}^1$ and either the set $\{h \geq C\} = \bigcup_{t \in \mathbb{S}^1} \{h_t \geq C\}$ is empty or $dh_t(R) \leq 0$ on $\{h_t \geq C\}$ for all $t \in \mathbb{S}^1$.

We will say that h is *cooperative with* Λ_0, Λ_1 if it is C-cooperative with Λ_0, Λ_1 for some $C > 0$.

5.1. Largeness of the conformal factor of φ_t

Theorem 5.2. *Assume that h is time-independent and compactly supported. Assume also that*

$$h|_{\Lambda_0} \geq 0, \;\; h|_{\Lambda_1} < 0.$$

and the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is weakly homologically bonded.

Then, the conformal factor of φ_t takes arbitrarily large values as t varies between 0 *and* $+\infty$:

$$\inf_{t \in (0, +\infty), y \in \Sigma} \frac{\left(\varphi_t^{-1}\right)^* \lambda\left(\varphi_t\left(y\right)\right)}{\lambda\left(\varphi_t\left(y\right)\right)} = +\infty.$$

Remark 5.3. It would be interesting to generalize Theorem 5.2 to contact Hamiltonians h that are not necessarily compactly supported but rather are constant outside a compact set $K \subset \Sigma$, meaning that the contact Hamiltonian flow of such an h outside K is a reparameterized Reeb flow and the conformal factor of the flow outside K is identically equal to 1, since the Reeb flow preserves the contact form. (Recall that, by our assumptions, the Reeb flow is defined for all times, so such an h would be complete). Such a generalization would be true if *in this particular setting*—for $Y_0 = \Sigma \times s_-$, $Y_1 = \Sigma \times s_+$ and the vector field *sgrad* (sh) on $(\Sigma \times \mathbb{R}_+, d(s\lambda))$—Fathi's Theorem 2.1 would provide a function $G : \Sigma \times \mathbb{R}_+ \to \mathbb{R}$, $G \in \mathcal{S}'(Y_0, Y_1)$ (see Proposition 2.2), so that the flow of *sgrad* G on $\Sigma \times [s_-, s_+]$ is complete.

Proof of Theorem 5.2. Pick $0 < s_- < s_+$. Let $X_0, X_1, Y_0, Y_1 \subset \Sigma \times [s_-, s_+]$ be the admissible quadruple defined for s_-, s_+ as in (25),(26).

Clearly, $H|_{X_0} \geq 0$, $\Delta(H; X_1, X_0) > 0$.

We claim that there is a chord of H from Y_0 to Y_1. Indeed, consider a cut-off function $\chi : \mathbb{R}_+ \to [0, 1]$ which equals 1 on $[s_-, s_+]$ and 0 outside $(s_- - \epsilon, s_+ + \epsilon)$ for some $0 < \epsilon < s_-$. Since $h : \Sigma \to \mathbb{R}$ is compactly supported, the Hamiltonian $\chi(s)h : \Sigma \times \mathbb{R}_+ \to \mathbb{R}$ is also compactly supported and coincides with $H = sh$ on $\Sigma \times [s_-, s_+]$. In particular, $\chi(s)H|_{X_0} \geq 0$, $\Delta(\chi(s)H; X_1, X_0) > 0$. Together with the assumption that the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is weakly homologically bonded, this allows to apply part B of Corollary 4.11 to the symplectic manifold $(N, \Omega) = (\Sigma \times \mathbb{R}_+, d(s\lambda))$ and the Hamiltonian $\chi(s)h$. This yields the existence of a chord of $\chi(s)h$ from Y_0 to Y_1 in N. An easy topological argument then yields that there exists a chord of $\chi(s)h$ from Y_0 to Y_1 in N that lies in $\Sigma \times [s_-, s_+]$ and therefore is a chord H. This proves the claim.

The existence of the chord of H from Y_0 to Y_1, together with (29), implies that

$$\left(\varphi_t^{-1}\right)^* \lambda\left(\varphi_t\left(y\right)\right) / \lambda\left(\varphi_t\left(y\right)\right) \geq s_+/s_-.$$

Since s_+/s_- can be made arbitrarily large, we get that

$$\inf_{t \in (0, +\infty), y \in \Sigma} \frac{\left(\varphi_t^{-1}\right)^* \lambda\left(\varphi_t\left(y\right)\right)}{\lambda\left(\varphi_t\left(y\right)\right)} = +\infty,$$

which finishes the proof of the theorem. □

5.2. Existence of chords of h from Λ_0 to Λ_1

Theorem 5.4. (Cf. Rem. 1.14 in [28]) *Assume*

$$0 < \inf_{\Sigma \times \mathbb{S}^1} h \leq \sup_{\Sigma \times \mathbb{S}^1} h < +\infty,$$

and let

$$s_+ > \frac{\sup_{\Sigma \times \mathbb{S}^1} h}{\inf_{\Sigma \times \mathbb{S}^1} h} \geq 1.$$

Then, the following claims hold:

A. Assume that h is time-independent and
$l_{\min, s_+}(\Lambda_0, \Lambda_1, \lambda) =: l_{\min, s_+} < +\infty$.

Then, there exists a chord of h *from* Λ_0 *to* Λ_1 *of time-length bounded from above by* $\dfrac{(s_+ - 1)l_{\min,s_+}}{s_+ \inf_\Sigma h - \sup_\Sigma h}$.

B. Assume that $\widehat{l}_{\min,s_+}(\Lambda_0, \Lambda_1, \lambda) =: \widehat{l}_{\min,s_+} < +\infty$. *Let*

$$\Delta_{s_+} := s_+ \inf_{\Sigma\times\mathbb{S}^1} h - \sup_{\Sigma\times\mathbb{S}^1} h.$$

Assume also that for some $0 < e < 1/2$

$$\sup_{\Sigma\times\mathbb{S}^1} |\partial h/\partial t| < \frac{(1-2e)e\Delta^2_{s_+} \inf_{\Sigma\times\mathbb{S}^1} h}{(s_+ - 1)(s_+ \max_{\Lambda_0\times\mathbb{S}^1} h + e\Delta_{s_+})\widehat{l}_{\min,s_+}}. \tag{30}$$

Then, there exists a chord of h *from* Λ_0 *to* Λ_1 *of time-length bounded from above by* $\dfrac{(s_+ - 1)\widehat{l}_{\min,s_+}}{(1-2e)\Delta_{s_+}}$.

Proof of Theorem 5.4. Pick $s_- := 1 < s_+$. Let $X_0, X_1, Y_0, Y_1 \subset \Sigma \times [1, s_+]$ be the admissible quadruple defined for $s_- = 1, s_+$ as in (25),(26). Clearly

$$\Delta(H; Y_0, Y_1) := \Delta_{s_+} = s_+ \inf_\Sigma h - \sup_\Sigma h.$$

Since by the hypothesis of the theorem

$$0 < \inf_{\Sigma\times\mathbb{S}^1} h \le \sup_{\Sigma\times\mathbb{S}^1} h < +\infty,$$

we get that $\Delta(H; Y_0, Y_1) = \Delta_{s_+} > 0$ if $s_+ > \dfrac{\sup_{\Sigma\times\mathbb{S}^1} h}{\inf_{\Sigma\times\mathbb{S}^1} h}$.

Let us now prove part A of the theorem. Its hypothesis allows to apply part A of Corollary 4.11 to $(N, \Omega) = (\Sigma \times \mathbb{R}_+, d(s\lambda))$ and the Hamiltonian H on it as long as $s_+ > \dfrac{\sup_{\Sigma\times\mathbb{S}^1} h}{\inf_{\Sigma\times\mathbb{S}^1} h}$. This yields the existence of a Hamiltonian chord of H from X_0 to X_1 of time-length bounded from above by

$$\frac{(s_+ - 1)l_{\min,s_+}(\Lambda_0, \Lambda_1, \lambda)}{s_+ \inf_\Sigma h - \sup_\Sigma h}.$$

The projection of this Hamiltonian chord to Σ is a chord of h from Λ_0 to Λ_1 of the same time-length. This finishes the proof of part A the theorem.

Let us prove part B of the theorem. Similarly to the setting of Corollary 4.12, for a given $0 < e < 1/2$, define

$$E := e\Delta_{s_+},$$
$$T := \frac{(s_+ - 1)\widehat{l}_{\min,s}}{(1-2e)\Delta_{s_+}},$$
$$c_{\min} := \min_{X_0\times\mathbb{S}^1} H = \min_{\Lambda_0\times\mathbb{S}^1} h.$$
$$c_{\max} := \max_{X_0\times\mathbb{S}^1} H = s_+ \max_{\Lambda_0\times\mathbb{S}^1} h.$$

Consider the set $S := \{c_{\min} - E \le H = sh \le c_{\max} + E\}$. We would like to apply Corollary 4.12, and in order to do this, we need to verify that the upper bound on the restriction of the function $|\partial H/\partial t|$ to S, required in Corollary 4.12, does hold in our case.

Note that on S

$$s \leq \frac{c_{\max} + E}{\inf_{\Sigma\times\mathbb{S}^1} h}.$$

Together with the upper bound on the restriction of the function $|\partial H/\partial t| = s|\partial h/\partial t|$ to S in the hypothesis of part B of the theorem, this yields the following upper bound on the function $|\partial H/\partial t| = s|\partial h/\partial t|$ on the set S:

$$\begin{aligned}
\sup_S s|\partial h/\partial t| &\leq \frac{c_{\max} + E}{\inf_{\Sigma\times\mathbb{S}^1} h} \cdot \sup_{\Sigma\times\mathbb{S}^1} |\partial h/\partial t| \\
&< \frac{c_{\max} + E}{\inf_{\Sigma\times\mathbb{S}^1} h} \cdot \frac{(1-2e)e\Delta_{s_+}^2 \inf_{\Sigma\times\mathbb{S}^1} h}{(s_+ - 1)(s_+ \max_{\Lambda_0\times\mathbb{S}^1} h + e\Delta_{s_+})\widehat{l}_{\min,s_+}} \\
&= \frac{c_{\max} + E}{\inf_{\Sigma\times\mathbb{S}^1} h} \cdot \frac{(1-2e)e\Delta_{s_+}^2 \inf_{\Sigma\times\mathbb{S}^1} h}{(s_+ - 1)(c_{\max} + E)\widehat{l}_{\min,s_+}} \\
&= \frac{(1-2e)e\Delta_{s_+}^2}{(s_+ - 1)\widehat{l}_{\min,s_+}} = \frac{E}{T},
\end{aligned}$$

yielding the bound required in Corollary 4.12. Thus, Corollary 4.12 can be applied to $(N,\Omega) = (\Sigma\times\mathbb{R}_+, d(s\lambda))$ and the Hamiltonian H on it (since, by our assumptions, $\widehat{l}_{\min,s_+} < +\infty$ and $\Delta_{s_+} > 0$), which yields the existence of a chord of h from Λ_0 to Λ_1 of time-length $\leq T = \dfrac{(s_+ - 1)\widehat{l}_{\min,s}}{(1-2e)\Delta_{s_+}}$.

This finishes the proof of part B of the theorem. □

Corollary 5.5. *Assume that h is time-independent and $\inf_\Sigma h > 0$.*

If h is C-cooperative with Λ_0, Λ_1 for some $C > \inf_\Sigma h$ (see Definition 1.8) and the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is weakly homologically bonded, then there exists a chord of h from Λ_0 to Λ_1 of time-length $\leq \inf_{s > C/\inf_\Sigma h} \dfrac{(s-1)l_{\min,s}(\Lambda_0,\Lambda_1,\lambda)}{s\inf_\Sigma h - C}$.

Furthermore, if h is cooperative with Λ_0, Λ_1 and the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is homologically bonded, the time-length of the chord can be also bounded from above by $\mu := l_{\min,\infty}(\Lambda_0,\Lambda_1,\lambda)/\inf_\Sigma h$. In particular, the pair (Λ_0,Λ_1) is μ-interlinked.

Proof of Corollary 5.5. Let us assume that condition (a) from Definition 1.8 of C-cooperativeness is satisfied—that is, $h < C$ on Λ_1 and either the set $\{h \geq C\}$ is empty or $dh(R) \geq 0$ on $\{h \geq C\}$ (the case of condition (b) from the same definition is similar).

Consider a smooth increasing function $\chi : \mathbb{R}_+ \to \mathbb{R}$, such that $\chi(s) = s$ for $s \in [0, C]$ and $\chi(s) = C + \epsilon$ for some $\epsilon > 0$ and all sufficiently large $s \in \mathbb{R}_+$. Consider the time-independent contact Hamiltonian $\widetilde{h} := \chi \circ h$. One readily sees that it is complete and satisfies $0 < \inf_\Sigma \widetilde{h} \leq \sup_\Sigma \widetilde{h} \leq C + \epsilon < +\infty$. Since the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is weakly homologically bonded, part A of Theorem 5.4, applied to $\widetilde{h}$, shows that there exists a chord $\gamma : [0, T] \to \Sigma$ of $\widetilde{h}$, such that

$\gamma(0) \in \Lambda_0$ and $\gamma(T) \in \Lambda_1$ for $T \le T_0 := \inf_s \frac{(s-1) l_{\min,s}(\Lambda_0, \Lambda_1, \lambda)}{s \inf_\Sigma \widetilde{h} - \sup_\Sigma \widetilde{h}}$, where the infimum is taken over all $s > \frac{\sup_\Sigma \widetilde{h}}{\inf_\Sigma \widetilde{h}}$.

We claim that $\gamma([0,T])$ lies in the set $\{\widetilde{h} \geq C\}$. Indeed, for all t

$$d(\widetilde{h} \circ \gamma)/dt = d\widetilde{h}(R) \cdot \widetilde{h} = (\chi' \circ h) \cdot dh(R) \cdot \widetilde{h}.$$

Therefore, if $\widetilde{h}(\gamma(t_0)) > C$ for some $t_0 \in [0,T]$, then $\widetilde{h}(\gamma(t)) \geq C$ for all $t \in [t_0, T]$, in contradiction to $\widetilde{h}(\gamma(T)) = h(\gamma(T)) < C$ (the latter holds, since $\gamma(T) \in \Lambda_1$ and $h < C$ on Λ_1). Thus, $\gamma([0,T])$ lies in $\{\widetilde{h} \leq C\}$ where $\widetilde{h}$ coincides with h, meaning that γ is, in fact, the chord of h of time-length T bounded from above by T_0. Since $\inf_\Sigma h = \inf_\Sigma \widetilde{h}$, $\sup_\Sigma \widetilde{h} \leq C + \epsilon$ and ϵ can be taken arbitrarily small, we get that $T_0 \leq \inf_{s > C/\inf_\Sigma h} \frac{(s-1) l_{\min,s}(\Lambda_0, \Lambda_1, \lambda)}{s \inf_\Sigma h - C}$, yielding the required upper bound on the time-length of the chord.

If the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is homologically bonded, then we can replace in the bound above $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda)$ by $l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda)$, remove the infimum and let s go to $+\infty$. This shows that the time-length of the chord is bounded from above by $l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda)/\inf_\Sigma h$.

This finishes the proof of the corollary. $\square$

Corollary 5.6. *Assume that* $\inf_{\Sigma \times \mathbb{S}^1} h > 0$, *$h$ is cooperative with* Λ_0, Λ_1 *and the pair* $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ *is stably homologically bonded. Denote* $\widehat{l}_{\min,\infty} := \widehat{l}_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) > 0$. *Assume also that for some* $0 < e < 1/2$

$$\sup_{\Sigma \times \mathbb{S}^1} |\partial h / \partial t| < \frac{(1-2e)e\big(\inf_{\Sigma \times \mathbb{S}^1} h\big)^3}{\big(\max_{\Lambda_0 \times \mathbb{S}^1} h + e \inf_{\Sigma \times \mathbb{S}^1} h\big) \widehat{l}_{\min,\infty}}.$$

Then, there exists a chord of h from Λ_0 to Λ_1 of time-length bounded from above by $\frac{\widehat{l}_{\min,\infty}}{(1-2e)\inf_{\Sigma \times \mathbb{S}^1} h}$.

Proof of Corollary 5.6. Let us assume that h is C-cooperative with Λ_0, Λ_1 for some $C > 0$ and condition (a) from Definition 1.8 of C-cooperativeness is satisfied (the case of condition (b) from the same definition is similar). Without loss of generality, assume $C > \inf_{\Sigma \times \mathbb{S}^1} h$.

Similarly to the proof of Corollary 5.5, consider a smooth increasing function $\chi : \mathbb{R}_+ \to \mathbb{R}$, such that $\chi(s) = s$ for $s \in [0, C]$ and $\chi(s) = C + \epsilon$ for some $\epsilon > 0$ and all sufficiently large $s \in \mathbb{R}_+$. Consider the time-dependent contact Hamiltonian $\widetilde{h} := \chi \circ h$. One readily sees that it is complete and satisfies

$$0 < \inf_{\Sigma \times \mathbb{S}^1} h = \inf_{\Sigma \times \mathbb{S}^1} \widetilde{h} \leq \sup_{\Sigma \times \mathbb{S}^1} \widetilde{h} \leq C + \epsilon < +\infty.$$

Note that for any sufficiently large $s_+ > 1$

$$\Delta_{s_+} := s_+ \inf_{\Sigma \times \mathbb{S}^1} \widetilde{h} - \sup_{\Sigma \times \mathbb{S}^1} \widetilde{h} > 0.$$

Also note that

$$\lim_{s_+\to+\infty} \frac{(1-2e)e\Delta_{s_+}^2 \inf_{\Sigma\times\mathbb{S}^1} h}{(s_+-1)(s_+ \max_{\Lambda_0\times\mathbb{S}^1} h + e\Delta_{s_+})} = \frac{(1-2e)e\big(\inf_{\Sigma\times\mathbb{S}^1} h\big)^3}{\max_{\Lambda_0\times\mathbb{S}^1} h + e\inf_{\Sigma\times\mathbb{S}^1} h}$$

and for all $s_+ > 1$

$$\widehat{l}_{\min,s_+}(\Lambda_0,\Lambda_1,\lambda) \le \widehat{l}_{\min,\infty}(\Lambda_0,\Lambda_1,\lambda).$$

Therefore, the upper bound on $\sup_{\Sigma\times\mathbb{S}^1} |\partial h/\partial t|$ in the hypothesis of the corollary implies that for any sufficiently large s_+, we can bound $\sup_{\Sigma\times\mathbb{S}^1} |\partial\widetilde{h}/\partial t|$ from above as in (30):

$$\sup_{\Sigma\times\mathbb{S}^1} |\partial\widetilde{h}/\partial t| < \frac{(1-2e)e\Delta_{s_+}^2 \inf_{\Sigma\times\mathbb{S}^1} h}{(s_+-1)(s_+ \max_{\Lambda_0\times\mathbb{S}^1} h + e\Delta_{s_+})\widehat{l}_{\min,s_+}}.$$

Since the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is stably homologically bonded, part B of Theorem 5.4, applied to $\widetilde{h}$, shows that there exists a chord $\gamma_{s_+} : [t_{s_+}, t_{s_+}+T_{s_+}] \to \Sigma$ of $\widetilde{h}$, for some $t_{s_+} \in \mathbb{R}$, so that $\gamma_{s_+}(t_{s_+}) \in \Lambda_0$ and $\gamma_{s_+}(t_{s_+}+T_{s_+}) \in \Lambda_1$ for $T_{s_+} \le \dfrac{(s_+-1)\widehat{l}_{\min,s_+}}{(1-2e)\Delta_{s_+}}$. Similarly to the proof of Corollary 5.5, we get that γ_{s_+} is in fact a chord of h.

Note that

$$\lim_{s_+\to+\infty} \frac{(s_+-1)\widehat{l}_{\min,s_+}}{(1-2e)\Delta_{s_+}} \le \lim_{s_+\to+\infty} \frac{(s_+-1)\widehat{l}_{\min,\infty}}{(1-2e)\Delta_{s_+}} = \frac{\widehat{l}_{\min,\infty}}{(1-2e)\inf_{\Sigma\times\mathbb{S}^1} h}.$$

Also note that since h is time-periodic with period 1, we can assume that $t_{s_+} \in [0,1]$ for all s_+. Now, since Λ_0 is compact, a standard compactness argument allows to obtain the existence of a chord of h from X_0 to X_1 of time-length $\le \dfrac{(s_+-1)\widehat{l}_{\min,\infty}}{(1-2e)\Delta_{s_+}}$. □

The following corollary yields the existence of a chord of h in case where h is not necessarily everywhere positive.

Corollary 5.7. *Assume there exists a (possibly non-compact or disconnected) closed codimension 0 submanifold $\Xi \subset \Sigma$ with a (possibly non-compact or disconnected) boundary $\partial\Xi$, so that*

1. $\inf_{\Xi\times\mathbb{S}^1} h > 0$ *(but h may be negative outside $\Xi \times \mathbb{S}^1$).*
2. $\sup_{\partial\Xi\times\mathbb{S}^1} h < +\infty$.
3. *For each $t \in \mathbb{S}^1$, the contact Hamiltonian vector field v_t of h is transverse to $\partial\Xi$ (in particular, $\partial\Xi$ is a convex surface in the sense of contact topology—see [37]) and either points inside Ξ everywhere on $\partial\Xi$ or points outside Ξ everywhere on $\partial\Xi$.*
4. *Both Λ_0 and Λ_1 lie in Ξ.*

Then, the following claims hold:

(I) *Assume the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is weakly homologically bonded. Assume also that h is time-independent and C-cooperative with Λ_0, Λ_1 for $C > \inf_\Xi h$.*

Then, there exists a chord of h from Λ_0 to Λ_1 of time-length bounded from above by $\inf_{s>C/\inf_\Xi h} \dfrac{(s-1)l_{\min,s}(\Lambda_0,\Lambda_1,\lambda)}{s\inf_\Xi h - C}$.

If the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is homologically bonded, then the time-length of the chord can be bounded from above by $\dfrac{l_{\min,\infty}(\Lambda_0,\Lambda_1,\lambda)}{\inf_\Xi h}$.

(II) *Assume that the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is stably homologically bonded and set $\widehat{l}_{\min,\infty} := \widehat{l}_{\min,\infty}(\Lambda_0,\Lambda_1,\lambda)$. Assume also that h is time-dependent and cooperative with Λ_0, Λ_1 and for some $0 < e < 1/2$*

$$\sup_{\Xi\times\mathbb{S}^1} |\partial h/\partial t| < \frac{(1-2e)e\big(\inf_{\Xi\times\mathbb{S}^1} h\big)^3}{\big(\max_{\Lambda_0\times\mathbb{S}^1} h + e\inf_{\Xi\times\mathbb{S}^1} h\big)\widehat{l}_{\min,\infty}}.$$

Then, there exists a chord of h from Λ_0 to Λ_1 whose time-length is bounded from above by $\dfrac{\widehat{l}_{\min,\infty}}{(1-2e)\inf_{\Xi\times\mathbb{S}^1} h}$.

Proof of Corollary 5.7. Let us assume that h is C-cooperative with Λ_0, Λ_1 for $C \geq \sup_{\partial\Xi\times\mathbb{S}^1} h$ (this is possible, since, by (2), $\sup_{\partial\Xi\times\mathbb{S}^1} h < +\infty$).

For any sufficiently small $\epsilon > 0$, one can find a new contact Hamiltonian $\widetilde{h}_\epsilon : \Sigma\times\mathbb{S}^1 \to \mathbb{R}$, so that $\widetilde{h} = h$ on a neighborhood of Ξ and $\inf_{\Xi\times\mathbb{S}^1} h - \epsilon \leq \widetilde{h}_\epsilon \leq \sup_{\Xi\times\mathbb{S}^1} h$ on $(\Sigma\backslash\Xi)\times\mathbb{S}^1$. Since $C \geq \sup_{\partial\Xi\times\mathbb{S}^1} h$, the contact Hamiltonian $\widetilde{h}_\epsilon$ satisfies the assumptions of Corollary 5.5 (in case (I)), or of Corollary 5.6 (in case (II)). Consequently, by these corollaries, there exists a chord $\gamma_\epsilon(t)$ of $\widetilde{h}_\epsilon$, $\gamma_\epsilon(t_\epsilon) \in \Lambda_0$, $\gamma_\epsilon(t_\epsilon + T_\epsilon) \in \Lambda_1$, so that

– In case (I):

$$T_\epsilon \leq \inf_s \frac{(s-1)l_{\min,s}(\Lambda_0,\Lambda_1,\lambda)}{s\inf_\Sigma \widetilde{h}_\epsilon - C} \leq \inf_s \frac{(s-1)l_{\min,s}(\Lambda_0,\Lambda_1,\lambda)}{s(\inf_\Xi h - \epsilon) - C}$$

(the infimum is taken over all $s > C/\inf_\Sigma \widetilde{h}_\epsilon$), and

$$T_\epsilon \leq \frac{l_{\min,\infty}(\Lambda_0,\Lambda_1,\lambda)}{\inf_\Sigma \widetilde{h}_\epsilon} \leq \frac{l_{\min,\infty}(\Lambda_0,\Lambda_1,\lambda)}{\inf_\Xi h - \epsilon}.$$

– In case (II):

$$T_\epsilon \leq \frac{\widehat{l}_{\min,\infty}}{(1-2e)\inf_{\Sigma\times\mathbb{S}^1} \widetilde{h}_\epsilon} \leq \frac{\widehat{l}_{\min,\infty}}{(1-2e)\inf_{\Xi\times\mathbb{S}^1} h - \epsilon}.$$

Here, in both cases, we have used that $\inf_\Sigma \widetilde{h}_\epsilon \geq \inf_\Xi h - \epsilon$ for all (sufficiently small) $\epsilon > 0$.

The chord γ_ϵ cannot cross $\partial\Xi$. Indeed, by (3), if it had crossed $\partial\Xi$, it would have had to cross it transversally from Ξ to $M \setminus \Xi$. This would mean that v_t (the contact Hamiltonian vector field of $\widetilde{h}_\epsilon|_\Xi = h|_\Xi$) points outside Ξ everywhere on $\partial\Xi$ for all $t \in \mathbb{S}^1$. But then, the chord would not have been able to return to Ξ to reach Λ_1. Thus, the chord lies in the interior of Ξ and is, in fact, a chord of h from Λ_0 to Λ_1.

We have such a chord γ_ϵ of h from Λ_0 to Λ_1 for any sufficiently small $\epsilon > 0$ and the time-lengths of the chords admit a bound continuous in ϵ.

We can also assume that for all $\epsilon > 0$ $t_{s_\epsilon} = 0$ in case (I) (since h is time-independent) and $t_{s_\epsilon} \in [0,1]$ in case (II) (since h is 1-periodic in time). Now, since Λ_0 is compact, a standard compactness argument allows to obtain the existence of the chord of h from X_0 to X_1 of time-length T, so that
– In case (I):

$$T \leq \inf_{s > C/\inf_{\Xi} h} \frac{(s-1) l_{\min,s}(\Lambda_0, \Lambda_1, \lambda)}{s \inf_{\Xi} h - C}$$

and if the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is homologically bonded, then

$$T \leq \frac{l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda)}{\inf_{\Xi} h}.$$

– In case (II)

$$T \leq \frac{\widehat{l}_{\min,\infty}}{(1-2e)\inf_{\Xi \times \mathbb{S}^1} h}.$$

This finishes the proof of the corollary. □

Remark 5.8. The claim of Theorem 5.4 is robust—for a fixed h—with respect to perturbations of $\Lambda = \Lambda_0 \sqcup \Lambda_1$ by Legendrian isotopies, as long as the perturbation is sufficiently C^1-small, depending on s_+—this follows from Remark 4.8.

Accordingly, the claims of Corollaries 5.5, 5.6, and 5.7 are robust—for a fixed h—with respect to perturbations of $\Lambda = \Lambda_0 \sqcup \Lambda_1$ by Legendrian isotopies, as long as the perturbation is sufficiently C^1-small, depending on $\dfrac{C}{\inf_{\Sigma \times \mathbb{S}^1} h}$, where C is the constant, such that h is C-cooperative with Λ_0, Λ_1.

Namely, if the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is weakly homologically bonded, then $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda) < +\infty$ for any $s > \dfrac{C}{\inf_{\Sigma \times \mathbb{S}^1} h}$. Fix such an s. Then, by Remark 4.8, $l_{\min,s}(\Lambda'_0, \Lambda'_1, \lambda)$ is finite and close to $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda)$ for any $\Lambda' = \Lambda'_0 \sqcup \Lambda'_1$ obtained from Λ by a Legendrian isotopy, as long as the isotopy is small (depending on the chosen s). The proofs of Corollaries 5.5, 5.6, and 5.7 then go through for Λ' instead of Λ and yield the existence of a chord of h between Λ'_0 and Λ'_1. The cases when the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is homologically bonded or stably homologically bonded are similar.

6. The case of J^1Q

In this section, let $\Sigma = J^1Q = T^*Q \times \mathbb{R}(z)$ be the 1-jet space of a smooth manifold Q, together with the standard contact form λ on it. The Reeb flow of λ is the shift in the z-coordinate. This is a nice contact manifold.

Let Λ_0 be the zero section of J^1Q.

Proposition 6.1. *Assume that $\Lambda_1 \subset \Sigma = J^1Q$ is a Legendrian submanifold, such that $V_\infty(\Lambda_1, \lambda) \neq 0$ and satisfying the following property: there is unique Reeb chord starting on Λ_0 and ending on Λ_1, and this chord is non-degenerate in the sense of (8).*

Then, the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ *is homologically bonded.*

Proof. Denote by a the unique Reeb chord going from Λ_0 to Λ_1.

For each $i, j = 0, 1$ write $\mathcal{A}_{ij}$ for the subalgebra of $\mathcal{A}(\Lambda_0 \sqcup \Lambda_1, \lambda)$ generated by all ij-composable monomials. We call a monomial in $\mathcal{A}(\Lambda_0 \sqcup \Lambda_1, \lambda)$ *good* if it is 11-composable and contains no a. Note that each such monomial lies in $\mathcal{A}(\Lambda_1, \lambda)$.

Let $J \in \mathcal{J}(\Lambda_0 \sqcup \Lambda_1)$. Then, J also lies in $\mathcal{J}(\Lambda_1)$. Consequently, J defines both a differential $d := \partial_J$ on $\mathcal{A}(\Lambda_0 \sqcup \Lambda_1, \lambda)$ and a differential d' on $\mathcal{A}(\Lambda_1, \lambda)$.

Note that $da = 0$. Indeed, d should map a into a sum of 01-composable monomials whose actions are smaller than the action of a, but there are no such monomials.

Lemma 6.2. *There is no* $x \in \mathcal{A}_{11}$*, such that* $dx = 1$.

Proof of Lemma 6.2. Since $V_\infty(\Lambda_1, \lambda) \neq 0$, there exists no $x' \in \mathcal{A}(\Lambda_1, \lambda)$, such that $d'x' = 1$. Indeed, otherwise, every d'-closed element y is d'-exact: $d'(yx') = y$.

Now, assume by contradiction that there exists $x \in \mathcal{A}_{11}$, such that $dx = 1$. Then, x is a sum of 11-composable monomials of the either of the following two types: either $b_1ab_2a \ldots b_ka$ (type I) for some $k > 0$, or $b_1ab_2a \ldots b_kac$ (type II) for some $k \geq 0$, where $b_1, \ldots, b_k$ are 10-composable monomials not containing a and c is a good monomial.

Since $da = 0$ and $\mathcal{A}_{10}$ is d-invariant, applying d to a monomial $b_1ab_2a \ldots b_ka$ of type I, we get that $d(b_1ab_2a \ldots b_ka)$ is again a sum of monomials of type I, or zero. Similarly, applying d to a monomial $b_1ab_2a \ldots b_kac$ of type II, we get that $d(b_1ab_2a \ldots b_kac)$ is a sum of monomials of type II, with at least one a in each of them, or zero—unless $k = 0$ and the original monomial of type II to which d is applied is just a good monomial c. Since $dx = 1$, we get that x can be written as $x = x' + x''$, where x' is a non-trivial sum of good monomials and x'' is a sum of monomials (of types I and II) containing a.

Each good monomial c lies in $\mathcal{A}(\Lambda_1, \lambda)$, and therefore, x', which is a sum of good monomials, also lies in $\mathcal{A}(\Lambda_1, \lambda)$. For each good monomial c, we can write $dc = d'c + y$, where y is a sum of 11-composable monomials containing a. At the same time, the discussion above shows that dx'' is a (possibly trivial) sum of 11-composable monomials containing a. Since $dx = 1$ and $d'c$ is a sum of 11-composable monomials that do not contain a and cannot cancel out monomials containing a, we get that $d'x' = 1$, which yields a contradiction.

This finishes the proof of the lemma. □

Let us now finish the proof of the proposition.

Note that there are no 00-chords (since Λ_0 is the zero section). Therefore, any 01-composable monomial has to be of either of the following two types: either $aB_1aB_2 \ldots aB_ka$ (type 1) for some $k > 0$, or $aB_1aB_2 \ldots aB_kaC$ (type 2) for some $k \geq 0$, where $B_1, \ldots, B_k$ are 10-composable monomials containing no a and C is a good monomial.

We claim that there exists no $z \in \mathcal{A}(\Lambda_0, \Lambda_1, \lambda)$, such that $dz = a$—since $da = 0$, this would readily imply the proposition.

Let us prove the claim. Assume, by contradiction, that such a z exists. It is a sum of 01-composable monomials, each of which is either of type 1 or of type 2.

Since $da = 0$, the differential of any monomial of type 1 or 2 is a (possibly zero) sum of monomials of the same type each of which contains *more than one* a, except for the case where the monomial to which d is applied is a monomial of type 2 of the form aC, where C is a good monomial—then $d(aC) = a \cdot dC$. Thus, the only way dz can contain a monomial a is that z is a sum of monomials one of which is of the form aC, where C is a good monomial, such that $dC = 1$. This leads to a contradiction with Lemma 6.2.

This finishes the proof of the claim and of the proposition. □

Further on in this section, assume that $l > 0$ and Λ_1 is the image of Λ_0 under the time-l Reeb flow.

Proposition 6.3. *The pair* $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ *is homologically bonded and*

$$l_{\min,s}(\Lambda_0, \Lambda_1, \lambda) \leq l$$

for all $s \in (1, +\infty]$.

Proof of Proposition 6.3. The pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is degenerate and we have to perturb, say, Λ_1 to make it non-degenerate.

Namely, consider a Morse function $f : Q \to \mathbb{R}$ which is a C^∞-small perturbation of the constant function z_+. Its 1-jet is a Legendrian submanifold $\Lambda_1' \subset J^1Q$ which is a small perturbation of Λ_1. It is not hard to check that the pair $(\Lambda_0 \sqcup \Lambda_1', \lambda)$ is non-degenerate: the Reeb chords of $\Lambda = \Lambda_0 \sqcup \Lambda_1'$ are then the Reeb chords from Λ_0 to Λ_1' corresponding to the critical points of f and their actions are the critical values of f shifted down by z_+. The 01-subspace is then the $\mathbb{Z}_2$-span of the Reeb chords of Λ.

One can show that, under the identification between the Reeb chords of Λ and the critical points of f, the differential in the Chekanov–Eliashberg algebra is identified with the Morse differential in the Morse chain complex of f (over $\mathbb{Z}_2$). More precisely, the differential in the Chekanov–Eliashberg algebra is identified with the differential in the Lagrangian Floer complex of the projections of Λ_0 and Λ_1' to T^*Q (that are *embedded* Lagrangian submanifolds) and the latter is identified with the Morse differential by the original work of Floer [33]—see, e.g., the comment preceding Cor. 1.10 in [9].

Thus, the persistence module associated with $(\Lambda_0, \Lambda_1', \lambda)$ is the Morse homology persistence module associated with f. The corresponding barcode contains infinite bars—their number is equal to the sum of the Betti numbers of Q over $\mathbb{Z}_2$. Hence, $l_{\min,s}(\Lambda_0, \Lambda_1', \lambda) \leq \max_Q$ for all $s \in (1, +\infty]$. Letting f converge uniformly to the constant function l, we readily get that $l_{\min,s}(\Lambda_0, \Lambda_1, \lambda) \leq l$ for all $s \in (1, +\infty]$, meaning that the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is homologically bonded. □

Proof of Theorem 1.5. The Legendrian submanifolds appearing in the formulation of the theorem are homologically bonded. For item (i) of the theorem, this follows from Proposition 6.3 combined with the fact that the property of being homologically bonded is invariant under a Legendrian isotopy of a pair.

For item (ii), this follows from Proposition 6.1. Therefore, the pair (Λ_0, Λ_1) is interlinked by Corollary 5.5. $\square$

Proposition 6.4. *The pair* $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ *is stably homologically bonded and* $\widehat{l}_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) \leq l$.

Proof of Proposition 6.4. There is a natural identification $\widehat{\Sigma} = J^1Q \times T^*\mathbb{S}^1 = J^1(Q \times \mathbb{S}^1)$ identifying the contact forms, and hence, the contact structures. The Legendrian submanifold $\widehat{\Lambda}_0$ is then the zero section of $J^1(Q \times \mathbb{S}^1)$ and $\widehat{\Lambda}_1$ is its image under the time-l Reeb flow. Now, the result follows from Proposition 6.3. $\square$

Let ψ be a positive function on Λ_0, and let $\Lambda := \{z = \psi(q), p = \psi'(q)\}$ be the graph of its 1-jet in J^1Q. Assume that K is a Legendrian submanifold of J^1Q Legendrian isotopic to Λ outside the zero section Λ_0, so that there exist exactly two Reeb chords A, a starting on K and ending on Λ_0, both non-degenerate, and with their time-lengths $|A|, |a|$ satisfying

$$0 < |A| - |a| < |b|, \tag{31}$$

for every Reeb chord b starting and ending on $K \sqcup \Lambda_0$.

Proposition 6.5. *For any* $s < |A|/|a|$,

$$l_{\min,s}(K, \Lambda_0, \lambda) = |a|. \tag{32}$$

Proof. For each $i, j = 0, 1$ write $\mathcal{A}_{ij}$ for the subalgebra of $\mathcal{A}(K \sqcup \Lambda_0, \lambda)$ generated by all ij-composable monomials. **Warning:** with our notation 01-chords are the ones going from K to Λ_0 etc. Observe that

$$\mathcal{A}_{01} = \mathrm{Span}\{ua, uA, u \in \mathcal{A}_{00}\}, \mathcal{A}_{11} = 0.$$

By the invariance of V_∞ under Legendrian isotopies of the two-part Legendrians and by Proposition 6.3, we get

$$V_\infty(K, \Lambda_0, \lambda) = V_\infty(\Lambda, \Lambda_0, \lambda) = 0,$$
$$V_\infty(\Lambda_0, K, \lambda) = V_\infty(\Lambda_0, \Lambda, \lambda) = V_\infty(\Lambda_0, \Lambda_1, \lambda) \neq 0,$$

where Λ_1 is defined before Proposition 6.3. Let $J \in \mathcal{J}(K \sqcup \Lambda_0)$ and let $d := \partial_J$ be the corresponding differential on $\mathcal{A}(K \sqcup \Lambda_0, \lambda)$.

Note that $da = 0$ as there is no 01-composable monomial with a smaller action. In what follows we write $|b|$ for the action of a monomial b.

Since there are no 11-chords, we can write $dA = ua + vA$ with $u, v \in \mathcal{A}_{00}$. Note that $v = 0$, since otherwise $|vA| \geq |A|$, while d lowers the action. By assumption (31), $|ua| > |A|$ for a non-scalar u, and hence, either $u = 0$ or $u = 1$ (recall that the base field is $\mathbb{Z}_2$).

Case 1 $u = 0$, i.e., $dA = 0$. We claim that in this case, $a \neq dx$ for any x. Indeed, otherwise, write $x = pa + qA$, where $p, q \in \mathcal{A}_{00}$. Then, $a = (dp)a + (dq)A$ yielding $dp = 1$. But then, for every closed $y \in \mathcal{A}_{10}$, we have $y = d(yp)$, meaning that y is exact, in contradiction to $V_\infty(\Lambda_0, K, \lambda) \neq 0$.

Case 2 $u = 1$, i.e. $dA = a$. Note that by assumption (31), A has the minimal action among all 01-monomials of action $> |a|$. It follows that the barcode of the persistence module $V(K, \Lambda_0, \lambda)$ contains a bar $(|a|, |A|]$. Moreover, since $|a|$ is the minimal action among all 01-chords, we get that (32). This completes the proof.

Proof of Theorem 1.7. Using if necessary small perturbations of the Legendrians, we may assume without loss of generality that all the chords are non-degenerate. The theorem immediately follows from Proposition 6.5 and Theorem 5.4A. □

7. The case of $ST^*\mathbb{R}^n$

In this section, let

$$\Sigma := ST^*\mathbb{R}^n, \; n > 1.$$

Denote by ξ the standard contact structure on Σ defined by the contact form

$$\lambda := pdq.$$

Let $x_0, x_1 \in \mathbb{R}^n(q)$, $x_0 \neq x_1$. Consider the Legendrian submanifolds

$$\Lambda_i := \{q = x_i, \; |p| = 1\}, \; i = 0, 1,$$

of (Σ, ξ). Set

$$\Lambda := \Lambda_0 \sqcup \Lambda_1.$$

One easily checks that the pair $(\Lambda = \Lambda_0 \sqcup \Lambda_1, \lambda)$ is non-degenerate.

Consider also the manifold

$$\widehat{\Sigma} := \Sigma \times T^*\mathbb{S}^1(r, \tau),$$

and the 1-form

$$\widehat{\lambda} := pdq - rd\tau,$$

where $r \in \mathbb{R}$, $\tau \in \mathbb{S}^1$, are the standard coordinates on $T^*\mathbb{S}^1 = \mathbb{R} \times \mathbb{S}^1$. Let $\mathbb{S}^1 := \{r = 0\} \subset T^*\mathbb{S}^1$ be the zero section. One easily sees that $\widehat{\lambda}$ is a contact form, defining a contact structure $\hat{\xi}$ on $\widehat{\Sigma}$ and the Reeb vector field of $\widehat{\lambda}$ can be described as follows: its projection to the $ST^*\mathbb{R}^n$ factor is the Reeb vector field of λ, while its projection to the $T^*\mathbb{S}^1$ factor is zero. Denote

$$\widehat{\Lambda}_0 := \Lambda_0 \times \mathbb{S}^1, \; \widehat{\Lambda}_1 := \Lambda_1 \times \mathbb{S}^1,$$
$$\widehat{\Lambda} := \widehat{\Lambda}_0 \sqcup \widehat{\Lambda}_1 = \Lambda \times \mathbb{S}^1.$$

For each $\delta > 0$, denote

$$\widehat{\Lambda}_{1,\delta} := \Lambda_1 \times \mathit{graph}\,(\delta df) \subset \widehat{\Sigma},$$
$$\widehat{\Lambda}_\delta := \widehat{\Lambda}_0 \sqcup \widehat{\Lambda}_{1,\delta},$$

where $\mathit{graph}\,(\delta df) \subset T^*\mathbb{S}^1$ is the graph of $d(\delta f)$ for a Morse function $f : \mathbb{S}^1 \to \mathbb{R}$ that has only two critical points. The set $\widehat{\Lambda}_\delta$ is a two-part Legendrian submanifold of $(\Sigma, \ker \widehat{\lambda})$. Since the pair (Λ, λ) is non-degenerate, so is the pair $(\widehat{\Lambda}_\delta, \widehat{\lambda})$.

Proposition 7.1. *The barcode of the persistence module $V(\Lambda_0, \Lambda_1, \lambda)$ consists of infinitely many infinite bars, of multiplicity* 1*, whose left ends are* $(2k-1)|x_0 - x_1|$, $k \in \mathbb{Z}_{>0}$.

Consequently, the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is homologically bonded and

$$l_{\min,s}(\Lambda_0, \Lambda_1, \lambda) = l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) = |x_0 - x_1|$$

for all $s \in (1, +\infty]$.

Proof of Proposition 7.1. The set $\mathcal{R}(\Lambda, \lambda)$ consists only of two Reeb chords: a 01-chord a and a 10-chord b, which are the lifts of the Euclidean geodesics in $\mathbb{R}^n$ going from x_0 to x_1 and from x_1 to x_0. Thus, $\mathcal{A}(\Lambda_0, \Lambda_1, \lambda)$ is spanned over $\mathbb{Z}_2$ by the monomials of the form $abab \cdot \ldots \cdot ba$, where each monomial contains k factors a and $k-1$ factors b, for $k \in \mathbb{Z}_{>0}$. For each $k \in \mathbb{Z}_{>0}$, there is exactly one such monomial of length $2k-1$. The actions of a and b are both equal to $|x_0 - x_1|$, and therefore, the action of the monomial of length $2k-1$ is $(2k-1)|x_0 - x_1|$.

Let $J \in \mathcal{J}(\Lambda)$. Since the differential ∂_J on $\mathcal{A}(\Lambda_0, \Lambda_1, \lambda)$ lowers the action, we immediately get that $\partial_J(a) = \partial_J(b) = 0$ and therefore ∂_J is identically zero on $\mathcal{A}(\Lambda_0, \Lambda_1, \lambda)$. Therefore, the barcode of the persistence module $V(\Lambda_0, \Lambda_1, \lambda)$ consists of infinitely many infinite bars of multiplicity 1 whose left ends are $(2k-1)|x_0 - x_1|$, $k \in \mathbb{Z}_{>0}$. This finishes the proof. □

Proposition 7.2. *The barcode of the persistence module $V(\widehat{\Lambda}_0, \widehat{\Lambda}_{1,\delta}, \widehat{\lambda})$ consists of infinitely many infinite bars whose left ends are $(2k-1)|x_0 - x_1|$, $k \in \mathbb{Z}_{>0}$. The multiplicity of the bar with the left end $(2k-1)|x_0 - x_1|$ is 2^{2k-1}.*

Consequently, the pair $(\Lambda_0 \sqcup \Lambda_1, \lambda)$ is stably homologically bonded and

$$\widehat{l}_{\min,s}(\Lambda_0, \Lambda_1, \lambda) = \widehat{l}_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) = |x_0 - x_1|$$

for all $s \in (1, +\infty]$.

Proof of Proposition 7.2. Let $a, b \in \mathcal{R}(\Lambda, \lambda)$ be the 01-chord and the 10-chord of Λ as above. The Reeb chords of $\widehat{\Lambda}_\delta$ are the direct products of the Reeb chords of $\mathcal{R}(\Lambda, \lambda)$ lying in Σ and the constant paths in $T^*\mathbb{S}^1$ corresponding to the 2 critical points of f (that is, the intersections of the graph of df with the zero section). Thus, $\mathcal{R}(\widehat{\Lambda}_\delta, \widehat{\lambda})$ consists only of exactly 4 Reeb chords: 01-chords a_1, a_2 and 10-chords b_1, b_2, where a_1, a_2 project onto a and b_1, b_2 project onto b under the projection $\widehat{\Sigma} = \Sigma \times T^*\mathbb{S}^1 \to \Sigma$.

Thus, $\mathcal{A}(\widehat{\Lambda}_0, \widehat{\Lambda}_{1,\delta}, \lambda)$ is spanned over $\mathbb{Z}_2$ by the monomials of the form $a_{j_1} b_{j_2} \cdot \ldots \cdot b_{j_{2k-2}} a_{j_{2k-1}}$, where each monomial contains k factors a_{j_i}, $j_i = 1, 2$, and $k-1$ factors b_{j_i}, $j_i = 1, 2$, for $k \in \mathbb{Z}_{>0}$. There are 2^{2k-1} such monomials of length $2k-1$. The actions of all a_j and b_j, $j = 1, 2$, are equal to $|x_0 - x_1|$, and therefore, the action of a monomial as above is $(2k-1)|x_0 - x_1|$.

Let $J \in \mathcal{J}(\widehat{\Lambda}_\delta)$. Since the differential ∂_J on $\mathcal{A}(\widehat{\Lambda}_0, \widehat{\Lambda}_{1,\delta}, \widehat{\lambda})$ lowers the action, we immediately get that $\partial_J(a_j) = \partial_J(b_j) = 0$ for all $j = 1, 2$, and therefore, ∂_J is identically zero on $\mathcal{A}(\widehat{\Lambda}_0, \widehat{\Lambda}_\delta, \widehat{\lambda})$. Therefore, the barcode of the persistence module $V(\widehat{\Lambda}_0, \widehat{\Lambda}_{1,\delta} \widehat{\lambda})$ consists of infinitely many infinite bars whose left ends are $(2k-1)|x_0 - x_1|$, $k \in \mathbb{Z}_{>0}$. The multiplicity of the bar with the left end $(2k-1)|x_0 - x_1|$ is 2^{2k-1}.

Consequently, for any $s \in (1, +\infty]$ and $\delta > 0$

$$l_{\min,s}(\widehat{\Lambda}_0, \widehat{\Lambda}_{1,\delta}, \widehat{\lambda}) = |x_0 - x_1|$$

and

$$\widehat{l}_{\min,s}(\Lambda_0, \Lambda_1, \lambda) = \liminf_{\delta \to 0} l_{\min,s}(\widehat{\Lambda}_0, \widehat{\Lambda}_{1,\delta}, \widehat{\lambda}) = |x_0 - x_1|.$$

This finishes the proof. □

Proof of Theorem 1.1. In the case $n = 1$, the claim follows from the results in [28]. Namely, in this case, the sets X_0, X_1, Y_0, Y_1 form a Lagrangian tetragon in $(\mathbb{R}^2(p, q), dp \wedge dq)$ built from the point $x_0 \in \mathbb{R}$ for $T = x_1 - x_0$ (see [28, Sec. 5.1]). In the terminology of [28], this Lagrangian tetragon is stably κ-interlinked, for $\kappa = |x_0 - x_1|(s_+ - s_-)$—this follows, e.g., from Cor. 5.3 and Thm. 5.8 in [28] (see [8, Thm. 1.20, Prop.1.21] for a different approach to the proof). By the definition of a κ-interlinked Lagrangian tetragon (see [28, Sec. 1.2]), this yields the dynamical claims of Theorem 1.1 in the case $n = 1$.

Assume now that $n > 1$.

By Propositions 7.1 and 7.2

$$l_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda) = \widehat{l}_{\min,s_+/s_-}(\Lambda_0, \Lambda_1, \lambda) = |x_0 - x_1|.$$

The symplectization $\Sigma \times \mathbb{R}_+(s)$ is identified symplectically with $\mathbb{R}^{2n}(p, q) \setminus \{p = 0\}$ by the map $(p, q, s) \mapsto (sp, q)$. Thus, $(\Sigma \times [s_-, s_+], d(s\lambda))$ can be viewed as a codimension-zero submanifold with boundary of $(N = \mathbb{R}^{2n}, \Omega = dp \wedge dq)$.

Now, the claims of the theorem follow from Corollaries 4.11, 4.12 applied to the Hamiltonian H on (N, Ω).

This finishes the proof of the theorem. □

Proofs of Theorems 1.13, 1.9 and Corollary 1.10. We need to prove Theorems 1.13, 1.9 and Corollary 1.10 in the following three cases:

(I) Λ_0 is the zero section of J^1Q and Λ_1 is its image under the time-l Reeb flow.

(II) Λ_0 is a cotangent unit sphere in $ST^*\mathbb{R}^n$ and Λ_1 is its image under the time-l Reeb flow.

(III) $\Lambda_0, \Lambda_1 \subset ST^*\mathbb{R}^n$ are the cotangent unit spheres at $x_0, x_1 \in \mathbb{R}^n$, $|x_0 - x_1| = l$.

The needed results in cases (I), (II), (III) follow from the corresponding general results in Theorem 5.2, Corollaries 5.5, 5.6 and Corollary 5.7, as soon as we check that in all the three cases $l_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) \geq l$, $\widehat{l}_{\min,\infty}(\Lambda_0, \Lambda_1, \lambda) \geq l$.

In case (I), these inequalities are proved in Propositions 6.3, 6.4.

In the case (II), they also follow from Propositions 6.3, 6.4, because the cases (I) and (II) can be identified by a contactomorphism preserving the contact forms—namely, the contact identification of $ST^*\mathbb{R}^n$ and $J^1\mathbb{S}^{n-1}$ (see (3)) can be easily adjusted, so that $\Lambda_0 \subset ST^*\mathbb{R}^n$ is identified with the zero section of $J^1\mathbb{S}^{n-1}$.

In case (III) the inequalities follow from Propositions 7.1, 7.2. □

Acknowledgements

We thank F. Bourgeois, B. Chantraine, G. Dimitroglou Rizell, T. Ekholm, Y. Eliashberg, E. Giroux, and M. Sullivan for useful discussions. We also thank A. Fathi for communicating to us his unpublished note [32] containing Theorem 2.1 and its proof. We thank T. Melistas, J. Zhang, and the anonymous referee for numerous corrections, and G. Dimitroglou Rizell for finding a mistake in the original version of Theorem 1.5(ii).

References

[1] Arnold, V.I.: Mathematical Methods of Classical Mechanics, 2nd edn. Springer, New York (1989)

[2] Arnold, V.: Topological Invariants of Plane Curves and Caustics. AMS, Providence (1994)

[3] Audin, M., Lalonde, F., Polterovich, L.: Symplectic rigidity: Lagrangian submanifolds. In: Holomorphic Curves in Symplectic Geometry, 271–321, Progr. Math., vol. 117. Birkhäuser, Basel (1994)

[4] Auslander, M.: Representation theory of Artin algebras. II. Commun. Algebr. **1**, 269–310 (1974)

[5] Azumaya, G.: Corrections and supplementaries to my paper concerning Krull–Remak–Schmidt's theorem. Nagoya Math. J. **1**, 117–124 (1950)

[6] Barannikov, S.A.: The framed Morse complex and its invariants. In: Singularities and Bifurcations, pp. 93–115, Adv. Soviet Math., vol. 21. AMS, Providence (1994)

[7] Bourgeois, F., Eliashberg, Y., Hofer, H., Wysocki, K., Zehnder, E.: Compactness results in symplectic field theory. Geom. Topol. **7**, 799–888 (2003)

[8] Buhovsky, L., Entov, M., Polterovich, L.: Poisson brackets and symplectic invariants. Selecta Math. **18**, 89–157 (2012)

[9] Chantraine, B., Colin, V., Dimitroglou Rizell, G.: Positive Legendrian isotopies and Floer theory. Ann. Inst. Fourier (Grenoble) **69**, 1679–1737 (2019)

[10] Chazal, F., de Silva, V., Glisse, M., Oudot, S.: The Structure and Stability of Persistence Modules. Springer, New York (2016)

[11] Chekanov, Y.: Differential algebra of Legendrian links. Invent. Math. **150**, 441–483 (2002)

[12] Chernov, V., Nemirovski, S.: Non-negative Legendrian isotopy in ST^*M. Geom. Topol. **14**, 611–626 (2010)

[13] Colin, V., Ferrand, E., Pushkar, P.: Positive isotopies of Legendrian submanifolds and applications. Int. Math. Res. Not., pp. 6231–6254 (2017)

[14] Crawley-Boevey, W.: Decomposition of pointwise finite-dimensional persistence modules. J. Algebra Appl. **14**, 1550066 (2015)

[15] Dimitroglou Rizell, G.: Lifting pseudo-holomorphic polygons to the symplectisation of $P \times \mathbb{R}$ and applications. Quantum Topol. **7**, 29–105 (2016)

[16] Dimitroglou Rizell, G.: Families of Legendrians and Lagrangians with unbounded spectral norm, preprint. arXiv:2012.15559 (2020)

[17] Dimitroglou Rizell, G., Sullivan, M.: An energy-capacity inequality for Legendrian submanifolds. J. Topol. Anal. **12**, 547–623 (2020)

[18] Dimitroglou Rizell, G., Sullivan, M.: The persistence of the Chekanov–Eliashberg algebra. Selecta Math. **26**, 1–32 (2020)

[19] Edelsbrunner, H., Harer, J.: Computational Topology: An Introduction. AMS, Providence (2010)

[20] Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. In: Proceedings of the 41st Annual Symposium on Foundations of Computer Science. IEEE, Redondo Beach, pp. 454–463 (2000)

[21] Ekholm, T., Etnyre, J., Sullivan, M.: Legendrian contact homology in $P \times \mathbb{R}$. Trans. AMS **359**, 3301–3335 (2007)

[22] Ekholm, T.: Rational symplectic field theory over $\mathbb{Z}_2$ for exact Lagrangian cobordisms. J. Eur. Math. Soc. **10**, 641–704 (2008)

[23] Ekholm, T., Honda, K., Kálmán, T.: Legendrian knots and exact Lagrangian cobordisms. J. Eur. Math. Soc. **18**, 2627–2689 (2016)

[24] Ekholm, T., Kálmán, T.: Isotopies of Legendrian 1-knots and Legendrian 2-tori. J. Symplectic Geom. **6**, 407–460 (2008)

[25] Eliashberg, Y., Gromov, M.: Convex symplectic manifolds. In: Several complex variables and complex geometry, Part 2 (Santa Cruz, CA, 1989), 135–162, Proc. Sympos. Pure Math., vol. 52, Part 2. AMS, Providence (1991)

[26] Eliashberg, Y.: Invariants in contact topology. In: Proc. Int. Congress Math., vol. II (Berlin, 1998), Extra vol. II, pp. 327–338, Doc. Math. (1998)

[27] Eliashberg, Y., Givental, A., Hofer, H., Introduction to symplectic field theory. In: GAFA 2000 (Tel Aviv, 1999). Geom. and Funct. Analysis 2000. Special Volume, Part II, 560–673 (2000)

[28] Entov, M., Polterovich, L.: Lagrangian tetragons and instabilities in Hamiltonian dynamics. Nonlinearity **30**, 13–34 (2017)

[29] Entov, M., Polterovich, L.: Contact topology and non-equilibrium thermodynamics (2021) **(preprint)**. arXiv:2101.03770

[30] Entov, M., Polterovich, L.: Filtered relative symplectic field theory **(in preparation)**

[31] Fathi, A.: An Urysohn-type theorem under a dynamical constraint. J. Mod. Dyn. **10**, 331–338 (2016)

[32] Fathi, A.: An Urysohn-type theorem under a dynamical constraint: non-compact case (2019) **(preprint)**

[33] Floer, A.: Morse theory for Lagrangian intersections. J. Differ. Geom. **28**, 513–547 (1988)

[34] Gabriel, P.: Unzerlegbare Darstellungen, I. Manuscr. Math. **6**, 71–103 (1972)

[35] Ganor, Y.: A homotopical viewpoint at the Poisson bracket invariants for tuples of sets. J. Symplectic Geom. **18**, 995–1026 (2020)

[36] Geiges, H.: An Introduction to Contact Topology. Cambridge University Press, Cambridge (2008)

[37] Giroux, E.: Convexité en topologie de contact. Comment. Math. Helv. **66**, 637–677 (1991)

[38] Moser, J.: On the volume elements on a manifold. Trans. AMS **120**, 288–294 (1965)

[39] Oudot, S.Y.: Persistence Theory: From Quiver Representations to Data Analysis. AMS, Providence (2015)

[40] Pardon, J.: Contact homology and virtual fundamental cycles. J. Am. Math. Soc. **32**, 825–919 (2019)

[41] Polterovich, L., Shelukhin, E.: Autonomous Hamiltonian flows, Hofer's geometry and persistence modules. Selecta Math. **22**, 227–296 (2016)

[42] Polterovich, L., Rosen, D., Samvelyan, K., Zhang, J.: Topological Persistence in Geometry and Analysis. AMS, Providence (2020)

[43] Ringel, C.M., Tachikawa, H.: QF-3 rings. J. Reine Angew. Math. **272**, 49–72 (1974)

[44] Webb, C.: Decomposition of graded modules. Proc. AMS **94**, 565–571 (1985)

[45] Zomorodian, A., Carlsson, G.: Computing persistent homology. Discrete Comput. Geom. **33**, 249–274 (2005)

Michael Entov
Department of Mathematics
Technion-Israel Institute of Technology
Haifa
Israel 32000
e-mail: `entov@technion.ac.il`

Leonid Polterovich
School of Mathematical Sciences
Tel Aviv University
Tel Aviv
Israel 69978
e-mail: `polterov@tauex.tau.ac.il`

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:47
https://doi.org/10.1007/s11784-022-00945-w
Published online June 4, 2022

Journal of Fixed Point Theory and Applications

Check for updates

A Lagrangian Klein bottle you can't squeeze

Jonathan David Evans

Dedicated to Claude Viterbo, on his fifty-tenth birthday.

Abstract. Suppose you have a nonorientable Lagrangian surface L in a symplectic 4-manifold. How far can you deform the symplectic form before the smooth isotopy class of L contains no Lagrangians? I solve this question for a particular Lagrangian Klein bottle. I also discuss some related conjectures.

Mathematics Subject Classification. 53D12, 53D35, 53D42.

Keywords. Lagrangian submanifold, symplectic geometry, Klein bottle, genus, pseudoholomorphic curve.

1. Introduction

Here are two overlapping questions:

Question 1.1. (Minimal nonorientable genus) *Given a symplectic 4-manifold (X, ω) and a $\mathbb{Z}/2$-homology class $\beta \in H_2(X; \mathbb{Z}/2)$, what is the minimal nonorientable genus of a nonorientable Lagrangian surface $L \subset X$ with $[L] = \beta$?*

Question 1.2. (Nonsqueezing) *Given a symplectic 4-manifold (X, ω) and a nonorientable Lagrangian surface $L \subset X$, how far can you deform ω in cohomology before there is no Lagrangian smoothly isotopic to L?*

If L is orientable then these questions are less interesting: the genus is determined by $[L]^2 = -\chi(L)$ and, in Question 1.2, it is necessary to deform ω subject to the cohomological condition $\int_L [\omega] = 0$. By contrast, if L is nonorientable, we have $H^2(L; \mathbb{R}) = 0$, which means that it is possible to deform ω, keeping L Lagrangian, in such a way that $[\omega]$ ranges over an open set in $H^2(X; \mathbb{R})$.

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

I will give some general discussion of these questions in turn, then give a concrete example of a Lagrangian Klein bottle for which Question 1.2 can be answered completely (Theorem 3.1).

One running theme throughout the discussion is the use of *visible* and *tropical* Lagrangians in almost toric 4-manifolds: these provide a rich source of Lagrangian submanifolds coming respectively from straight lines and tropical curves in integral affine surfaces. I have found them useful for thinking about some of the phenomena under discussion, and for formulating conjectures. Visible Lagrangians were introduced in Symington's work [16]; tropical Lagrangians were introduced independently by Mikhalkin [9] and Matessi [8].

2. The minimal genus question

2.1. Review

Definition 2.1. Define the *nonorientable genus* of the nonorientable surface $\#_k\mathbf{RP}^2$ to be k. Proposition 1.1 of [4] shows that any $\mathbb{Z}/2$-homology class in a symplectic 4-manifold can be represented by some embedded nonorientable Lagrangian, so Question 1.1 has a well-defined answer, which I will denote[1] by $\mathfrak{y}(X,\omega,\beta)$.

Remark 2.2. Audin [1] showed that

$$P_2(\beta)=\chi(L)=2-k \mod 4,$$

where P_2 denotes the Pontryagin square operation and χ is the Euler characteristic. If you find a Lagrangian with nonorientable genus k then you can perform a Hamiltonian finger move locally to introduce pairs of intersections with index difference 1 and then perform Polterovich surgery [13] on these self-intersections to get an embedded Lagrangian with nonorientable genus $k+4$. This means that the set of genera which can be realised is $\{\mathfrak{y}(X,\omega,\beta),\mathfrak{y}(X,\omega,\beta)+4,\ldots\}$.

Remark 2.3. The quantity $\mathfrak{y}(X,\omega,\beta)$ is known in a small range of cases, the lower bound being the principal difficulty.

1. When X satisfies $[\omega]\cdot c_1(X)>0$, we know that $\mathfrak{y}(X,\omega,0)=6$. This follows from Givental's construction [5] of a Lagrangian $\#_6\mathbf{RP}^2$ in the 4-ball and from the fact, proved by Shevchishin [14] that X contains no nullhomologous Lagrangian Klein bottles (see also the beautiful papers by Nemirovski [11,12]).
2. Let $X_{a,b,c}$ be the blow-up of the 4-ball in three subballs so that the symplectic areas of the exceptional spheres E_1,E_2,E_3 are a,b,c. Shevchishin and Smirnov [15] show that $E_1+E_2+E_3$ contains a Lagrangian $\mathbf{RP}^2$ if and only if the following inequalities all hold

   $$a<b+c,\quad b<c+a,\quad c<a+b.$$

   They call these the *symplectic triangle inequalities*. This gives the lower bound $\mathfrak{y}(X_{a,b,c},\omega,E_1+E_2+E_3)\geq 5$ when a,b,c violate the triangle inequalities.

[1] "ng" is the International Phonetic Alphabet symbol for the "ng" sound.

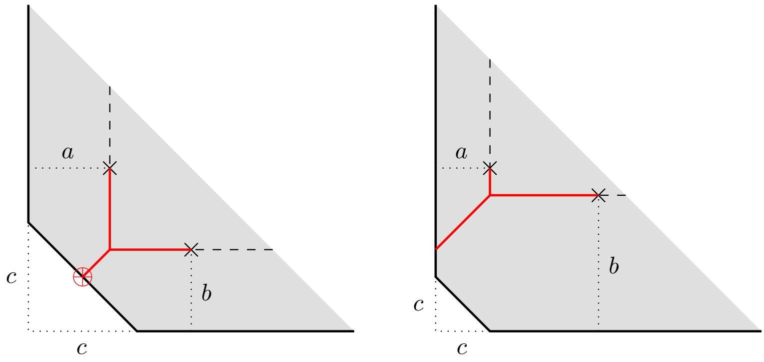


Figure 1. Almost toric base diagrams for $X_{a,b,c}$ with a tropical curve in red. Left: The symplectic triangle inequalities and the associated tropical Lagrangian is diffeomorphic to $\mathbf{RP}^2$ (with the core circle of a cross-cap living over the point marked by the cross-hair symbol). Right: The symplectic triangle inequalities are violated and the associated tropical Lagrangian is diffeomorphic to a disc

Remark 2.4. After the fact, we see that there is a *tropical* or *almost toric* motivation for the Shevchishin-Smirnov triangle inequalities. The almost toric base diagram in Fig. 1 depicts the blow-up $X_{a,b,c}$; the affine lengths a, b, c indicated correspond to the sizes of the exceptional spheres E_1, E_2, E_3. In red you can see a tropical curve; using the ideas of Mikhalkin [9] and Matessi [8], we can construct a Lagrangian submanifold L living over a (small thickening of a) tropical curve. This tropical Lagrangian is diffeomorphic to $\mathbf{RP}^2$ if and only if the inequalities all hold: the preimage of the point marked with cross-hairs is a circle in L whose neighbourhood is a Möbius strip.

2.2. $S^2 \times S^2$

Let $X = S^2 \times S^2$. Modulo an overall scale factor, any symplectic form on X is diffeomorphic to one from the family $\lambda p_1^*\sigma + p_2^*\sigma$, where $p_1, p_2 \colon X \to S^2$ are the two projections and σ is an area form on S^2. We know that $\mathfrak{y}(X, \omega, 0) = 6$, which leaves two interesting $\mathbb{Z}/2$-homology classes up to diffeomorphism: $\beta = [\star \times S^2]$ and the class Δ of the diagonal. The Pontryagin squares are $P_2(\beta) = 0$ and $P_2(\Delta) = 2$, so there is a chance to represent β by Lagrangian Klein bottles.

Lemma 2.5. *If $\lambda < 2$ then β is represented by a Lagrangian Klein bottle.*

Proof. The rectangle in Fig. 2 is the toric moment polygon for the standard Hamiltonian torus action on $S^2 \times S^2$ with symplectic form ω_λ. There is a Lagrangian Klein bottle living over the line ℓ (slope 1/2) in the diagram. To see this, consider the two S^2 factors sitting inside $\mathbb{R}^3$ and let (p_j, θ_j) be cylindrical coordinates on the jth factor $(j = 1, 2)$. These are action-angle coordinates, so $\omega = \sum dp_j \wedge d\theta_j$. The line ℓ is given by $2p_2 = p_1$ and the

Lagrangian Klein bottle is cut out by this equation together with $\theta_2 = -2\theta_1$. This is certainly Lagrangian for this symplectic form. To see that L is a Klein bottle, notice that the regular level sets of p_1 restricted to L are circles $\theta_2 = -2\theta_1$ in the (θ_1, θ_2)-torus, which collapse 2-to-1 onto the circles of maxima and minima at $p_1 = \pm\lambda$ (as the torus collapses to the circle with coordinate θ_2). The projections of these circles are denoted with cross-hairs in Fig. 2. □

Remark 2.6. This L is a *visible Lagrangian* in the sense of Symington [16] as well as being a tropical Lagrangian in the sense of Matessi [8] and Mikhalkin [9]. This Klein bottle is well-known: it appears in [3] as a Hamiltonian minimal Lagrangian, in [6] as a Hamiltonian suspension, and in [4] as a fibre connect-sum of $\mathbf{RP}^2$s. It has minimal Maslov number 1 and has a monotone representative in its Lagrangian isotopy class if $\lambda = 1$.

If $\lambda \geq 2$ then the line ℓ does not fit into the rectangle. The following conjecture seems natural; while I cannot prove it, it inspired Theorem 3.1 below.

Conjecture 2.7. *There is no Lagrangian Klein bottle in the class β if $\lambda \geq 2$.*

It is interesting to consider what happens for large λ. We have essentially no tools to prove lower bounds when the Lagrangians are of high genus and may be Floer-theoretically obstructed. The most pessimistic conjecture is that Lagrangians with high genus become flexible enough that:

Conjecture 2.8. $\lim_{\lambda\to\infty} \mathfrak{y}(X, \omega_\lambda, \beta) < \infty$.

The following lemma gives an upper bound on $\mathfrak{y}(X, \omega_\lambda, \beta)$, but it goes to infinity with λ.

Lemma 2.9. *We have $\mathfrak{y}(X, \omega_\lambda, \beta) \leq 20\ell + 2$ when $\lambda < 10\ell + 2$.*

Proof. If $\lambda < 10\ell + 1$ then there is a tropical Lagrangian in the class β with nonorientable genus $20\ell + 2$. We show the tropical curve for $\ell = 2$ in Fig. 3 below; for general ℓ we simply repeat the pattern between the vertical blue bars as often as required to get from the left-hand side to the right-hand side of the rectangle.

The edges of this tropical curve are:

- internal edges parallel to either $(3, 1)$ or $(2, -1)$,

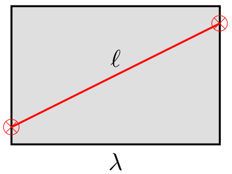


Figure 2. A visible Lagrangian Klein bottle in $(S^2 \times S^2, \omega_\lambda)$ for $\lambda < 2$. The cores of two cross-caps are indicated with cross-hairs

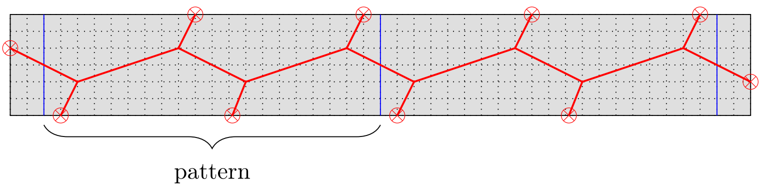


FIGURE 3. A tropical curve giving a Lagrangian of genus $20\ell + 2$ in the case $\ell = 2$

- external edges parallel to $(2, -1)$ or $(1, 2)$.

The corresponding tropical Lagrangian intersects the horizontal spheres with even multiplicity and the vertical spheres with odd multiplicity, so it inhabits the class β. The vertices of the tropical curve are not smooth[2]: each has self-intersection equal to 2. By [9, Theorem 3.2], this tropical curve therefore yields an immersed Lagrangian with 8ℓ double points and $2 + 4\ell$ cross-caps where it hits the boundary (marked with cross-hairs in Fig. 3). When we perform Polterovich surgery at the double points, we obtain a Lagrangian which is topologically a surface of genus 8ℓ with $4\ell + 2$ cross-caps. This has Euler characteristic $2 - 16\ell - 4\ell - 2 = -20\ell$, so the nonorientable genus is $2 + 20\ell$. □

Remark 2.10. It seems harder to make the genus significantly smaller using tropical Lagrangians, but there is no reason to believe that tropical Lagrangians should give a sharp upper bound for η.

3. Nonsqueezing

3.1. Statement

For each connected open interval $I \subset \mathbb{R}$ (length $|I|$), let C_I denote the cylinder $I \times (\mathbb{R}/2\pi\mathbb{Z})$ with coordinates (p, θ), equipped with the symplectic form $\frac{1}{2\pi} dp \wedge d\theta$; this has total area $|I|$. Let S^2 denote the 2-sphere equipped with its area form σ satisfying $\int_{S^2} \sigma = 2$.

Let $U_I = S^2 \times C_I$. Note that U_I is obtained from $(S^2 \times S^2, \omega_{|I|})$ by excising the spheres $S^2 \times \{n, s\}$, where n, s denote the poles of the second factor. Arguing as in Lemma 2.5, we see that if $|I| > 1$, the only nontrivial class $\beta \in H_2(U_I; \mathbb{Z}/2)$ is represented by a Lagrangian Klein bottle (see Fig. 4).

Theorem 3.1. *Suppose that $|I| \leq 1$. If $\iota \colon K \to U_I$ is a Lagrangian embedding of the Klein bottle in the class β then $\iota_* \colon \mathbb{Q} = H_1(K; \mathbb{Q}) \to H_1(U_I; \mathbb{Q}) = \mathbb{Q}$ is the zero map.*

Remark 3.2. The proof of Theorem 3.1 will occupy the rest of the paper. It uses SFT and neck-stretching.

[2] At each vertex of a tropical curve, the outgoing edges v_1, v_2, v_3 must sum to zero; if we write m for the determinant $|v_1 \wedge v_2| = |v_2 \wedge v_3| = |v_3 \wedge v_1|$ then the self-intersection of this vertex is defined to be $\frac{m-1}{2}$. Smoothness means all vertices have self-intersection zero.

Remark 3.3. Note that if $|I| > 1$ then $H_1(L;\mathbb{Q}) \to H_1(U_I;\mathbb{Q})$ is an isomorphism for the Lagrangian Klein bottle L coming from Lemma 2.5. To see this, take either one of the circles living over the points marked with cross-hairs in Fig. 4; this is a generator for both $H_1(L;\mathbb{Q})$ and $H_1(U_I;\mathbb{Q})$. We deduce:

Corollary 3.4. *The Lagrangian Klein bottle in $U_{(0,1+\epsilon)}$ from Lemma 2.5 cannot be squeezed into $U_{(0,1)}$.*

Remark 3.5. To reduce Conjecture 2.7 to this result, you would need to produce a pair of symplectic spheres in the class $[S^2 \times \star]$ which "link" your Lagrangian Klein bottle in an appropriate way. Since this class has non-minimal symplectic area, it is difficult to control the SFT limit of such spheres.

We now proceed to the proof of Theorem 3.1.

3.2. Mohnke's almost complex structure

Pick a flat metric g on the Klein bottle. There is a contact form (the canonical 1-form) on the unit cotangent bundle $M \subset T^*K$ whose closed Reeb orbits correspond to closed geodesics on K. We will not distinguish notationally between geodesics and the corresponding Reeb orbits and we will write $-\gamma$ for the geodesic obtained by reversing γ. There are two isolated simple geodesics γ_0, γ_1 which are the core circles for two disjoint embedded Möbius strips in K. Any isolated geodesic is a multiple cover of one of these and all other geodesics occur in one-parameter families. We call the isolated geodesics *odd* and the other geodesics *even*.

Theorem 3.6. (Mohnke [10, Section 2.1]) *There exists an almost complex structure J^- on the cotangent bundle T^*K with the following properties:*

1. *J^- is cylindrical at infinity and suitable for neck-stretching.*
2. *For any geodesic γ there is a finite-energy J^--holomorphic cylinder f_γ in T^*K asymptotic to γ and $-\gamma$.*
3. [10, Lemma 7(2)] *Any J^--holomorphic cylinder in T^*K which intersects the zero-section is one of these f_γ for some closed geodesic γ.*

Remark 3.7. If we let $W := \overline{T^*K}$ denote the compactification of the cotangent bundle obtained by gluing on its ideal contact boundary M then there is a well-defined intersection pairing $H_2(W,M;\mathbb{Z}/2) \otimes H_2(W;\mathbb{Z}/2) \to \mathbb{Z}/2$. The cylinders f_γ define elements of $H_2(W,M;\mathbb{Z}/2)$ and we have [10, Lemma

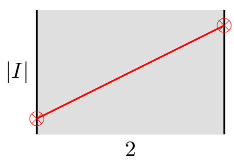


FIGURE 4. The visible Lagrangian Klein bottle in U_I when $|I| > 1$

7(3)]

$$f_\gamma \cdot K = \begin{cases} 1 \text{ if } \gamma \text{ is odd} \\ 0 \text{ if } \gamma \text{ is even.} \end{cases}$$

Remark 3.8. [10, Lemma 7(1)] Note that there are also no finite energy planes in T^*K, nor in the symplectisation $\mathbb{R} \times M$, for *any* cylindrical almost complex structure adapted to our chosen contact form. This is because there are no contractible Reeb orbits, and a finite energy plane would provide a nullhomotopy of its asymptote.

3.3. Neck-stretching

Let $I = (0,1)$ and $\bar{I} = [0,1]$. Suppose there is a Lagrangian Klein bottle $K \subset U_I$ such that $\mathbb{Q} = H_1(K;\mathbb{Q}) \to H_1(U_I;\mathbb{Q}) = \mathbb{Q}$ is nonzero (in particular, it is injective). Think of K sitting inside $U_{\bar{I}}$ and make symplectic cuts to $U_{\bar{I}}$ at $p = 0,1$ to obtain a Lagrangian Klein bottle K living in the manifold $X = S^2 \times S^2$ equipped with the product symplectic form giving the factors areas 2 and 1 respectively. Crucially, the symplectic cut introduces symplectic spheres S_0 and S_1 (at the $p = 0,1$ cuts respectively) which are disjoint from K.

Pick a sequence of almost complex structures J_t, $t \in \mathbb{R}$, on X with the following properties:

- on a Weinstein neighbourhood of K, J_t coincides with Mohnke's almost complex structure J^-;
- on a neck-stretching region $(a_t, b_t) \times M$ around K, J_t is a neck-stretching sequence;
- the spheres S_0, S_1 are J_t-holomorphic for all $t \in \mathbb{R}$.

Pick a point k on K which does not lie on any of the cylinders f_γ for an odd geodesic γ. Let $u_t\colon S^2 \to X$ be a J_t-holomorphic curve representing the class $\alpha = [\star \times S^2]$ and such that $u_t(0) = k$; there is a unique such u_t up to reparametrisation by a theorem of Gromov [7, 2.4.C], since α is a minimal area sphere class in X.

By the SFT compactness theorem [2], there is a sequence t_i such that u_{t_i} converges (after reparametrisations) to a holomorphic building with components in T^*K (the completion of the Weinstein neighbourhood of K), components in $\mathbb{R} \times M$ (the completion of the neck) and components in $X \backslash K$ (the completion of the complement of the Weinstein neighbourhood).

3.4. SFT limit analysis

The components $v_1, \ldots, v_n$ of the SFT limit building living in $X \backslash K$ can be compactified, yielding topological surfaces in X with boundary on K; we will still denote these by $v_1, \ldots, v_n$. The sum of the ω-areas of the v_i (weighted by multiplicities if the SFT limit involves a branched cover) equals the ω-area of α, which is 1.

Lemma 3.9. *There must be at least two planar components amongst the v_i, possibly geometrically indistinct* (i.e. having the same image).

Proof. First note that the limit building intersects K because $u_t(0) = k \in K$ for all t. It also necessarily has at least one component in $X \backslash K$ because T^*K is exact and so contains no closed holomorphic curves. A genus zero holomorphic building with at least two levels must have two planar components (just for topological reasons) though these could be geometrically indistinct. Any planar components live in $X \backslash K$. □

Lemma 3.10. *There are two components v_0, v_1 of the limit building such that $v_i \cdot S_j = \delta_{ij}$. These components are planar and there are no further components of the limit building in $X \backslash K$.*

Proof. Since α intersects S_0 and S_1 there must be components of the limit building which intersect S_0 and S_1. By positivity of intersections, either:

(A) there is one component v_1 which hits both S_0 and S_1 once transversely and all other components are disjoint from S_0, S_1.

(B) there are two components v_0, v_1 such that v_0 intersects S_0 once transversely and is disjoint from S_1 and vice versa for v_1.

Moreover, each of these components occurs with multiplicity one in the SFT limit in order to get the correct intersection numbers $\alpha \cdot S_0, \alpha \cdot S_1$.

If v_2 is a component which does not intersect S_0 or S_1 then it defines a class in $H_2(U_I, K; \mathbb{Z})$. By assumption, the kernel of the map $\mathbb{Z} \oplus \mathbb{Z}/2 = H_1(K; \mathbb{Z}) \to H_1(U_I; \mathbb{Z}) = \mathbb{Z}$ is precisely the torsion part. Therefore the long exact sequence

$$\cdots \to H_2(U_I; \mathbb{Z}) \to H_2(U_I, K; \mathbb{Z}) \to H_1(K; \mathbb{Z}) \to H_1(U_I; \mathbb{Z}) \to \cdots$$

splits off a sequence

$$\cdots \to \mathbb{Z} \to H_2(U_I, K; \mathbb{Z}) \to \mathbb{Z}/2 \to 0.$$

This implies that the areas of classes in $H_2(U_I, K; \mathbb{Z})$ are half-integer multiples of the area of the generator $\beta \in H_2(U_I; \mathbb{Z})$, which is 2. Therefore v_2 has integer area.

The area of α is 1, so the (weighted) sum of the ω areas of the v_i equals 1. Since v_1 has positive area, v_2 must have positive area strictly less than 1, but this is not possible if v_2 has integer area. Therefore there cannot be any component v_2 disjoint from S_0 and S_1.

By Lemma 3.9, there are at least two planar components (or one planar component with multiplicity two) in the limit building. This is not compatible with Case (A), so we must be in Case (B) and v_0, v_1 must additionally be planes. □

Lemma 3.11. 1. *All the remaining parts of the limit building are cylinders.*

2. *At least one of these cylinders lives in T^*K and has the form f_γ for an odd geodesic γ.*
3. *There are no other cylindrical components of the SFT limit building in T^*K.*

Proof. 1 If a component has three or more punctures then the limit building must contain at least three planar components (counted with multiplicity) but we have seen that all the planar components must live in

$X \backslash K$ (Remark 3.8) and that there are precisely two such components (Lemma 3.10).

2 Since $u_t(0) = k$ for all t, the limit building contains a component in T^*K, which must be a cylinder of the form f_γ by Theorem 3.6(3). At least one of these cylindrical components must correspond to an odd geodesic because α has odd intersection with K in $H_2(X; \mathbb{Z}/2)$ and the intersection number picks up contributions from each component of the building inside T^*K, which are nontrivial if and only if γ is odd (Remark 3.7).

3 If there are two or more cylindrical components in T^*K then there must be a further cylindrical component in $\mathbb{R} \times M$ which connects the asymptotes of two of these cylinders. Since this cylinder has no positive asymptote, this cannot exist by the maximum principle. □

Proof of Theorem 3.1. We chose $k \in K$ not to lie on any of the cylinders f_γ for γ an odd geodesic, but we have showed that these are the only cylinders which can arise as components of the SFT limit building. Since the SFT limit building must pass through k, we get a contradiction. □

Acknowledgements

I would like to thank Ivan Smith, Emily Maw, Georgios Dimitroglou Rizell, and Daniel Cavey for helpful conversations around this topic. My research is supported by EPSRC Grant EP/P02095X/2.

References

[1] Audin, M.: Fibrés normaux d'immersions en dimension double, points doubles d'immersions lagragiennes et plongements totalement réels. Comment. Math. Helv. **63**(4), 593–623 (1988)

[2] Bourgeois, F., Eliashberg, Y., Hofer, H., Wysocki, K., Zehnder, E.: Compactness results in symplectic field theory. Geom. Topol. **7**, 799–888 (2003)

[3] Castro, I., Urbano, F.: Minimal Lagrangian surfaces in $\mathbb{S}^2 \times \mathbb{S}^2$. Commun. Anal. Geom. **15**(2), 217–248 (2007)

[4] Dai, B., Ho, C.-I., Li, T.-J.: Nonorientable Lagrangian surfaces in rational 4-manifolds. Algebr. Geom. Topol. **19**(6), 2837–2854 (2019)

[5] Givental, A.B.: Lagrangian imbeddings of surfaces and the open Whitney umbrella. Funktsional. Anal. i Prilozhen. **20**(3), 35–41 (1986). (**96**)

[6] Goodman, E.: Lagrangian tori in $/{>}R^4$ and $S^2 \times S^2$. PhD thesis, Stanford University (2015)

[7] Gromov, M.: Pseudo holomorphic curves in symplectic manifolds. Invent. Math. **82**(2), 307–347 (1985)

[8] Matessi, D.: Lagrangian pairs of pants. arXiv:1802.02993 (2018)

[9] Mikhalkin, G.: Examples of tropical-to-Lagrangian correspondence. Eur. J. Math. **5**(3), 1033–1066 (2019)

[10] Mohnke, K.: How to (symplectically) thread the eye of a (Lagrangian) needle. arXiv:math/0106139 (2001)

[11] Nemirovski, S.: Homology class of a Lagrangian Klein bottle (transl.). Izv. Math. **73**, 689–698 (2009)

[12] Nemirovski, S.: Lagrangian Klein bottles in $\mathbb{R}^{2n}$. Geom. Funct. Anal. **19**(3), 902–909 (2009)

[13] Polterovich, L.: The surgery of Lagrange submanifolds. Geom. Funct. Anal. **1**(2), 198–210 (1991)

[14] Shevchishin, V.: Lagrangian embeddings of the Klein bottle and the combinatorial properties of mapping class groups. Izv. Ross. Akad. Nauk Ser. Mat. **73**(4), 153–224 (2009)

[15] Shevchishin, V., Smirnov, G.: Symplectic triangle inequality. Proc. Am. Math. Soc. **148**(4), 1389–1397 (2020)

[16] Symington,M.: Four dimensions from two in symplectic topology. In: Topology and Geometry of Manifolds (Athens, GA, 2001), Volume 71 of Proc. Sympos. Pure Math., pp. 153–208. Amer. Math. Soc., Providence (2003)

Jonathan David Evans
Department of Mathematics and Statistics
University of Lancaster
Bailrigg LA1 4YW
UK
e-mail: j.d.evans@lancaster.ac.uk

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:39
https://doi.org/10.1007/s11784-022-00960-x
Published online April 18, 2022

Journal of Fixed Point Theory and Applications

Construction of a linear K-system in Hamiltonian Floer theory

Kenji Fukaya, Yong-Geun Oh, Hiroshi Ohta and Kaoru Ono

In celebration of Claude Viterbo's 60th birthday.

Abstract. The notion of linear K-system was introduced by the present authors as an abstract model arising from the structure of compactified moduli spaces of solutions to Floer's equation in the book (Fukaya et al. in Springer monographs in mathematics, Springer, Berlin, 2020). The purpose of the present article is to provide a geometric realization of the linear K-system associated with solutions to Floer's equation in the Morse–Bott setting. Immediate consequences [when combined with the abstract theory from Fukaya et al. (Springer monographs in mathematics, Springer, Berlin, 2020)] are the construction of Floer cohomology for periodic Hamiltonian systems on general compact symplectic manifolds without any restriction, and the construction of an isomorphism over the Novikov ring between the Floer cohomology and the singular cohomology of the underlying symplectic manifold. The present article utilizes various analytical results on pseudoholomorphic curves established in our earlier papers and books. However, the paper itself is geometric in nature, and does not presume much prior knowledge of Kuranishi structures and their construction but assumes only the elementary part thereof, and results from Fukaya et al. (Surv Differ Geom 22:133–190, 2018) and Fukaya et al. (Exponential decay estimate and smoothness of the moduli space of pseudoholomorphic curves) on their construction, and the standard knowledge on Hamiltonian Floer theory. We explain the general procedure of the construction of a linear K-system by explaining in detail the inductive steps of ensuring the compatibility conditions for the system of Kuranishi structures leading to a linear K-system for the case of Hamiltonian Floer theory.

Mathematics Subject Classification. 53D40, 53D35, 58D27, 57P99.

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.
Kenji Fukaya is supported partially by Simons Collaboration on homological mirror symmetry, Yong-Geun Oh by the IBS project IBS-R003-D1, Hiroshi Ohta by JSPS Grant-in-Aid for Scientific Research Nos. 15H02054, 21H00983, 21K18576, 21H00985 and Kaoru Ono by JSPS Grant-in-Aid for Scientific Research, Nos. 26247006, 19H00636.

Contents

1. Introduction

The technique of virtual fundamental cycles or chains now provides us with a general and powerful method for studying certain moduli spaces in symplectic geometry and gauge theory. In [25], the first author and the fourth of this article introduced the notion of Kuranishi structure and constructed the virtual fundamental cycle of the moduli space of stable maps from a marked semi-stable curve to a general closed symplectic manifold based on the theory of Kuranishi structures, and applied it to Floer theory for periodic Hamiltonian systems. After that, in [10,11], the authors of the present article constructed the virtual fundamental chain of the moduli space of stable maps from a marked bordered semi-stable curve to a closed symplectic manifold with Lagrangian boundary condition using the theory of Kuranishi structures, and applied it to Lagrangian intersection Floer theory.

In these two studies, we had to construct a certain compatible system of Kuranishi structures on these moduli spaces.

Recently in [23], the authors developed a systematic foundation of the theory of Kuranishi structures and virtual fundamental chains in a general abstract setting. In particular, we axiomatized the properties of two systems, called a *linear K-system* and a *tree-like K-system*, consisting of abstract paracompact metrizable spaces equipped with Kuranishi structures satisfying certain compatibility conditions.

A tree-like K-system [23, Definition 21.9] is a model arising from the moduli spaces of stable maps from a marked bordered semi-stable curve with Lagrangian boundary condition which are used in Lagrangian Floer theory. In fact, in [20,21], we carried out a geometric realization of the tree-like K-system using the moduli spaces of stable maps from a marked bordered semi-stable curve. Namely, we constructed a system of Kuranishi structures on the moduli spaces which indeed satisfies the axioms of the tree-like K-system given in [23]. In the procedure of constructing the required Kuranishi structures we introduced the notion of *obstruction bundle data* assigned to each point of the moduli space [20, Definition 5.1] and showed that we can associate a Kuranishi structure on each moduli space to any obstruction bundle data in a canonical way [20, Theorem 7.1]. In addition, if the system of the obstruction bundle data satisfies certain compatibility conditions (*disk-component-wise system of obstruction bundle data* in the sense of [21, Definition 5.1]), then the obtained system of Kuranishi structures associated to the system of obstruction bundle data defines a tree-like K-system [21, Theorem 5.3].

On the other hand, a linear K-system [23, Definition 16.6] is the model that is used in Floer theory for periodic Hamiltonian systems which arises from the moduli spaces of solutions to Floer's equation. The purpose of the present article is to give a geometric realization of the linear K-system using the moduli spaces of solutions to Floer's equation. The strategy to achieve this realization lies on the line similar to that for the tree-like K-system. Namely, we introduce the notion of obstruction bundle data at each point for the moduli spaces of solutions to Floer's equation in Definition 4.11. Since we explained the construction of a Kuranishi structure associated to obstruction bundle data in [20,21] in detail and since such a construction is similar for the case of a linear K-system, we mainly present the way to equip moduli spaces with suitable obstruction bundle data in this article. In Sect. 5, we will construct a system of obstruction bundle data satisfying certain compatibility conditions which gives rise to a linear K-system. Comparing to [21], we take a slightly different way for constructing the compatible system of obstruction bundle data in order to demonstrate that both methods are applicable to both cases. Here we use *outer collars* of the moduli spaces to describe the compatibility. This method is outlined in [17, Remark 4.3.89]. The notion of outer collars is introduced and used in [23, Chapter 17] in the abstract setting.

Now our main theorem in this article is summarized as follows.

Theorem 1.1. (Theorem 2.9) *Let* (X, ω) *be a closed symplectic manifold and* $H : X \times S^1 \to \mathbb{R}$ *a smooth function. Suppose that the set* $\mathrm{Per}(H)$ *of all contractible 1-periodic orbits of the time dependent Hamiltonian vector field of* H *is Morse–Bott non-degenerate* (see Condition 2.2)*. Then, we can construct a linear K-system* $\mathcal{F}_X(H)$ *such that the critical submanifolds are connected components of* $\widetilde{\mathrm{Per}}(H)$ (which are copies of connected components of $\mathrm{Per}(H)$, see Definition 2.3)*, and the spaces of connecting orbits are the outer collared spaces of the moduli spaces of solutions to Floer's equation* (2.4)*.*

We also construct morphisms of linear K-systems. Combining these results with general properties concerning linear K-systems proved in [23], we have the following theorems:

Theorem 1.2. (Theorem 9.1) *Under the assumption of Theorem 1.1, we can define the Floer cohomology* $HF(X, H; \Lambda_{0,\mathrm{nov}})$ *of a periodic Hamiltonian system, also called the* Hamiltonian Floer cohomology, *which is independent of various choices involved in the definition.*

Here $\Lambda_{0,\mathrm{nov}}$ is the Novikov ring defined by

$$\Lambda_{0,\mathrm{nov}} = \left\{ \sum_{i=0}^{\infty} a_i T^{\lambda_i} \;\middle|\; a_i \in \mathbb{C}, \lambda_i \in \mathbb{R}_{\geq 0}, \lim_{i\to\infty} \lambda_i = \infty \right\}. \tag{1.1}$$

We define the Novikov field Λ_{nov} as its field of fractions by allowing λ_i to be negative.

Theorem 1.3. (Theorem 9.2, Corollary 10.3) *Suppose that two functions* $H^1, H^2 : X \times S^1 \to \mathbb{R}$ *satisfy the assumption in Theorem 1.1. Then the Floer cohomologies* $HF(X, H^r; \Lambda_{\mathrm{nov}})$ *over the Novikov field* Λ_{nov} *associated to* H^r $(r = 1, 2)$ *are isomorphic as* Λ_{nov}*-modules:*

$$HF(X, H^1; \Lambda_{\mathrm{nov}}) \cong HF(X, H^2; \Lambda_{\mathrm{nov}}).$$

Moreover, they are isomorphic to $H(X; \Lambda_{\mathrm{nov}})$*.*

More specifically, in Theorem 2.9, we associate an inductive system of linear K-systems in the sense of [23, Section 16] to a periodic Hamiltonian system. Together with [23, Theorem 16.39], we obtain the Floer cohomology of a periodic Hamiltonian system in Theorem 9.1. In Sect. 6 we construct a morphism between two linear K-systems associated to different Hamiltonians in Theorem 6.4. Together with [23, Theorem 16.39], it implies that the Floer cohomology of a periodic Hamiltonian system is independent of the Hamiltonian function (Theorem 9.2). We then calculate the Floer cohomology of a periodic Hamiltonian system on any compact symplectic manifold in Sect. 10. See Corollary 10.3. It gives a proof of the homological version of Arnold's conjecture [2] about the number of periodic orbits of a periodic Hamiltonian system. This proof is the same as those in the literature [25,29,33], modulo technical detail. All the proofs in the literature as well as the one in this article, first define the Floer cohomology of a periodic Hamiltonian system. In the generality we discuss here, we need to use virtual fundamental chain techniques. The references [25,29,33] use various versions of virtual fundamental chain techniques ([23,28,36] etc.) for this purpose.

A significant difference lies in the way we prove that the Floer cohomology of a periodic Hamiltonian system on a symplectic manifold X is isomorphic to the cohomology of X itself. In the literature there appeared basically three different ways to prove it. (We mention only the methods which are established to work for arbitrary compact symplectic manifolds.)

(1) Morse–Bott method:
   (a) We first include the case when the set $\mathrm{Per}(H)$ of periodic orbits of the periodic Hamiltonian $H : X \times S^1 \to \mathbb{R}$ does not necessarily consist of isolated points but is a submanifold. Technically speaking we include the case when Floer's functional $\mathcal{A}_H$ (see (2.9)) is not necessarily a Morse function but only a Morse–Bott function.
   (b) We prove the independence of the Floer cohomology of the periodic Hamiltonian H.
   (c) We next consider the case when $H \equiv 0$ and prove that the Floer cohomology of the periodic Hamiltonian system is isomorphic to the cohomology of X.

(2) Method to reduce to a small Morse function:
   We consider the case when the Hamiltonian H is sufficiently small in the C^2 sense and is time independent. We prove that the Floer cohomology of the periodic Hamiltonian system is isomorphic to the (Morse) homology of X in that case.

(3) Reduction to Lagrangian Floer cohomology ([8]) of the diagonal:
   (a) We consider the diagonal $\Delta \subset X \times X$ in the direct product of X with itself with symplectic form $-\omega \oplus \omega$.
   (b) We prove that Lagrangian intersection Floer cohomology
$$HF((\Delta, b), (\Delta, b); \Lambda_{0,\mathrm{nov}})$$
   is well-defined in the sense of [10].
   (c) We prove that $HF((\Delta, b), (\Delta, b); \Lambda_{0,\mathrm{nov}})$ is isomorphic to $H(X; \Lambda_{0,\mathrm{nov}})$ for a certain choice of a bounding cochain b.

The method of this article (Sect. 10) is a variation of the Morse–Bott method (1). We do not use S^1 equivariant Kuranishi structures explicitly. This is the point where our discussion is slightly different from those in the literature but is the same as in [31]. The method (2) was used in [25]. We provided its technical details in [13, Part 5]. The method (3) is [10, Theorem H]. See also [18, Subsection 6.3.3].

Any of those methods implies the following inequality due to [25,29,33]:

$$\#\mathrm{Per}(H) \geq \sum_k \mathrm{rank}\, H_k(X; \mathbb{Q}), \tag{1.2}$$

which was proved by several people for some special cases of X, for example, [7,9,26,30]. The case when the coefficient field $\mathbb{Q}$ is replaced by a finite field is studied in a recent paper by Abouzaid–Blumberg [1].

In the present paper, we presume that readers have only basic knowledge concerning the theory of Kuranishi structures, for example, some basic definitions and terminology contained in the quick survey [22, Part 7] or in the much shorter summary [20, Section 6]. While we use some part of [23], we refer to it in a pinpointed way, so readers do not have to understand the details of the proof therein. We also use the notion of outer collaring introduced in [23, Chapter 12], but do not assume readers' knowledge of the contents thereof. As for the analytic arguments, we employ the results from [16]. We also use the arguments in [16, Chapter 8] to prove smoothness of the coordinate change. In other words, we use the exponential decay estimates but do not use its proof. In this way [16] (except Chapter 8) and most parts of [23] are used as a 'black box' in this article.

Then, based on the arguments and results in [20] (See also [15,19].), we will construct the desired Kuranishi structures. Although there are some parts, especially Part 5 in [13] related to this article, we do not assume the contents of [13] for reading this article. Indeed we will repeat and describe the arguments here if necessary.

Throughout this paper, a *K-space* means a paracompact metrizable space with a Kuranishi structure. A *K-system* is an abbreviation for a system of K-spaces. We assume that X is any closed symplectic manifold, unless otherwise mentioned.

The authors would like to thank a referee for careful reading and useful comments.

2. Floer's equation and moduli space of solutions

We review Floer's moduli space in this section. Our discussion is mostly the same as [13, Section 29], except that we include the case when the space of contractible periodic orbits has positive dimension. We repeat several parts for the reader's convenience.

Let $H : X \times S^1 \to \mathbb{R}$ be a smooth function on a symplectic manifold (X, ω). We put $H_t(x) = H(x, t)$ where $t \in S^1$ and $x \in X$. The function H_t generates the Hamiltonian vector field $\mathfrak{X}_{H_t}$ defined by

$$i_{\mathfrak{X}_{H_t}}\omega = \mathrm{d}H_t.$$

It defines a one parameter family of diffeomorphisms $\exp_t^H : X \to X$ by

$$\begin{aligned} \exp_0^H(x) &= x, \\ \left(\frac{\mathrm{d}\exp_t^H}{\mathrm{d}t}\right)(x, t_0) &= \mathfrak{X}_{H_{t_0}}(\exp_{t_0}^H(x)). \end{aligned} \tag{2.1}$$

We denote by $\mathrm{Per}(H)$ the set of all 1-periodic orbits of the time dependent vector field $\mathfrak{X}_{H_t}$. We can identify

$$\mathrm{Per}(H) \cong \mathrm{Fix}(\exp_1^H) = \{x \in X \mid \exp_1^H(x) = x\}. \tag{2.2}$$

From now on, $\mathrm{Per}(H)$ denotes the set of contractible 1-periodic orbits. Our assumption that $\mathrm{Per}(H)$ is non-degenerate in the Morse–Bott sense is as follows:

Condition 2.1. We say that $\mathrm{Per}(H)$ is *Morse–Bott non-degenerate* if the following holds.

(1) $\mathrm{Fix}(\exp_1^H)$ is a smooth submanifold of X.

(2) Let $x \in \mathrm{Fix}(\exp_1^H)$ and consider the linear map: $d_x\exp_1^H : T_xX \to T_xX$. We require

$$T_x(\mathrm{Fix}(\exp_1^H)) = \{V \in T_xX \mid (d_x\exp_1^H)(V) = V\}. \tag{2.3}$$

Remark 2.2. The typical examples where Condition 2.1 is satisfied are the following two cases.

(1) $\mathrm{Per}(H)$ is discrete. In this case, Condition 2.1 is equivalent to the condition that the graph of $\exp_1^H$ is transversal to the diagonal in $X \times X$.

(2) The case $H = 0$.

To prove (1.2) it suffices to study these two cases only.

Definition 2.3. We put

$$\widetilde{\mathrm{Per}}(H) = \{(\gamma, w) \mid \gamma \in \mathrm{Per}(H),\ w : D^2 \to X,\ w(e^{2\pi it}) = \gamma(t)\}/\sim,$$

where $(\gamma, w) \sim (\gamma', w')$ if and only if $\gamma = \gamma'$ and

$$\omega([w] - [w']) = 0, \quad c_1(TX)([w] - [w']) = 0.$$

We have a natural surjection $\varpi : \widetilde{\mathrm{Per}}(H) \ni [(\gamma, w)] \to \gamma \in \mathrm{Per}(H)$. We put $\mathfrak{G} = \pi_2(X)/\sim$, where $\alpha \sim \alpha'$ if and only if $\omega[\alpha] = \omega[\alpha']$ and $c_1(TX)[\alpha] = c_1(TX)[\alpha']$. Then the group $\mathfrak{G}$ acts on $\widetilde{\mathrm{Per}}(H)$ by changing the bounding disk $w : D^2 \to X$ so that the action on the fiber $\varpi^{-1}(*)$ is simply transitive.

Following [9], we consider maps $u : \mathbb{R} \times S^1 \to X$ satisfying the equation

$$\frac{\partial u}{\partial \tau} + J\left(\frac{\partial u}{\partial t} - \mathfrak{X}_{H_t} \circ u\right) = 0 \tag{2.4}$$

which we call *Floer's equation.* Here τ and t are the coordinates of $\mathbb{R}$ and $S^1 = \mathbb{R}/\mathbb{Z}$, respectively. For $\tilde{\gamma}_\pm = (\gamma_\pm, w_\pm) \in \widetilde{\mathrm{Per}}(H)$ we consider the boundary condition

$$\lim_{\tau \to \pm\infty} u(\tau, t) = \gamma_\pm(t). \tag{2.5}$$

Proposition 2.4. *We assume Condition 2.1. Then for any solution u of (2.4) with*

$$\int_{\mathbb{R}\times S^1} \left\| \frac{\partial u}{\partial \tau} \right\|^2 \mathrm{d}\tau \mathrm{d}t < \infty$$

there exists $\gamma_\pm \in \mathrm{Per}(H)$ such that (2.5) is satisfied.

Proof. In the case of Remark 2.2 (1) this is proved by Floer [9, Proposition 3b]. In the case Remark 2.2 (2) this follows from the removable singularity theorem for pseudo-holomorphic curves. The general case can be proved in the same way as in [25, Lemma 11.2]. We omit the details of the proof of the general case since to prove (1.2) it suffices to consider the two cases in Remark 2.2. □

Remark 2.5. The convergence (2.5) is of exponential order. Namely we have

$$\|u(\tau, t) - \gamma_\pm(t)\|_{C^k} \le C_k e^{-c_k|\tau|}$$

for some $c_k > 0, C_k > 0$.

We decompose $\widetilde{\mathrm{Per}}(H)$ into connected components and write

$$\widetilde{\mathrm{Per}}(H) = \coprod_{\alpha \in \mathfrak{A}} R_\alpha. \tag{2.6}$$

Here $\mathfrak{A}$ is the index set for connected components of $\widetilde{\mathrm{Per}}(H)$. We denote by $\overline{R_\alpha} \subset X$ the image of R_α under the projection $\varpi : [(\gamma, w)] \in \widetilde{\mathrm{Per}}(H) \to \gamma(0) \in \mathrm{Fix}(\exp_1^H)(\cong \mathrm{Per}(H))$, see (2.2). In the Morse–Bott situation,[1] we need to use certain local systems on the space of 1-periodic orbits to equip the moduli spaces of solutions to Floer's equation for defining Floer complexes and chain homomorphisms with the orientation isomorphisms in the sense of linear K-systems [23, Condition 16.1 (VII)]. Gluing D^2 and $[0,\infty) \times S^1$ by identifying $e^{2\pi\sqrt{-1}t} \in \partial D^2$ and $(0,[t]) \in \{0\} \times S^1$, we obtain a capped half cylinder Z.

For each 1-periodic orbit $\gamma \in \overline{R_\alpha}$, let $\mathcal{P}_\gamma(\overline{R_\alpha})$ be the set of trivializations of γ^*TX as a symplectic vector bundle and write $\mathcal{P}(\overline{R_\alpha}) = \cup_{\gamma \in \overline{R_\alpha}} \mathcal{P}_\gamma(\overline{R_\alpha})$. Pick $\mathfrak{t} \in \mathcal{P}_\gamma(\overline{R_\alpha})$. Using $\mathfrak{t}$, we extend γ^*TX to a symplectic vector bundle over D^2. Gluing it with $(\gamma \circ \mathrm{pr}_2)^*TX$ on $[0,\infty) \times S^1$, we obtain a symplectic vector bundle $E(\mathfrak{t})$ on Z. Here $\mathrm{pr}_2 : [0,\infty) \times S^1 \to S^1$ is the second factor projection. We extend $\gamma^* J_{\partial D^2 = \{0\} \times S^1}$ to a complex structure, which is also denoted by J of the vector bundle $E(\mathfrak{t})$. The pull-back of the unitary connection induces a holomorphic vector bundle structure on $E(\mathfrak{t}) \to Z$. Pick a cut-off function $\chi : [0,\infty) \to \mathbb{R}$ such that $\chi = 0$ near 0 and $\chi(\tau) = 1$ for sufficiently large τ.

Based on [9, section 2e] and [25, section 21], we define the map $P(\gamma;\mathfrak{t}) : \Gamma(Z; E(\mathfrak{t})) \to \Gamma(Z; E(\mathfrak{t}) \otimes \Lambda^{0,1})$ by

$$P(\gamma;\mathfrak{t})\xi = \begin{cases} \overline{\partial}_{E(\mathfrak{t})}\xi & \text{on } D^2, \\ (1-\chi(\tau))\,\overline{\partial}_{E(\mathfrak{t})}\xi(\mathrm{d}\tau - \sqrt{-1}\mathrm{d}t) + \chi(\tau) D_\gamma \overline{\partial}_{J,H}\xi & \text{on } [0,\infty) \times S^1. \end{cases} \tag{2.7}$$

Here $D_\gamma \overline{\partial}_{J,H}$ is the linearization operator for Floer's equation of connecting orbits at the stationary solution $u(\tau,t) = \gamma(t)$, i.e.,

$$D_\gamma \overline{\partial}_{J,H}\xi = \left(\nabla_{\frac{\partial}{\partial \tau}}\xi + \nabla_{\frac{\partial}{\partial t}}(J\xi) - \nabla_\xi(J\mathfrak{X}_H)\right)(\mathrm{d}\tau - \sqrt{-1}\mathrm{d}t).$$

Pick a positive number δ such that δ is less than the absolute values of the non-zero eigenvalues of the self adjoint differential operator $\zeta \mapsto A\zeta = \nabla_{\frac{\partial}{\partial t}}(J\zeta) - \nabla_\zeta(J\mathfrak{X}_H)$ on S^1. We use $e^{\delta\tau}$ for defining weighted Sobolev spaces and regard the operator $P(\gamma;\mathfrak{t})$ as an operator

$$P(\gamma;\mathfrak{t}) : W^{1,p}_\delta(Z; E(\mathfrak{t})) \to L^p_\delta(Z; E(\mathfrak{t}) \otimes \Lambda^{0,1}).$$

For two trivializations $\mathfrak{t}_1, \mathfrak{t}_2$, the operators $P(\gamma;\mathfrak{t}_1)$ and $P(\gamma;\mathfrak{t}_2)$ may have different Fredholm indices, in general. However, the real determinant bundles as $O(1)$-bundles are canonically identified. These operators coincide on $[0,\infty) \times S^1$, where we do not use trivializations. They may differ on D^2. Note

[1] In the non-degenerate case, the issue discussed here was written in [9,25]. Here we extend the argument to the Morse–Bott case.

that there exists a complex vector bundle F on $\mathbb{C}P^1$ such that $E(\mathfrak{t}_2)$ is isomorphic to the gluing and smoothing of $E(\mathfrak{t}_1)$ and F along the fibers on $0 \in D^2$ and $\infty \in \mathbb{C}P^1$. Since the Dolbeault operator $\overline{\partial}_F$ on $\mathbb{C}P^1$ with coefficients in F is a complex Fredholm operator, its index is a virtual complex vector space and the coincidence condition for sections of $E(\mathfrak{t}_1)$ and F at $0 \in D^2$ and $\infty \in \mathbb{C}P^1$ is required in the complex vector space $E(\mathfrak{t}_1)|_0 = F|_\infty$. Hence its real determinant is canonically oriented (complex orientation). Thus the real determinant lines for the operator $P(\gamma; \mathfrak{t}_1)$ and $P(\gamma; \mathfrak{t}_2)$ are canonically isomorphic. The determinant line bundle for the family $P(\gamma; \mathfrak{t})$ on $\mathcal{P}(\overline{R_\alpha})$ descends to an $O(1)$-bundle on $\overline{R_\alpha}$, which we denote by o_{R_α}. (In the case of Floer theory for Lagrangian intersections, see [11, Proposition 8.8.1].)

Definition 2.6. We call o_{R_α} on R_α the *orientation system of the critical submanifold* R_α (see [23, Condition 16.1 (VII) (i)]).

Definition 2.7. Let $\alpha_1, \alpha_2 \in \mathfrak{A}$. We denote by $\widetilde{\mathcal{M}}^{\mathrm{reg}}(X, H; \alpha_-, \alpha_+)$ the set of all maps $u : \mathbb{R} \times S^1 \to X$ with the following properties:

(1) u satisfies (2.4).
(2) There exist $\tilde{\gamma}_\pm = (\gamma_\pm, w_\pm) \in R_{\alpha_\pm}$ such that (2.5) and

$$w_- \# u \sim w_+$$

are satisfied. Here # is the obvious concatenation.

The translation along $\tau \in \mathbb{R}$ defines an $\mathbb{R}$ action on $\widetilde{\mathcal{M}}^{\mathrm{reg}}(X, H; \alpha_-, \alpha_+)$. This $\mathbb{R}$ action is free unless $\alpha_- = \alpha_+$. We denote by $\mathcal{M}^{\mathrm{reg}}(X, H; \alpha_-, \alpha_+)$ the quotient space of this action. For the case $\alpha_- = \alpha_+$, the set $\mathcal{M}^{\mathrm{reg}}(X, H; \alpha_-, \alpha_+)$ is the empty set by definition.

We define the evaluation map

$$\mathrm{ev} = (\mathrm{ev}_-, \mathrm{ev}_+) : \mathcal{M}^{\mathrm{reg}}(X, H; \alpha_-, \alpha_+) \to R_{\alpha_-} \times R_{\alpha_+} \tag{2.8}$$

by

$$\mathrm{ev}(u) = ((\gamma_-, w_-), (\gamma_+, w_+)).$$

Remark 2.8. Note that w_- can not be determined by the map u only. In fact if $[v] \in \pi_2(X)$ then u may be also regarded as an element of $\mathcal{M}^{\mathrm{reg}}(X, H; [v]\#\alpha_-, [v]\#\alpha_+)$. More precisely, an element of $\widetilde{\mathcal{M}}^{\mathrm{reg}}(X, H; \alpha_-, \alpha_+)$ should be regarded as a pair (u, α_-). We write u instead of (u, α_-) in the case no confusion can occur.

The main result we will prove in Sects. 3–5 is the following.

Theorem 2.9. *We assume Condition* 2.1.

(1) *The space $\mathcal{M}^{reg}(X, H; \alpha_-, \alpha_+)$ has a compactification $\mathcal{M}(X, H; \alpha_-, \alpha_+)$.*
(2) *The compact space $\mathcal{M}(X, H; \alpha_-, \alpha_+)$ has a Kuranishi structure with corners. The evaluation map* ev *is extended to it as a strongly smooth map in the sense of* [23, Definition 3.40].
(3) *There exists a linear K-system $\mathcal{F}_X(H)$, whose critical submanifold is R_α ($\alpha \in \mathfrak{A}$) and whose space of connecting orbits is $\mathcal{M}(X, H; \alpha_-, \alpha_+)^{\boxplus 1}$. Here $\mathcal{M}(X, H; \alpha_-, \alpha_+)^{\boxplus 1}$ is the outer collaring of $\mathcal{M}(X, H; \alpha_-, \alpha_+)$.*

See Definition 5.1 for the notion of the outer collaring. (We use o_{R_α} in Definition 2.6 as the orientation on R_α.)

(4) *The Kuranishi structure on $\mathcal{M}(X,H;\alpha_-,\alpha_+)^{\boxplus 1}$ in (3) coincides with one in (2) on $\mathcal{M}(X,H;\alpha_-,\alpha_+) \subset \mathcal{M}(X,H;\alpha_-,\alpha_+)^{\boxplus 1}$.*

In later sections, we consider a slightly general case when we include additional interior marked points on the domain. We will denote such a moduli space by $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ where ℓ is a number of interior marked points. See Sect. 3 for the precise definition.

Definition 2.10. Theorems 2.9 and [23, Theorem 16.39] define a $\Lambda_{0,\mathrm{nov}}$-module. We call the resulting $\Lambda_{0,\mathrm{nov}}$-module the *Floer cohomology of a periodic Hamiltonian system* and write $HF(X,H;\Lambda_{0,\mathrm{nov}})$.

See Theorem 9.1 for the well-definedness of the Floer cohomology group $HF(X,H;\Lambda_{0,\mathrm{nov}})$.

The proof of Theorem 2.9 occupies Sects. 3–5. In the rest of this section, we discuss an easy part of the construction.

Definition 2.11. We define group homomorphisms $E : \pi_2(X) \to \mathbb{R}$ and $\mu : \pi_2(X) \to \mathbb{Z}$ by $E(\beta) = \omega[\beta]$, $\mu(\beta) = 2c_1(TX)[\beta]$. We recall that the image of $(E,\mu) : \pi_2(X) \to \mathbb{R} \times \mathbb{Z}$ is isomorphic to the group $\mathfrak{G}$ in Definition 2.3. We also denote by $E : \mathfrak{G} \to \mathbb{R}$ and $\mu : \mathfrak{G} \to \mathbb{Z}$ the induced homomorphisms, respectively.

The map $E : \mathfrak{A} \to \mathbb{R}$ is defined by $E(\alpha) = \omega[w]$ where $(\gamma, w) \in R_\alpha$.

Remark 2.12. It is easy to see that $E : \mathfrak{A} \to \mathbb{R}$ is well-defined. Namely $\omega[w]$ is independent of the element (γ, w) but depends only on R_α. In the case of Remark 2.2 (1) this is trivial. In the case of Remark 2.2 (2) this follows from the fact that $\mathfrak{G} = \mathfrak{A}$ in this case. In the general case, it follows from the fact that R_α is a critical submanifold of Floer's functional $\mathcal{A}_H$ defined by

$$\mathcal{A}_H(\gamma, w) = \int_{D^2} w^*\omega + \int_{S^1} H(\gamma(t),t)\mathrm{d}t. \tag{2.9}$$

To define $\mu : \mathfrak{A} \to \mathbb{Z}$ we recall that the linearized operator of the Eq. (2.4) is given by

$$\begin{aligned} D_u\overline{\partial} - \nabla.(J\mathfrak{X}_H) \otimes (\mathrm{d}\tau - \sqrt{-1}\mathrm{d}t) \ &: \ L^2_{m+1,\delta}(\mathbb{R} \times S^1; u^*TX) \\ &\to L^2_{m,\delta}(\mathbb{R} \times S^1; u^*TX \otimes \Lambda^{0,1}). \end{aligned} \tag{2.10}$$

Here $V \mapsto \nabla_V(J\mathfrak{X}_H)$ is the covariant derivative of $J\mathfrak{X}_H$ with respect to the tangential vector V. Here $L^2_{m,\delta}$ is the completion of the space of smooth sections with respect to the weighted L^2_m norm with weight $e^{\delta|\tau|}$. (Here $\delta > 0$ and τ is the coordinate of $\mathbb{R}$. We assume that m is sufficiently large, say > 100.)

Definition 2.13. We define the *virtual dimension* of $\mathcal{M}^{\mathrm{reg}}(X,H;\alpha_-,\alpha_+)$ by

$$\dim \mathcal{M}^{\mathrm{reg}}(X,H;\alpha_-,\alpha_+) = \mathrm{Index}(2.10) + \dim R_{\alpha_-} + \dim R_{\alpha_+} - 1.$$

Here u is an element of $\mathcal{M}^{\mathrm{reg}}(X,H;\alpha_-,\alpha_+)$.

Let $(\gamma, w) \in R_{\mathfrak{a}}$. Recall $w : D^2 \to X$. We take a smooth map $\mathbb{R} \times S^1 \to D^2$ such that it is the projection to $\partial D^2 = S^1$ on $\mathbb{R}_{\geq 0} \times S^1$, is constantly equal to 0 on $\mathbb{R}_{\leq -T} \times S^1$ and defines a diffeomorphism from $(-T, 0) \times S^1$ onto $\operatorname{Int} D^2 \backslash \{0\}$. We compose it with w to obtain $u_w : \mathbb{R} \times S^1 \to X$. Let $\chi : \mathbb{R} \to [0, 1]$ be a smooth function such that $\chi(\tau) = 0$ for $\tau < -1$ and $\chi(\tau) = 1$ for $\tau > 1$. We modify (2.10) to define

$$\begin{aligned} D_u\overline{\partial} - \chi(\tau)\nabla_\cdot(J\mathfrak{X}_H) \otimes (\mathrm{d}\tau - \sqrt{-1}\mathrm{d}t) \ : \ & L^2_{m+1,\delta}(\mathbb{R} \times S^1; u_w^*TX) \\ \to L^2_{m,\delta}(\mathbb{R} \times S^1; u_w^*TX \otimes \Lambda^{0,1}). & \end{aligned} \tag{2.11}$$

Let $(\gamma, w) \in R_{\mathfrak{a}}$. Note that $w : D^2 \to X$ induces a trivialization of $w^*TX \to D^2$, hence the one of γ^*TX, which is denoted by $\mathfrak{t}_w$.

Definition 2.14. We define $\mu : \mathfrak{A} \to \mathbb{Z}$ by

$$\mu(\alpha) = \operatorname{Index} P(\gamma, \mathfrak{t}_w).$$

Lemma 2.15. (1) *For $\beta \in \pi_2(X)$ and $\alpha \in R_\alpha$. We have*

$$\mu(\beta\alpha) = \mu(\alpha) + 2c_1(TX)[\beta].$$

(2) *In the case $H \equiv 0$, $\mathfrak{A} = \mathfrak{G}$ and $\mu(\beta) = 2c_1(TX)[\beta]$.*
(3) *We have*

$$\dim \mathcal{M}^{reg}(X, H; \alpha_-, \alpha_+) = \mu(\alpha_+) - \mu(\alpha_-) - 1 + \dim R_{\alpha_+}.$$

Proof. (1) follows from the excision property of the index.

Let $(\gamma_\pm, w_\pm) \in R_{\alpha_\pm}$. To prove (3) it suffices to show that

$$\operatorname{Index}(2.10) + \dim R_{\alpha_-} + \operatorname{Index} P(\gamma_-, \mathfrak{t}_{w_-}) = \operatorname{Index} P(\gamma_+, \mathfrak{t}_{w^+}). \tag{2.12}$$

We can prove (2.12) as follows. We remark that our operator is of the form $\frac{\mathrm{d}}{\mathrm{d}\tau} + Q_\tau$ where Q_τ is a family of elliptic differential operators on S^1. If we write $P(\gamma_-, \mathfrak{t}_{w^-})$ in this form, then in the limit $\tau \to \infty$, the multiplicity of zero eigenvalue of the operator Q_τ is exactly equal to $\dim R_{\alpha_-}$. If we write (2.10) in this form, then in the limit $\tau \to -\infty$, the operator Q_τ has exactly $\dim R_{\alpha_-}$ as the multiplicity of zero eigenvalue. (2.12) is a consequence of this observation and a well established result on spectral flow [4, section 7].

We finally prove (2). In view of (1)(3), it suffices to show the case $\beta = 0$. Namely, γ_p is the constant periodic orbit at p for $H = 0$ and the constant bounding disk $w_p : D^2 \to X$ induces the canonical trivialization $\mathfrak{t}_{\mathrm{tri}}$, i.e., $(\gamma_p)^*TX \cong S^1 \times T_pX$. In that case we can see directly that the index of $P(\gamma_p, \mathfrak{t}_{\mathrm{tri}})$ is 0 by, e.g., the Atiyah–Patodi–Singer index formula [3]. □

3. Stable map compactification of Floer's moduli space

3.1. Definition of $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$

We define a compactification of the space $\mathcal{M}^{\mathrm{reg}}(X, H; \alpha_-, \alpha_+)$ in this section. This compactification is classical. The description of this section is mostly the same as that of [13, Section 30]. We repeat the argument for reader's convenience and to fix the notation. We need to stabilize the domain sometimes in later constructions. For this purpose we treat the case when we include interior marked points on the domain. We will denote it by $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ where ℓ is a number of interior marked points.

Remark 3.1. Including marked points also has applications to symplectic topology [12]. Especially it is used in the construction of the spectral invariant with bulk, which is now known to contain more informations than the version without bulk. (See [22,37].)

We consider $(\Sigma, (z_-, z_+, \vec{z}))$, a genus zero semistable curve Σ with $2 + \ell$ marked points z_-, z_+ and $\vec{z} = (z_1, \ldots, z_\ell)$. Two marked points z_-, z_+ will play different roles.

Definition 3.2.
- Let Σ_0 be the union of the irreducible components of Σ such that
  (1) $z_-, z_+ \in \Sigma_0$.
  (2) Σ_0 is connected.
  (3) Σ_0 is the smallest among those satisfying (1),(2) above.

  We call Σ_0 the *mainstream* of $(\Sigma, (z_-, z_+, \vec{z}))$, or simply, of Σ. An irreducible component of the mainstream Σ_0 is called a *mainstream component*. Other irreducible components of Σ are called *bubble components*.
- Let $\Sigma_a \subset \Sigma$ be a mainstream component. If $z_- \notin \Sigma_a$, then there exists a unique singular point $z_{a,-}$ of Σ contained in Σ_a such that
  (1) z_- and $\Sigma_a \backslash \{z_{a,-}\}$ belong to the different connected component of $\Sigma \backslash \{z_{a,-}\}$.
  (2) z_+ and $\Sigma_a \backslash \{z_{a,-}\}$ belong to the same connected component of $\Sigma \backslash \{z_{a,-}\}$.

  In case $z_- \in \Sigma_a$ we set $z_- = z_{a,-}$. We define $z_{a,+}$ in the same way.

  We call $z_{a,\pm}$ the *transit points*. We call other singular points *non-transit singular points*.
- A *parametrization of the mainstream* of $(\Sigma, (z_-, z_+, \vec{z}))$ is $\varphi = \{\varphi_a\}$, where $\varphi_a : \mathbb{R} \times S^1 \to \Sigma_a$ for each irreducible component Σ_a of the mainstream such that:
  (1) φ_a is a biholomorphic map $\varphi_a : \mathbb{R} \times S^1 \cong \Sigma_a \backslash \{z_{a,-}, z_{a,+}\}$.
  (2) $\lim_{\tau \to \pm\infty} \varphi_a(\tau, t) = z_{a,\pm}$.
- A union of one mainstream component and all the trees of the bubble components rooted on it is called an *extended mainstream component*. We sometimes denote by $\widehat{\Sigma}_a$ the extended main stream component containing the mainstream component Σ_a. Here and hereafter we call each of $\widehat{\Sigma}_a \backslash \{z_{a,-}, z_{a,+}\}$ an *interior of extended mainstream component*. See Fig. 1.

Let $\mathfrak{A}$ be as in (2.6) and $\alpha_\pm \in \mathfrak{A}$.

Definition 3.3. The set $\widehat{\mathcal{M}}_\ell(X, H; \alpha_-, \alpha_+)$ consists of triples $((\Sigma, (z_-, z_+, \vec{z})), u, \varphi)$ satisfying the following conditions: Here $\ell = \#\vec{z}$.

(1) $(\Sigma, (z_-, z_+, \vec{z}))$ is a genus zero semistable curve with $\ell+2$ marked points.
(2) φ is a parametrization of the mainstream.

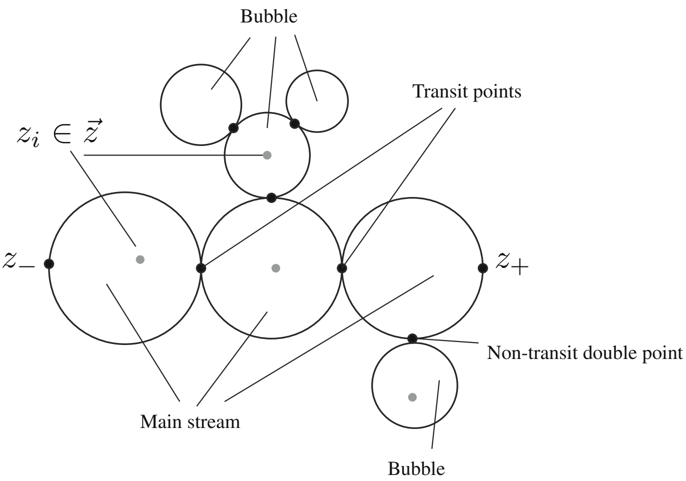


FIGURE 1. $(\Sigma, (z_-, z_+, \vec{z}))$

(3) For each extended main stream component $\widehat{\Sigma}_a$, the map u induces $u_a : \widehat{\Sigma}_a \setminus \{z_{a,-}, z_{a,+}\} \to X$ which is a continuous map.[2]
(4) If Σ_a is a mainstream component and $\varphi_a : \mathbb{R} \times S^1 \to \Sigma_a$ is as above, then the composition $u_a \circ \varphi_a$ satisfies the Eq. (2.4).
(5)
$$\int_{\mathbb{R}\times S^1} \left\| \frac{\partial(u \circ \varphi_a)}{\partial \tau} \right\|^2 \mathrm{d}\tau \mathrm{d}t < \infty.^3$$
(6) If Σ_b is a bubble component, then u is pseudo-holomorphic on it.
(7) If $\widehat{\Sigma}_{a_1}$ and $\widehat{\Sigma}_{a_2}$ are extended mainstream components and if $z_{a_1,+} = z_{a_2,-}$, then
$$\lim_{\tau\to+\infty} (u_{a_1} \circ \varphi_{a_1})(\tau, t) = \lim_{\tau\to-\infty} (u_{a_2} \circ \varphi_{a_2})(\tau, t)$$
holds for each $t \in S^1$. ((5) and Proposition 2.4 imply that both of the left and right hand sides converge.)
(8) If $\widehat{\Sigma}_a$, $\widehat{\Sigma}_{a'}$ are extended mainstream components and $z_{a,-} = z_-$, $z_{a',+} = z_+$, then there exist $(\gamma_\pm, w_\pm) \in R_{\alpha_\pm}$ such that
$$\lim_{\tau\to-\infty} (u_a \circ \varphi_a)(\tau, t) = \gamma_-(t),$$
$$\lim_{\tau\to+\infty} (u_{a'} \circ \varphi_{a'})(\tau, t) = \gamma_+(t).$$
Moreover,
$$[u_*[\Sigma]] \# w_- = w_+$$
where # is the obvious concatenation.

[2] In other words u is a continuous map from the complement of the set of the transit points.
[3] Condition (5) follows from the rest of the conditions in Definition 3.3.

(9) We assume $((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)$ is stable in the sense of Definition 3.5 below.

To define stability we first define the group of automorphisms.

Definition 3.4. We assume that $((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)$ satisfies (1)–(8) above. The *extended automorphism group* $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)$ of $((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)$ consists of maps $v:\Sigma\to\Sigma$ with the following properties:

(1) $v(z_-) = z_-$ and $v(z_+) = z_+$. In particular, v preserves each of the mainstream component Σ_a of Σ. Moreover, v fixes each of the transit points.
(2) $u = u\circ v$ holds outside the set of the transit points.
(3) If Σ_a is a mainstream component of Σ, then there exists $\tau_a\in\mathbb{R}$ such that

$$(v\circ\varphi_a)(\tau,t) = \varphi_a(\tau+\tau_a,t). \tag{3.1}$$

on $\mathbb{R}\times S^1$.
(4) There exists $\sigma\in\mathrm{Perm}(\ell)$ such that $v(z_i) = z_{\sigma(i)}$. Here $\mathrm{Perm}(\ell)$ denotes the group of permutations of ℓ elements.

The *automorphism group* $\mathrm{Aut}((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)$ of $((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)$ consists of the elements of $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)$ such that σ in Item (4) is the identity.

Definition 3.5. We say $((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)$ is *stable* if $\mathrm{Aut}((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)$ is a finite group. (This is equivalent to the finiteness of $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)$.)

Lemma 3.6. *An element $((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)$ satisfying (1)–(8) of Definition 3.3 is stable if and only if, for each irreducible component Σ_i of Σ, one of the following holds.*

(1) *Σ_i is a bubble component, and u is nonconstant on Σ_i.*
(2) *Σ_i is a bubble or a mainstream component, and*

$$\#((\vec{z}\cup\{z_-,z_+\})\cap\Sigma_i) + \#(\textit{Singular points on }\Sigma_i)\ge 3.$$

(3) *Σ_i is a mainstream component Σ_a, and*

$$\frac{\mathrm{d}}{\mathrm{d}\tau}(u\circ\varphi_a)$$

is nonzero at some point. Here τ is the coordinate of the $\mathbb{R}$ factor in $\mathbb{R}\times S^1$.

The proof is left to the reader.

Definition 3.7. On the set $\widehat{\mathcal{M}}_\ell(X,H;\alpha_-,\alpha_+)$ we define two equivalence relations $\sim_1$, $\sim_2$ as follows.

$((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)\sim_1((\Sigma',(z'_-,z'_+,\vec{z}')),u',\varphi')$ if and only if there exists a biholomorphic map $v:\Sigma\to\Sigma'$ with the following properties:

(1) $v(z_-) = z'_-$ and $v(z_+) = z'_+$. In particular v sends the mainstream of Σ to the mainstream of Σ'. Moreover, v sends transit points to transit points.

(2) $u' = u \circ v$ holds outside the transit points.
(3) If Σ_a is a mainstream component of Σ and $v(\Sigma_a) = \Sigma'_{a'}$, then we have

$$v \circ \varphi_a = \varphi'_{a'} \tag{3.2}$$

on $\Sigma_a \backslash \{z_{a,-}, z_{a,+}\}$.
(4) $v(z_i) = z'_i$.

The equivalence relation $\sim_2$ is defined replacing (3.2) by existence of $\tau_a \in \mathbb{R}$ such that

$$(v \circ \varphi_a)(\tau, t) = \varphi'_{a'}(\tau + \tau_a, t). \tag{3.3}$$

We put

$$\widetilde{\mathcal{M}}_\ell(X, H; \alpha_-, \alpha_+) = \widehat{\mathcal{M}}_\ell(X, H; \alpha_-, \alpha_+)/\sim_1,$$
$$\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+) = \widehat{\mathcal{M}}_\ell(X, H; \alpha_-, \alpha_+)/\sim_2 .$$

In case $\ell = 0$ we write $\mathcal{M}(X, H; \alpha_-, \alpha_+)$ etc.

Definition 3.8. For the case $X =$ one point and $H \equiv 0$, we obtain the space $\mathcal{M}_\ell(\text{one point}, 0; \alpha_0, \alpha_0)$. Here α_0 is the unique point in $\mathrm{Per}(0)$. We denote this space by $\mathcal{M}_\ell(\text{source})$.

Remark 3.9. The parametrization $\varphi_a : \mathbb{R} \times S^1 \to \Sigma_a \backslash \{z_{a,-}, z_{a,+}\}$ of each mainstream component Σ_a is automatically determined by the marked curve $(\Sigma, z_-, z_+, \vec{z})$ up to the ambiguity $\varphi_a(\tau, t) \mapsto \varphi_a(\tau + \tau_0, t + t_0)$ where $(\tau_0, t_0) \in \mathbb{R} \times S^1$. In other words, the choice of parametrization of the mainstream one to one corresponds to the choice of one additional marked point on each mainstream component. We can use this fact to define the structure of an orbifold with corner on $\mathcal{M}_\ell(\text{source})$.

Example 3.10. (1) $\mathcal{M}_0(\text{source}) = \emptyset$. $\mathcal{M}_1(\text{source}) \cong S^1$. In fact, its point is determined by the S^1 factor of the marked point.
(2) $\mathcal{M}_2(\text{source}) \cong S^1 \times S^1 \times [0, 1]$. To see this let us first consider the case when there is only one mainstream component. In that case let $\varphi(\tau_i, t_i)$ $(i = 1, 2)$ be the marked points. Because of translation symmetry t_1, t_2 and $\tau_2 - \tau_1$ parametrize such elements of $\mathcal{M}_2(\text{source})$. (Note in case these two marked points coincide there occurs bubble.) The boundary $S^1 \times S^1 \times \partial[0, 1]$ then corresponds to the situation where there are two mainstream components.

Definition 3.11. An element $[(\Sigma, (z_-, z_+, \vec{z})), u, \varphi]$ of $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ is said to be *source stable* if $(\Sigma, (z_-, z_+, \vec{z}))$ is stable as genus zero marked curve.

3.2. Topology on $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$

We next define a topology on the moduli space $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$. It is mostly the same as the topology of the moduli space of stable maps which was introduced in [25, Definitions 10.2 and 10.3]. However, since the notion of equivalence relation $\sim_2$ is slightly different (namely the condition (3.3) is included) it is slightly different.

We here recall the following well-known fact. Suppose that a sequence of genus 0 stable curves $(\Sigma^j,(z^j_-,z^j_+,\vec z^j))$ converges to $(\Sigma,(z_-,z_+,\vec z))$ in the moduli space of stable curves. Then for a sufficiently small $\epsilon>0$ and a large j we have a biholomorphic embedding

$$\psi_{j,\epsilon}:\Sigma\backslash(\epsilon \text{ neighborhood of singular points})\to\Sigma^j, \tag{3.4}$$

$\vec z^{j,\epsilon}\subset\Sigma$ and $R(\epsilon)>0$ such that the following holds.

Condition 3.12. (1) $\lim_{j\to\infty,\epsilon\to0} z^{j,\epsilon}_i=z_i$ for each $i\in\{1,\dots,\ell,+,-\}$.
(2) $\lim_{\epsilon\to0}R(\epsilon)=\infty$.
(3) $\psi_{j,\epsilon}(z^{j,\epsilon}_i)=z^j_i$.
(4) Any connected component of the complement of the image of $\psi_{j,\epsilon}$ is biholomorphic to one of $[0,R]\times S^1$ with $R>R(\epsilon)$, $(-\infty,0]\times S^1$, or $[0,\infty)\times S^1$.

We can use Margulis' lemma to prove it. (See [5, Chapter 4], [32, Chapter 11], [27, Chapter IV], for example.) In fact, we can use this condition as the definition of the topology of the moduli space of marked stable curves of genus 0.

Definition 3.13. Suppose that elements $[(\Sigma,(z_-,z_+,\vec z)),u,\varphi]$ and $[(\Sigma^j,(z^j_-,z^j_+,\vec z^j)),u^j,\varphi^j]$ of $\mathcal M_\ell(X,H;\alpha_-,\alpha_+)$ are source stable. We say $[(\Sigma^j,(z^j_-,z^j_+,\vec z^j)),u^j,\varphi^j]$ *converges* to $[(\Sigma,(z_-,z_+,\vec z)),u,\varphi]$ and write

$$\operatorname*{lims}_{j\to\infty}[(\Sigma^j,(z^j_-,z^j_+,\vec z^j)),u^j,\varphi^j]=[(\Sigma,(z_-,z_+,\vec z)),u,\varphi]$$

if there exist $\psi_{j,\epsilon}$, $\vec z^{j,\epsilon}\subset\Sigma$ and $R(\epsilon)>0$ as in (3.4) with the following properties.

(1) Condition 3.12 (1)–(4) are satisfied.
(2) For each $\epsilon>0$

$$\lim_{j\to\infty}\sup\{d(u^j(\psi_{j,\epsilon}(z)),u(z))\mid z\in\mathrm{Dom}(\psi_{j,\epsilon})\}=0. \tag{3.5}$$

Here $\mathrm{Dom}(\psi_{j,\epsilon})=\Sigma\backslash(\text{Union of }\epsilon\text{ neighborhoods of singular points})$ is the domain of $\psi_{j,\epsilon}$ as in (3.4) and d is a metric on X.[4]
(3) For each mainstream component Σ_a of Σ there exist a mainstream component $\Sigma^j_{a(j)}$ of Σ^j and $T_j\to\infty$, $\tau_j\in\mathbb R$ such that

$$\lim_{j\to\infty}\sup\{d((\varphi_{a_j}^{-1}\circ\psi_{j,\epsilon}\circ\varphi_a)(\tau-\tau_j,t),(\tau,t))\mid(\tau-\tau_j,t)\in[-T_j,T_j]\times S^1\}=0. \tag{3.6}$$

Here d is the standard metric on $\mathbb R\times S^1$.
(4) The 'diameter' of the image of u^j of each connected component $\mathfrak W$ of $\Sigma^j\backslash\psi_{j,\epsilon}(\mathrm{Dom}(\psi_{j,\epsilon}))$ converges to 0 in the sense of Condition 3.16 below.

To state Condition 3.16 below we need some notation. Let

$$((\Sigma,(z_-,z_+,\vec z)),u,\varphi)\in\widehat{\mathcal M}_\ell(X,H;\alpha_-,\alpha_+)$$

[4]This condition is independent of d since X is compact.

and define

$$\hat{u} : \Sigma \to X$$

as follows. We use the flow map $\exp_t^H : X \to X$ defined in (2.1). We identify $t \in [0,1) \subset S^1 \cong \mathbb{R}/\mathbb{Z}$.

(1) If $z \in \Sigma$ is in the mainstream component Σ_a and is not a transit point, then we take (τ, t) with $\varphi_a(\tau, t) = z$ and put

$$\hat{u}(z) = (\exp_t^H)^{-1}(u(z)).$$

(2) Suppose $z \in \Sigma$ is in a bubble component. Let z_0 be the root of the bubble tree containing z. We take (τ, t) with $\varphi_a(\tau, t) = z_0$. Then we put

$$\hat{u}(z) = (\exp_t^H)^{-1}(u(z)).$$

(3) (1) and (2) define $\hat{u}$ outside the set of the singular points. It is easy to see that it extends to a map $\hat{u} : \Sigma \to X$.

Remark 3.14. (1) The map u does *not* extend continuously to the transit point. In fact, the image of the neighborhood of the transit point contains a periodic orbit which may not consist of a point. After composing $(\exp_t^H)^{-1}$ it can be extended to a continuous map because of (2.5).

(2) The map $\hat{u}$ is *not* continuous at $t = [0] = [1] \in S^1$. This is because $\varphi_1^H \neq \varphi_0^H =$ identity. However, it is continuous at the transit point.

Definition 3.15. We call $\hat{u}$ the *redefined connecting orbit map.*

Condition 3.16. In the situation of Definition 3.13 we require the following. There exists $o(\epsilon, j) > 0$ with $\lim_{j\to\infty,\epsilon\to 0} o(\epsilon, j) \to 0$ such that for each connected component $\mathfrak{W}$ of $\Sigma^j \backslash \psi_{j,\epsilon}(\mathrm{Dom}(\psi_{j,\epsilon}))$ we have either

(i) $\mathrm{Diam}(\hat{u}^j(\mathfrak{W})) < o(\epsilon, j)$ for the case when $\mathfrak{W}$ corresponds to a transit point of $(\Sigma, (z_-, z_+, \vec{z})), u, \varphi)$, or

(ii) $\mathrm{Diam}(u^j(\mathfrak{W})) < o(\epsilon, j)$ for the case when $\mathfrak{W}$ corresponds to a non-transit singular point of $(\Sigma, (z_-, z_+, \vec{z})), u, \varphi)$.

Here $\hat{u}^j$ is the redefined connecting orbit map of $((\Sigma^j, (z_-^j, z_+^j, \vec{z}^j)), u^j, \varphi^j)$.

Definition 3.17. We say a sequence $[(\Sigma^j, (z_-^j, z_+^j, \vec{z}^j)), u^j, \varphi^j]$ in $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ *converges* to $[(\Sigma, (z_-, z_+, \vec{z})), u, \varphi] \in \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ and write

$$\lim_{j\to\infty} [(\Sigma^j, (z_-^j, z_+^j, \vec{z}^j)), u^j, \varphi^j] = [(\Sigma, (z_-, z_+, \vec{z})), u, \varphi],$$

if there exist $\vec{w}^j \subset \Sigma^j$ and $\vec{w} \subset \Sigma$ such that $[(\Sigma^j, (z_-^j, z_+^j, \vec{z}^j \cup \vec{w}^j)), u^j, \varphi^j]$ and $[(\Sigma, (z_-, z_+, \vec{z} \cup \vec{w})), u, \varphi]$ are source stable and

$$\operatorname*{lims}_{j\to\infty} [(\Sigma^j, (z_-^j, z_+^j, \vec{z}^j \cup \vec{w}^j)), u^j, \varphi^j] = [(\Sigma, (z_-, z_+, \vec{z} \cup \vec{w})), u, \varphi]$$

in the sense of Definition 3.13.

Remark 3.18. We define the topology by defining the notion of convergence of sequences. We can do so by the following reason. We can define the notion of combinatorial or topological type of an element of $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$. (See [25, Section 19] for example.) If we fix a combinatorial type, we can embed the set of elements $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ to a space obtained as a stable pair of a curve and a map from it with the fixed combinatorial type. The topology in Definition 3.17 coincides with one which is defined by the standard topology of the moduli space of marked curves and, say, the C^∞ topology of the maps. Obviously, the latter topology is metrizable. Also there exist only countably many combinatorial types.

Theorem 3.19. *The space $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ is compact and Hausdorff with respect to the topology of Definition 3.17.*

The proof is the same as that of [25, Lemma 10.4 and Theorem 11.1], which relies on Proposition 2.4, and so is omitted.

We close this section with a few small remarks.

The evaluation map (2.8) is extended to

$$(\mathrm{ev}_-,\mathrm{ev}_+):\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)\to R_{\alpha_-}\times R_{\alpha_+}. \tag{3.7}$$

Moreover, we define evaluation maps

$$(\mathrm{ev}_1,\dots,\mathrm{ev}_\ell):\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)\to X^\ell \tag{3.8}$$

by

$$\mathrm{ev}_i([(\Sigma,(z_-,z_+,\vec{z}\cup\vec{w})),u,\varphi])=u(z_i).$$

They are continuous.

Let $\alpha_-=\alpha_0,\alpha_1,\dots,\alpha_{m-1},\alpha_m=\alpha_+\in\mathfrak{A}$. We consider the fiber product

$$\begin{aligned}&\mathcal{M}_{\ell_1}(X,H;\alpha_0,\alpha_1)\,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\mathcal{M}_{\ell_2}(X,H;\alpha_1,\alpha_2)\,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\cdots{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\\ &\quad\mathcal{M}_{\ell_{i+1}}(X,H;\alpha_i,\alpha_{i+1})\,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\cdots{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\mathcal{M}_{\ell_m}(X,H;\alpha_{m-1},\alpha_m)\end{aligned} \tag{3.9}$$

and denote it by

$$\mathcal{M}_{(\ell_1,\dots,\ell_m)}(X,H;\alpha_0,\alpha_1.,\dots,\alpha_m).$$

See Fig. 2. In case ℓ_i are all 0 we write $\mathcal{M}(X,H;\alpha_0,\alpha_1.,\dots,\alpha_m)$.

We also define a map

$$\mathcal{M}_{(\ell_1,\dots,\ell_m)}(X,H;\alpha_0,\alpha_1.,\dots,\alpha_m)\to\mathcal{M}_{\ell_1+\dots+\ell_m}(X,H;\alpha_-,\alpha_+) \tag{3.10}$$

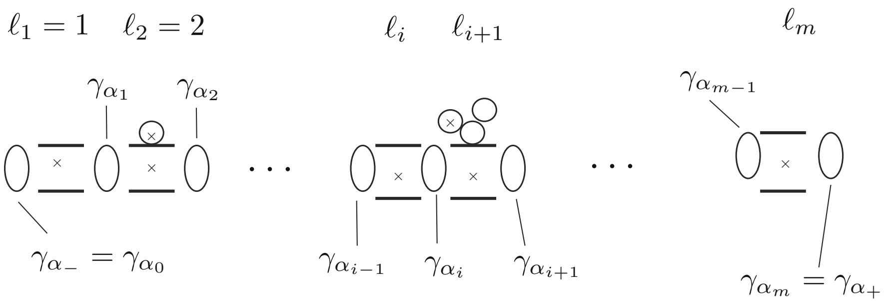


FIGURE 2. $\mathcal{M}_{(\ell_1,\dots,\ell_m)}(X,H;\alpha_0,\alpha_1.,\dots,\alpha_m)$

as follows. Let $[(\Sigma^a,(z_-^a,z_+^a,\vec{z}^a)),u^a,\varphi^a] \in \mathcal{M}_{\ell_a}(X,H;\alpha_{a-1},\alpha_a)$ for $a = 1,\dots,m$. We assume

$$\begin{aligned} &\mathrm{ev}_+([(\Sigma^a,(z_-^a,z_+^a,\vec{z}^a)),u^a,\varphi^a]) \\ &\quad = \mathrm{ev}_-([(\Sigma^{a+1},(z_-^{a+1},z_+^{a+1},\vec{z}^{a+1})),u^{a+1},\varphi^{a+1}]). \end{aligned} \tag{3.11}$$

On the disjoint union $\bigsqcup_{a=1}^m \Sigma_a$ we identify z_+^a and z_-^{a+1} to obtain Σ. We put $z_- = z_-^0$ and $z_+ = z_+^m$. We also put $\vec{z} = \bigcup_{a=1}^m \vec{z}^a$, and u and φ to be the union of u^a and φ^a, respectively. They obviously satisfy Definition 3.3 (1)–(6). Definition 3.3 (7) is a consequence of (3.11). We thus obtain the map (3.10). This map is obviously continuous, injective and is a homeomorphism onto its image.

4. Construction of Kuranishi structure

4.1. Statement

In this section we present technical details of the proof of the next theorem.

Theorem 4.1. (1) *The space $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ has a Kuranishi structure with corners together with an isomorphism*

$$\mathrm{ev}_+^*(\det TR_{\alpha_+}) \otimes \mathrm{ev}_+^*(o_{R_{\alpha_+}}) \cong o_{\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)} \otimes \mathrm{ev}_-^*(o_{R_{\alpha_-}})$$

of principal $O(1)$-bundles. Here $o_{\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)}$ is the orientation bundle defined by the Kuranishi structure as in [23, Definition 3.10 (1)] *and $o_{R_{\alpha_\pm}}$ are defined in* Definition 2.6.

(2) *Its codimension k normalized corner*[5] *is the union of the images of the map* (3.10) *with a certain Kuranishi structure on $\mathcal{M}_{(\ell_1,\dots,\ell_m)}(X,H;\alpha_0,\alpha_1.,\dots,\alpha_k)$.*

(3) *The evaluation maps* (3.7) *and* (3.8) *are induced from a strongly smooth map*[6] *of Kuranishi structures. The map* ev_+ *is weakly submersive.*

(4) *The dimension (as a K-space) is given by*

$$\dim \mathcal{M}_\ell(X,H;\alpha_-,\alpha_+) = \mu(\alpha_+) - \mu(\alpha_-) - 1 + \dim R_{\alpha_+} + 2\ell.$$

Remark 4.2. The Kuranishi structure on $\mathcal{M}_{(\ell_1,\dots,\ell_m)}(X,H;\alpha_0,\alpha_1.,\dots,\alpha_m)$ mentioned in Theorem 4.1 (2) is not in general the fiber product Kuranishi structure. The method we give in this section does not provide such a system of Kuranishi structures compatible with the fiber product yet. In Sect. 5 we will modify the Kuranishi structures on $\mathcal{M}_{\ell_i}(X,H;\alpha_{i-1},\alpha_i)$ at their outer collars so that they are consistent with the fiber product. Then we will obtain a required K-system so that $\mathcal{M}_{(\ell_1,\dots,\ell_m)}(X,H;\alpha_0,\alpha_1.,\dots,\alpha_m)$ has indeed a fiber product Kuranishi structure. We will carry out this procedure in Sect. 5.

Remark 4.3. The group $\mathfrak{G}$ in Definition 2.3 acts on $\widetilde{\mathrm{Per}}(H)$ so that it induces a simply transitive action on the index set $\mathfrak{A}$ in (2.6). Then we have a natural identification between $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ and $\mathcal{M}_\ell(X,H;g(\alpha_-),g(\alpha_+))$ for any $g \in \mathfrak{G}$ and $\alpha_\pm \in \mathfrak{A}$. Since our construction of Kuranishi structures

[5] See [23, Definition 24.18] for the definition of normalized corner.

[6] See [23, Definition 3.40] for the definition of strongly smooth map.

given below is independent of the choice of the bounding disk $w : D^2 \to X$ in the definition of $\widetilde{\mathrm{Per}}(H)$, the resulting Kuranishi structures for other choices of w's are isomorphic under this identification.

4.2. Obstruction bundle data

The detail of the proof of Theorem 4.1 (1)(2) given in this section is mostly the same as the one given in [20] (see also [13, Parts 4 and 5] if necessary). We repeat the proof for the completeness and also to prepare notations for the discussion in the next section. More specifically, in [20, Definition 5.1] we introduced the notion of *obstruction bundle data* for the moduli space of bordered stable maps of genus 0 and showed the existence [20, Theorem 11.1] and that we can associate a Kuranishi structure for any obstruction bundle data in a canonical way [20, Theorem 7.1]. The strategy of this article is the same. So we first discuss a version of the obstruction bundle data in our situation with Hamiltonian perturbation.

Definition 4.4. A *symmetric stabilization* of $((\Sigma, (z_-, z_+, \vec{z})), u, \varphi)$ is an ordered set $\vec{w}$ of points in Σ with the following properties.

(1) The set $\vec{w}$ is contained in the union of bubble components.
(2) $\vec{w} \cap \vec{z} = \emptyset$. None of the points in $\vec{w}$ is a singular point.
(3) For each bubble component Σ_b we have the inequality
$$\#(\vec{w} \cap \vec{z}) + \#(\text{Singular point on } \Sigma_b) \geq 3.$$
(4) For any $v \in \mathrm{Aut}^+((\Sigma, (z_-, z_+, \vec{z})), u, \varphi)$ we have $\sigma \in \mathrm{Perm}(\#\vec{w})$ such that $v(w_i) = w_{\sigma(i)}$.
(5) If $v \in \mathrm{Aut}((\Sigma, (z_-, z_+, \vec{z})), u, \varphi)$ and $\sigma = \mathrm{id}$ in (4), then $v = \mathrm{id}$ also.
(6) For each w_i the map u is an immersion at w_i.

Remark 4.5. Condition (5) is assumed only to simplify the notations in later discussion. By this condition we have an embedding $\mathrm{Aut}((\Sigma, (z_-, z_+, \vec{z})), u, \varphi) \to \mathrm{Perm}(m)$ where $m = \#\vec{w}$.

Lemma 4.6. *There exists a symmetric stabilization for any* $((\Sigma, (z_-, z_+, \vec{z})), u, \varphi)$.

The proof is similar to that for the case of the moduli space of stable maps, which is now standard. So we omit the proof.

Note that $(\Sigma, \{z_-, z_+\} \cup \vec{z} \cup \vec{w})$ may not yet be a stable marked curve because there can be a mainstream component Σ_a which contains only transit points or $z_\pm$. (Namely Σ_a may not contain any of non-transit singular point or a point of $\vec{z}$.) In Definition 4.4 we do not put additional marked points $\vec{w}$ on such a component. Actually it is in general impossible to find w_i on such a component satisfying Definition 4.4 (6). However, we can find a 'canonical' marked point on such a component. Namely we use a function f_{H,u,Σ_a} to find the canonical position in the $\mathbb{R}$ coordinate and the parametrization φ_a to find the S^1 coordinate. This is the way taken in [13, page 204], which we repeat below.

Let Σ_a be a mainstream component of Σ which does not contain non-transit singular point or marked point $\in \vec{z}$. We define a function f_{H,u,Σ_a} :

$\mathbb{R} \to \mathbb{R}$ as follows. We first define an element $(\gamma_{a,-}, w_{a,-}) \in \widetilde{\mathrm{Per}}(H)$. Here $\gamma_{a,-}$ is a closed orbit which corresponds to the transit point $z_{a,-}$ by

$$\gamma_{a,-}(t) = \lim_{\tau \to -\infty} u(\varphi_a(\tau, t)).$$

We consider the connected component of $\Sigma \setminus \{z_{a,-}\}$ which contains z_-. The restriction of u to this connected component together with w_- (which is a disk that bounds a curve, an element in R_{α_-}) determines a homotopy class of maps $w_{a,-} : D^2 \to X$ which bounds $\gamma_{a,-}$. We thus obtain $(\gamma_{a,-}, w_{a,-}) \in \widetilde{\mathrm{Per}}(H)$.

We define a function

$$f_{H,u,\Sigma_a}(\tau_0) = \int_{D^2} w_{a,-}^* \omega + \int_{\tau=-\infty}^{\tau=\tau_0} \int_{t \in S^1} (u \circ \varphi_a)^* \omega + \int_{t \in S^1} H(u(\tau_0, t), t) \mathrm{d}t. \tag{4.1}$$

Proposition 2.4 implies that this integral converges and both limits $\lim_{\tau \to \pm\infty} f_{H,u,\Sigma_a}(\tau)$ exist. Furthermore, using the fact that $u \circ \varphi_a$ satisfies (2.4) we can show that f_{H,u,Σ_a} is nondecreasing. Since Σ_a does not contain non-transit singular points or marked points $\in \vec{z}$, the stability implies that f_{H,u,Σ_a} is *strictly* increasing. In fact, the first derivative $\mathrm{d}f_{H,u,\Sigma_a}(\tau)/\mathrm{d}\tau$ is strictly positive unless $(\gamma_{a,-}, w_{a,-}) = (\gamma_{a,+}, w_{a,+})$. Thus we have proved:

Lemma 4.7. *There exists a unique $\tau_a \in \mathbb{R}$ such that*

$$f_{H,u,\Sigma_a}(\tau_a) = \frac{1}{2} \left(\lim_{\tau \to -\infty} f_{H,u,\Sigma_a}(\tau) + \lim_{\tau \to +\infty} f_{H,u,\Sigma_a}(\tau) \right). \tag{4.2}$$

Definition 4.8. We call the point $w_{a,\mathrm{can}} = \varphi_a(\tau_a, 0)$ the *canonical marked point on* Σ_a. (Note that by our assumption that Σ_a has no marked or singular points other than transit points or $z_\pm$, the point $w_{a,\mathrm{can}}$ is neither a marked nor a singular point.)

We denote by $\vec{w}_{\mathrm{can}}$ the totality of $w_{a,\mathrm{can}}$ for each mainstream component Σ_a which does not contain any non-transit singular point or a point of $\vec{z}$.

The above discussion also proves the next lemma.

Lemma 4.9. *If $((\Sigma, (z_-, z_+, \vec{z})), u, \varphi) \in \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ and $\vec{w}$ is its symmetric stabilization, then $(\Sigma, \{z_-, z_+\} \cup \vec{z} \cup \vec{w} \cup \vec{w}_{\mathrm{can}}))$ is stable.*

Now we define the notion of obstruction bundle data.[7]

Notation 4.10.
- We denote by $\mathcal{M}_\ell^{\mathrm{cl}}$ the moduli space of stable curves of genus zero without boundary and with ℓ marked points, and by $\overset{\circ}{\mathcal{M}}{}_\ell^{\mathrm{cl}}$ its subset consisting of elements with only one irreducible component.
- We denote by $\overset{\circ}{\mathcal{M}}_\ell(\mathrm{source})$ the set of points of $\mathcal{M}_\ell(\mathrm{source})$ which have only one irreducible component, (which is necessarily a mainstream component).

[7] This notion was originally introduced in [13, Definition 31.1] The difference is that we include the Morse–Bott case in this article, while in [13, Definition 31.1] we assumed [13, Definition 29.4] which implies that all the periodic orbits are isolated.

- We denote by $\pi : \mathcal{C}_\ell^{\mathrm{univ}} \to \mathcal{M}_\ell^{\mathrm{cl}}$ the universal family of stable curves of genus zero with ℓ marked points. See [20, Theorem 2.2] for example.

Definition 4.11. *Obstruction bundle data* $\mathfrak{E}_{\mathbf{p}}$ *centered at*

$$\mathbf{p} = [(\Sigma, (z_-, z_+, \vec{z})), u, \varphi] \in \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$$

are the data

$$\left(\vec{w}, \{\mathcal{V}(\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can,v}})\}_{\mathrm{v}}, \{(\psi_{\mathrm{v}}, \phi_{\mathrm{v}})\}_{\mathrm{v}}, \{K_{\mathrm{v}}^{\mathrm{obst}}\}_{\mathrm{v}}, \{E_{\mathbf{p},\mathrm{v}}(\mathfrak{y}_{\mathrm{v}})\}_{\mathrm{v}}, \{\mathcal{D}_i\}_{w_i \in \vec{w}}\right) \tag{4.3}$$

satisfying the conditions described below. We put

$$\mathfrak{x} = (\Sigma, \{z_-, z_+\} \cup \vec{z}).$$

Let $\mathrm{Irr}(\Sigma)$ be the set of the irreducible components of Σ and, for $\mathrm{v} \in \mathrm{Irr}(\Sigma)$, we denote by Σ_{v} the corresponding irreducible component.

(1) A symmetric stabilization $\vec{w}$ of $((\Sigma, z_-, z_+, \vec{z}), u, \varphi)$.

(2) A neighborhood $\mathcal{V}(\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can,v}})$ of

$$\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can,v}} = (\Sigma_{\mathrm{v}}, \{z_{\mathrm{v},-}, z_{\mathrm{v},+}\} \cup \vec{z}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can,v}})$$

for each $\mathrm{v} \in \mathrm{Irr}(\Sigma)$: Here
- $\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can,v}}$ is an irreducible component of $\mathfrak{x}$.
- $\{z_{\mathrm{v},-}, z_{\mathrm{v},+}\} = \{\text{transit points on } \Sigma\}$ [8] $\cap \Sigma_{\mathrm{v}}$.
- $\vec{z}_{\mathrm{v}} = (\vec{z} \cup \{\text{non-transit singular points on } \Sigma\}) \cap \Sigma_{\mathrm{v}}$.
- $\vec{w}_{\mathrm{v}} = \vec{w} \cap \Sigma_{\mathrm{v}}$ and $\vec{w}_{\mathrm{can,v}} = \vec{w}_{\mathrm{can}} \cap \Sigma_{\mathrm{v}}$.

Namely:

(a) In the case Σ_{v} is a mainstream component Σ_a, we include the parametrization φ_a to $\mathfrak{x}_{\mathrm{v}}$ and $\mathcal{V}(\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can,v}})$ is an open subset of $\overset{\circ}{\mathcal{M}}_{\ell_{\mathrm{v}}+\ell'_{\mathrm{v}}+\ell''_{\mathrm{v}}}(\text{source})$ where $\ell_{\mathrm{v}} = \#\vec{z}_{\mathrm{v}}, \ell'_{\mathrm{v}} = \#\vec{w}_{\mathrm{v}}$ and $\ell''_{\mathrm{v}} = \#\vec{w}_{\mathrm{can,v}}$.

(b) In the case Σ_{v} is a bubble component, $\mathcal{V}(\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can,v}})$ is an open subset of $\overset{\circ}{\mathcal{M}^{\mathrm{cl}}}_{\ell_{\mathrm{v}}+\ell'_{\mathrm{v}}+\ell''_{\mathrm{v}}}$ where $\ell_{\mathrm{v}} = \#\vec{z}_{\mathrm{v}}, \ell'_{\mathrm{v}} = \#\vec{w}_{\mathrm{v}}$ and $\ell''_{\mathrm{v}} = \#\vec{w}_{\mathrm{can,v}} = 0$.

(3) *Local trivialization data* $\{(\psi_{\mathrm{v}}, \phi_{\mathrm{v}})\}_{\mathrm{v}}$ at $\mathfrak{x} \cup \vec{w} \cup \vec{w}_{\mathrm{can}}$ in the sense of [20, Definition 3.8]: Namely, denoting by $\mathcal{V}_{\mathrm{v}}$ the neighborhood $\mathcal{V}(\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can,v}})$ taken in (2) above, $\psi_{\mathrm{v},z} : \mathcal{V}_{\mathrm{v}} \times \mathrm{Int}\, D^2 \to \mathcal{C}_\ell^{\mathrm{univ}}$ is an *analytic family of coordinates* at each non-transit singular point z of Σ which is contained in Σ_{v} (see [20, Definition 3.1]), and $\phi_{\mathrm{v}} : \mathcal{V}_{\mathrm{v}} \times (\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can,v}}) \to \pi^{-1}(\mathcal{V}_{\mathrm{v}})$ is a C^∞ trivialization over $\mathcal{V}_{\mathrm{v}}$ of the universal family $\pi : \mathcal{C}_\ell^{\mathrm{univ}} \to \mathcal{M}_\ell^{\mathrm{cl}}$ of marked stable curves (see [20, Definition 3.6]), which is compatible with the analytic family of coordinates in the sense of [20, Definition 3.7 (1)]. We also require the following additional conditions on a part of the local trivialization data.

(a) Let z_a be a transit point contained in Σ_a and Σ_{a+1}. Then the coordinate near z_a given by the local trivialization data is the parametrization φ_a or φ_{a+1} up to the $\mathbb{R}$ action. Namely it is $(\tau, t) \mapsto \varphi_a(\tau + \tau_0, t)$ (resp. $\varphi_{a+1}(\tau + \tau_0, t)$) for some $\tau_0 \in \mathbb{R}$.

[8] Containing $z_\pm$, see Definition 3.2.

(b) Let z be a non-transit singular point contained in $\Sigma_{\rm v}$. Since $\Sigma_{\rm v}$ is a sphere, there exists a biholomorphic map

$$\phi : \Sigma_{\rm v} \cong \mathbb{C} \cup \{\infty\}$$

such that $\phi(z) = 0$. Then the coordinate around z given by the local trivialization data is

$$(\tau, t) \mapsto \phi^{-1}(e^{-2\pi(\tau+\sqrt{-1}t)}),$$

for some choice of ϕ.

(4) A compact subset $K_{\rm v}^{\rm obst}$ of $\Sigma_{\rm v}$ such that $\mathcal{V}(\mathfrak{x}_{\rm v}\cup\vec{w}_{\rm v}\cup\vec{w}_{\rm can,v})\times K_{\rm v}^{\rm obst}$ is contained in a compact subset of $\pi^{-1}(\mathcal{V}_{\rm v})$ under the C^∞ trivialization $\phi_{\rm v}$: We assume that $\cup_{\rm v\in Irr(\Sigma)}K_{\rm v}^{\rm obst}$ is invariant under the $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)$ action. We call $K_{\rm v}^{\rm obst}$ the *support of the obstruction bundle*.

(5) A $\mathfrak{y}_{\rm v} \in \mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec{w}_{\rm v} \cup \vec{w}_{\rm can,v})$-parametrized smooth family of finite dimensional complex linear subspaces $E_{\mathbf{p},{\rm v}}(\mathfrak{y}_{\rm v}) \subset C_0^\infty(\,\mathrm{Int}\,K_{\rm v}^{\rm obst}; u^*TX \otimes \Lambda^{0,1}\Sigma_{\mathfrak{y}_{\rm v}})$: Here C_0^∞ denotes the set of smooth sections with compact support in $\mathrm{Int}\,K_{\rm v}^{\rm obst}$. We also regard $K_{\rm v}^{\rm obst}$ as a subset of $\Sigma_{\mathfrak{y}_{\rm v}}$ by using the C^∞ trivialization of the analytic family of coordinates given as a part of the local trivialization data. (Here $\Sigma_{\mathfrak{y}_{\rm v}}$ is the source curve of $\mathfrak{y}_{\rm v}$.)

We assume that the direct sum $\bigoplus_{\rm v\in Irr(\Sigma)} E_{\mathbf{p},{\rm v}}(\mathfrak{y}_{\rm v})$ is invariant under the $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec{z})),u,\varphi)$ action in the sense of [20, Definition 5.5].

When $\mathfrak{y}_{\rm v} = \mathfrak{x}_{\rm v}\cup\vec{w}_{\rm v}\cup\vec{w}_{\rm can,v}$ which is the center of the neighborhood $\mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec{w}_{\rm v} \cup \vec{w}_{\rm can,v})$, we denote $E_{\mathbf{p},{\rm v}}(\mathfrak{y}_{\rm v})$ by $E_{\mathbf{p},{\rm v}}$.

(6) For each $\mathbf{p}$ we consider a linear differential operator

$$D_{\mathbf{p}}\overline{\partial}_{J,H}^{\rm whole} \;:\; \bigoplus_{\rm v\in Irr(\Sigma)} L^2_{m+1,\delta}(\Sigma_{\rm v}; u^*TX) \to \bigoplus_{\rm v\in Irr(\Sigma)} L^2_{m,\delta}(\Sigma_{\rm v}; u^*TX \otimes \Lambda^{0,1}\Sigma_{\rm v}) \tag{4.4}$$

defined as follows. Here we use the above weighted Sobolev spaces in the same way as in [16, Definition 3.4] for the case $\partial\Sigma_i = \emptyset$ there.

(a) On $L^2_{m+1,\delta}(\Sigma_{\rm v}; u^*TX)$ for ${\rm v} \in \mathrm{Irr}(\Sigma)$ being a mainstream component and ${\rm v} \in \mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec{w}_{\rm v} \cup \vec{w}_{\rm can,v})$, it is the linearized operator of Floer's equation (see (2.10) for this operator)

$$\begin{aligned} D_u^{\rm Floer} &:= D_u\overline{\partial} - \nabla_{\cdot}(J\mathfrak{X}_H) \otimes (\varphi_{\rm v}^*)^{-1}(\mathrm{d}\tau - \sqrt{-1}\mathrm{d}t) \;:\; L^2_{m+1,\delta}(\Sigma_{\rm v}; u^*TX) \\ &\to L^2_{m,\delta}(\Sigma_{\rm v}; u^*TX \otimes \Lambda^{0,1}\Sigma_{\rm v}). \end{aligned} \tag{4.5}$$

(b) On $L^2_{m+1,\delta}(\Sigma_{\rm v}; u^*TX)$ for ${\rm v} \in \mathrm{Irr}(\Sigma)$ being a bubble component and ${\rm v} \in \mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec{w}_{\rm v} \cup \vec{w}_{\rm can,v})$, it is the linearization

$$D_u\overline{\partial} \;:\; L^2_{m+1,\delta}(\Sigma_{\rm v}; u^*TX) \to L^2_{m,\delta}(\Sigma_{\rm v}; u^*TX \otimes \Lambda^{0,1}\Sigma_{\rm v}) \tag{4.6}$$

of the nonlinear Cauchy–Riemann operator.

We consider the following subspace

$$L^2_{m+1,\delta}(\mathbf{p}) := \{s = (s_{\rm v}) \in \bigoplus_{\rm v} L^2_{m+1,\delta}(\Sigma_{\rm v}; u^*TX) \mid s \,\text{satisfies the condition}\ (\heartsuit)\}$$

of the domain of the operator in (4.4), where $(\heartsuit)$ is the coincidence condition at each singular point $z \in \Sigma_{\rm v^+} \cap \Sigma_{\rm v^-}$ on Σ such that

$$(\mathrm{ev}_+^{\mathrm{v}^+})_*(s_{\mathrm{v}^+}) = (\mathrm{ev}_-^{\mathrm{v}^-})_*(s_{\mathrm{v}^-})$$

for a transit point $z = z_{\mathrm{v}^+,+} = z_{\mathrm{v}^-,-}$ and

$$(\mathrm{ev}_z^{\mathrm{v}^+})_*(s_{\mathrm{v}^+}) = (\mathrm{ev}_z^{\mathrm{v}^-})_*(s_{\mathrm{v}^-})$$

for a non-transit point z. Here at a transit point the maps

$$(\mathrm{ev}_\pm^{\mathrm{v}^\pm})_* \; : \; L^2_{m+1,\delta}(\Sigma_{\mathrm{v}}; u^*TX) \to T_{\gamma^\pm_{\mathrm{v}^\pm}} \mathrm{Per}(H)$$

are differentials of the evaluation maps as in (3.7) and

$$\gamma^\pm_{\mathrm{v}^\pm} = \lim_{\tau \to \pm\infty} u(\varphi_{\mathrm{v}^\pm}(\tau, t)).$$

Restricting the domain to the subspace $L^2_{m+1,\delta}(\mathbf{p})$, we have a Fredholm operator

$$D_{\mathbf{p}}\overline{\partial}_{J,H} \; : \; L^2_{m+1,\delta}(\mathbf{p}) \to \bigoplus_{\mathrm{v}} L^2_{m,\delta}(\Sigma_{\mathrm{v}}; u^*TX \otimes \Lambda^{0,1}\Sigma_{\mathrm{v}}) \tag{4.7}$$

and call it the *linearization operator at* $\mathbf{p}$. Then in this item we require that the linearization operator $D_{\mathbf{p}}\overline{\partial}_{J,H}$ is surjective mod $\oplus_{\mathrm{v}} E_{\mathbf{p},\mathrm{v}}$, and $\mathrm{Aut}^+((\Sigma, (z_-, z_+, \vec{z})), u, \varphi)$ acts on $(D_{\mathbf{p}}\overline{\partial}_{J,H})^{-1}(\oplus_{\mathrm{v}} E_{\mathbf{p},\mathrm{v}})$ effectively. Here $E_{\mathbf{p},\mathrm{v}}$ is introduced at the end of (5) above.

(7) For each $w_i \in \vec{w} \in \Sigma$ we take a codimension 2 submanifold $\mathcal{D}_i$ of X such that $u(w_i) \in \mathcal{D}_i$ and

$$u_* T_{w_i}\Sigma + T_{u(w_i)}\mathcal{D}_i = T_{w_i} X.$$

Moreover, the set $\{\mathcal{D}_i\}$ is invariant under the $\mathrm{Aut}^+((\Sigma, (z_-, z_+, \vec{z})), u, \varphi)$ action in the following sense. Let $v \in \mathrm{Aut}^+((\Sigma, (z_-, z_+, \vec{z})), u, \varphi)$ and $v(w_i) = w_{\sigma(i)}$ then

$$\mathcal{D}_i = \mathcal{D}_{\sigma(i)}.$$

(Note $u(w_i) = u(w_{\sigma(i)})$ since $u \circ v = u$.) We note that we do *not* take such submanifolds for the canonical marked points $\in \vec{w}_{\mathrm{can}}$. (In fact, since u is not necessarily an immersion at the canonical marked points, we can not choose such submanifolds.)

Remark 4.12. (1) The condition (6) and the coincidence condition $(\heartsuit)$ therein imply that the restrictions of differential of the evaluation map $\oplus_{z,\mathrm{v}}(\mathrm{ev}_z^{\mathrm{v}})_*$, where z is any singular point, to $(D_u^{\mathrm{Floer}})^{-1}(\oplus_{\mathrm{v}} E_{\mathbf{p},\mathrm{v}})$ and $(D\overline{\partial}_u)^{-1}(\oplus_{\mathrm{v}} E_{\mathbf{p},\mathrm{v}})$ are surjective. Here D_u^{Floer} and $D\overline{\partial}_u$ are the linearized operators in (4.5) and (4.6), respectively.

(2) In Item (5), we consider the family of finite dimensional complex subsapaces $E_{\mathbf{p},\mathrm{v}}(\mathfrak{y}_{\mathrm{v}})$ over $\mathfrak{y}_{\mathrm{v}} \in \mathcal{V}(\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can},\mathrm{v}})$, while we impose the condition in (6) only at $\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can},\mathrm{v}}$. However, if we start with the finite dimensional complex subspace $E_{\mathbf{p},\mathrm{v}}$ in (6), we can obtain a smooth family of finite dimensional complex subspaces $E_{\mathbf{p},\mathrm{v}}(\mathfrak{y}_{\mathrm{v}})$ of $C_0^\infty(\mathrm{Int}\, K_{\mathrm{v}}^{\mathrm{obst}}; u^*TX \otimes \Lambda^{0,1}\Sigma_{\mathfrak{y}_{\mathrm{v}}})$ at $\mathfrak{y}_{\mathrm{v}}$ in the neighborhood $\mathcal{V}(\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can},\mathrm{v}})$ satisfying the property in (6) as well using the composition of

the inclusion map $\Lambda^{0,1}_x(\Sigma_{\rm v}) \to \Lambda^1_x(\Sigma_{\rm v})\otimes\mathbb{C}$ and the projection $\Lambda^1_x(\Sigma_{\rm v}) \to \Lambda^{0,1}_x(\Sigma_{\mathfrak{y}_{\rm v}})$ for $x \in K^{\rm obst}_{\rm v}$.

Note in [13, page 204 line 6 to 4 from the bottom] we wrote

<< We require that the data $K^{\rm obst}_{\rm v}$, $E_{\mathfrak{p},{\rm v}}(\mathfrak{y}, u)$ depend only on the mainstream component $\mathfrak{p}_i = [(\Sigma_i, z_{i-1}, z_i), u, \varphi]$ (where z_i is the i-th transit point) that contains the v-th irreducible component. We call this condition *mainstream-component-wise.* >>

This point is very much related to the main theme of the study in Sect. 5, which is a construction of Kuranishi structure compatible with the fiber product description of the boundary and corners. As we mentioned in Remark 4.2, the Kuranishi structure we construct in this section is *not* compatible with fiber product description of the boundary and corners. The same proof given in [20, Theorem 11.1] (see also [13, Lemma 17.11]) yields the following.

Lemma 4.13. *For each* $\mathbf{p} \in \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ *there exist obstruction bundle data* $\mathfrak{E}_{\mathbf{p}}$ *centered at* $\mathbf{p}$ *in the sense of Definition* 4.11.

4.3. Smoothing singularities and ϵ-closeness

We put

$$\mathfrak{x} = [\Sigma, (z_-, z_+, \vec{z}), \varphi] \in \mathcal{M}_\ell(\text{source}).$$

Let $\mathfrak{x}_{\rm v}$ be an irreducible component $\Sigma_{\rm v}$ of Σ together with marked and singular points on it. It is an element of an appropriate moduli space of marked curves of genus zero if $\Sigma_{\rm v}$ is a bubble component. It is in $\mathcal{M}_m(\text{source})$ if it is a mainstream component. (More precisely, they may not be stable. They become stable after we add $\vec{w}_{\rm v} \cup \vec{w}_{{\rm can},{\rm v}}$ that are parts of $\vec{w} \cup \vec{w}_{\rm can}$ on this irreducible component.)

We recall from [20, Lemma 3.9][9] the way how to smooth the singularity of the curve Σ and fix the local trivialization of the universal family (outside the node). Namely it determines a map:

$$\begin{aligned}\Phi_{\mathbf{p}} : \prod_{\rm v} \mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec{w}_{\rm v} \cup \vec{w}_{{\rm can},{\rm v}}) \times D(k; \vec{T}_0) \times \prod_{j=1}^{m} \left(((T_{0,j}, \infty] \times S^1)/\sim\right) \\ \to \mathcal{M}_{\ell+\ell'+\ell''}(\text{source})\end{aligned} \tag{4.8}$$

that is a homeomorphism onto an open neighborhood of $[\mathfrak{x} \cup \vec{w} \cup \vec{w}_{\rm can}]$ in $\mathcal{M}_{\ell+\ell'+\ell''}(\text{source})$. (See Definition 3.8 for this notation.) Here $\ell' = \#\vec{w}$ and $\ell'' = \#\vec{w}_{\rm can}$. The map $\Phi_{\mathbf{p}}$ is defined in [13, (31.4)]. We recall its definition together with the other notations appearing in (4.8) from [13, Section 30] below:

[9] [20, Lemma 3.9] treats the case of bordered stable curves with interior and boundary marked points. In our case, smoothing at a non-transit point corresponds to that at an interior marked point which has a two dimensional parameter space. On the other hand, at a transit point the parameter space of smoothing is only one dimensional, because the equation on the main stream component is not S^1 invariant of the rotational action. In this way smoothing at a transit point can be treated similarly to the case of smoothing at a boundary marked point of a bordered stable curve described in [20].

- k is the number of transit points except $\{z_-, z_+\}$ of Σ.
- A manifold with corner $\widetilde{D}(k;\vec{T}_0)$ is defined as follows. For any $\vec{T}_0 = (T_{0,(1)}, \dots, T_{0,(k)}) \in \mathbb{R}^k_{>1}$ we put

$$\overset{\circ}{\widetilde{D}}(k;\vec{T}_0) = \{(T_1, \dots, T_{k+1}) \in \mathbb{R}^{k+1} \mid T_{a+1} - T_a > T_{0,(a)}\} \tag{4.9}$$

and partially compactify it to $\widetilde{D}(k;\vec{T}_0)$ by admitting $T_{a+1} - T_a = \infty$ as follows. We put $s'_a = 1/\log(T_{a+1} - T_a)$. Then T_1 and $s'_1, \dots, s'_{k-1}$ define another parameters. So (4.9) is identified with $\mathbb{R} \times \prod_{i=1}^k (0, 1/\log T_{0,(a)})$. We partially compactify it to $\mathbb{R} \times \prod_{a=1}^k [0, 1/\log T_{0,(a)})$. By taking the quotients of $\overset{\circ}{\widetilde{D}}(k;\vec{T}_0)$ and $\widetilde{D}(k;\vec{T}_0)$ by the $\mathbb{R}$ action $T(T_1, \dots, T_{k+1}) = (T_1 + T, \dots, T_{k+1} + T)$, we obtain $\overset{\circ}{D}(k;\vec{T}_0)$ and $D(k;\vec{T}_0)$, respectively.

Remark 4.14. We take the logarithm of $T_{a+1} - T_a$ to define our coordinate s'_a. By doing so we can stay in the category of admissible orbifolds or admissible Kuranishi structures in the sense of [23, Chapter 25]. We also take $e^{2\pi t\sqrt{-1}}/\log T$ for the coordinate of the $((T_0, \infty] \times S^1)/\sim$ factor. This is a slightly different choice from [11, Appendix A1.4].

- The space $D(k;\vec{T}_0)$ is used to parametrize the ways of smoothing the transit point singularities as follows. We consider the case when the parameter $\vec{T}$ is in $\overset{\circ}{D}(k;\vec{T}_0)$. Taking a section of the projection $\widehat{\mathcal{M}}_*(\text{source}) \to \mathcal{M}_*(\text{source})$, we have a parametrization $\varphi_a : \mathbb{R} \times S^1 \to \Sigma_a$ for each mainstream component Σ_a with $a = 1, \dots, k+1$ as in Definition 3.3 (4). Let us consider

$$[-5T_{0,(a-1)}, 5T_{0,(a)}]_a \times S^1_a$$

where $T_{0,(0)} = T_{0,(k+1)} = +\infty$ as convention, and regard it as a subset of the domain of the parametrization $\varphi_a : \mathbb{R} \times S^1 \to \Sigma_a$. We define

$$\varphi_0 : \bigcup_a ([-5T_{0,(a-1)}, 5T_{0,(a)}]_a \times S^1_a) \to \mathbb{R} \times S^1$$

as follows. If $(\tau, t) \in [-5T_{0,(a-1)}, 5T_{0,(a)}]_a \times S^1_a$, then

$$\varphi_0(\tau, t) = (\tau + 10T_a, t).$$

We use $\varphi_0 \circ \varphi_a^{-1}$ to identify (a part of) Σ_a with a subset of $\mathbb{R} \times S^1$. Under this identification marked points on Σ_a can be moved to $\mathbb{R} \times S^1$. Adding z_-, z_+, we have a marked Riemann surface. Taking the equivalence class by $\sim_2$ in Definition 3.7, we obtain an element of $\mathcal{M}_*(\text{source})$ so the map $\Phi_{\mathbf{p}}$ in the case $m = 0$. See Fig. 3.
- m in (4.8) is the number of non-transit singular points. The factor $((T_{0,j}, \infty] \times S^1)/\sim$ $(j = 1, \dots, m)$ is the space to parametrize the way to smooth these singular points. Here $\sim$ is the equivalence relation such that $(T, t) \sim (T', t')$ if and only if $T = T' = \infty$ or $(T, t) = (T', t')$. The way to use this parameter to smooth non-transit singular points is written in [20, Lemma 3.9]

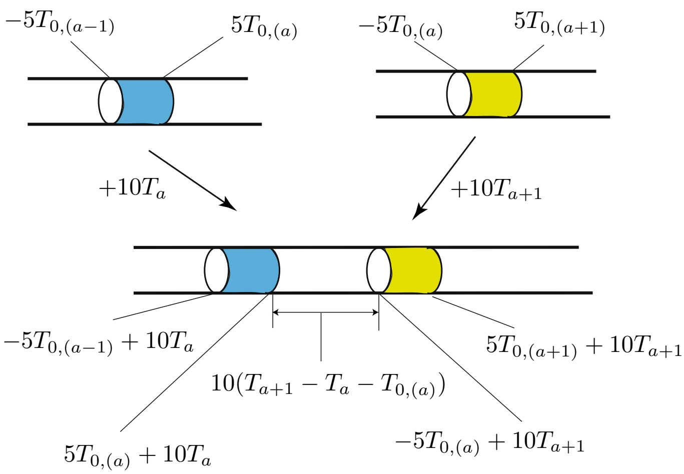


FIGURE 3. The map φ_0

We have thus defined all notations appearing in (4.8).

Notation 4.15. Suppose

$$(\mathfrak{Y} \cup \vec{w}', \varphi') = \Phi_{\mathbf{p}}(\mathfrak{y}, \vec{T}, \vec{\theta}) \in \mathcal{M}_{\ell+\ell'+\ell''}(\text{source}). \tag{4.10}$$

Here

- $\mathfrak{y} = (\mathfrak{y}_{\rm v})$ where v is the index in the set of irreducible components of Σ in $\mathbf{p}$ as in Definition 4.11.
- $\vec{w}'$ is the set of the additional marked points corresponding to $\vec{w}$ and $\vec{w}_{\rm can}$.
- The notation $\mathfrak{Y}$ includes the marked points corresponding to $\vec{z}$ and $z_{\pm}$.
- The pair of parameters $(\vec{T}, \vec{\theta}) \in D(k; \vec{T}_0) \times \prod_{j=1}^{m} \left(((T_{0,j}, \infty] \times S^1)/\sim\right)$ and the map φ' is a parametrization of the mainstream of Σ'. Here Σ' is the source curve of $\mathfrak{Y}$.

Of course, the left hand side depends on $\mathbf{p}$ and $\mathfrak{y}, \vec{T}, \vec{\theta}$ in the right hand side also depend on $\mathbf{p}$ as well as the left hand side.

In the situation of Notation 4.15, since Σ' is obtained from the source curves $\Sigma_{\mathfrak{y}_{\rm v}}$ by first removing neighborhoods of singular points and then gluing them, there exists an embedding

$$\mathfrak{v}_{\mathfrak{Y},\mathfrak{y};{\rm v}} : K_{\rm v}^{\rm obst} \to \Sigma'. \tag{4.11}$$

Actually the embedding to Σ' is defined in a larger region called the *core* of the source curve $\Sigma_{\mathfrak{y}}$. (It is the complement of the *neck region*. See [20, Definition 4.12] for the definition of neck region.)

Definition 4.16. Let $(\mathfrak{Y} \cup \vec{w}', \varphi') = \Phi_{\mathbf{p}}(\mathfrak{y}, \vec{T}, \vec{\theta}) \in \mathcal{M}_{\ell+\ell'+\ell''}(\text{source})$ and $u' : \Sigma' \setminus \{\text{transit points}\} \to X$. We assume that $(\mathfrak{Y}, u', \varphi')$ satisfies Definition 3.3 (1)(2)(3)(7) and (8). We say that $(\mathfrak{Y}, u', \varphi') \cup \vec{w}'$ is *ϵ-close* to $\mathbf{p} \cup \vec{w} \cup \vec{w}_{\rm can}$ if the following holds.

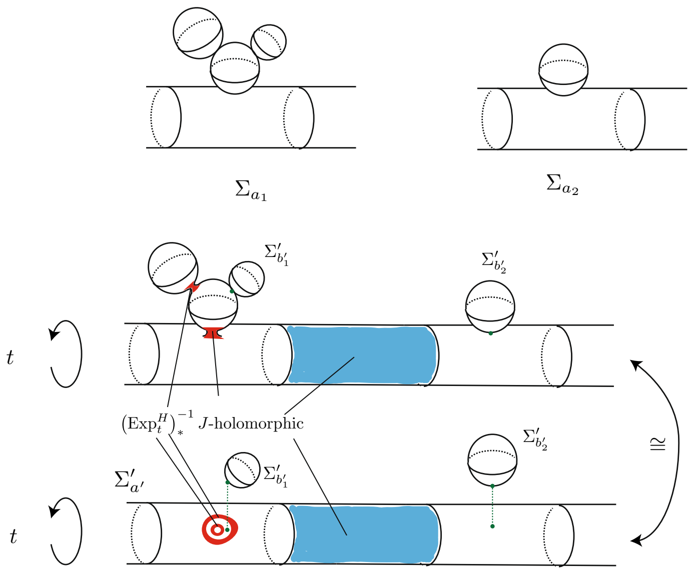


FIGURE 4. Neck regions on mainstream components

(1) $\|u' \circ \mathfrak{v}_{\mathfrak{Y},\mathfrak{y};\mathrm{v}} - u\| < \epsilon$ on the core of $\Sigma_{\mathfrak{y}}$. Here $\|\cdot\|$ is the C^{10} norm.
(2) The map $u' \circ \varphi'$ satisfies the Eq. (2.4) in the neck regions corresponding to transit points. For a bubble component $\Sigma'_{b'}$, there is a non-transit point on a mainstream component $\Sigma'_{a'}$ such that $\Sigma'_{b'}$ is joined to $\Sigma'_{a'}$ by a tree of bubble components. Then the map $u'|_{\Sigma'_{b'}}$ is J-holomorphic on some neighborhoods of nodes. When a non-transit point on a mainstream component $\Sigma'_{a'}$ is smoothed, $u' \circ \varphi'$ satisfies the Eq. (2.4) on the corresponding neck region on the mainstream component. See Fig. 4.
(3) Let $\hat{u}'$ be the redefined connecting orbit map of u' (see Definition 3.15). For a non-transit point $\varphi_{a'}(\tau_0, t_0)$ on a mainstream component $\Sigma'_{a'}$, set $u'^{\#}(z) = (\exp_t^H)^{-1}(u(z))$ for $z = \varphi_a(\tau, t)$, $t_0 - 1/2 < t < t_0 + 1/2$. Then for each connected component $\mathfrak{W}$ of the complement of the core, we have either
$$\mathrm{Diam}(\hat{u}'(\mathfrak{W})) < \epsilon, \quad \text{for } \mathfrak{W} \text{ around transit points,} \tag{4.12}$$
or
$$\mathrm{Diam}(u'^{\#}(\mathfrak{W})) < \epsilon, \quad \text{for } \mathfrak{W} \text{ around non-transit points.} \tag{4.13}$$
(4) For each component $T_{0,(a)}$ of $\vec{T}_0$ we have $T_{0,(a)} > \epsilon^{-1}$ and $T_{0,j} > \epsilon^{-1}$ for any $j = 1, \dots, m$.

Remark 4.17. Definition 4.16 is the same as the definition of ϵ-closeness appearing in [13, page 215], [20, Definition 4.12]. Note that we do not assume Condition (5)' in [13, page 215], since it is a consequence of Definition 4.16 (1) above, and so is unnecessary to be assumed.

Situation 4.18. Let $\mathbf{p} \in \mathcal{M}_{\ell+\ell'+\ell''}(X, H; \alpha_-, \alpha_+)$. We fix obstruction bundle data $\mathfrak{C}_{\mathbf{p}}$ centered at $\mathbf{p}$. Let $(\mathfrak{Y} \cup \vec{w}', \varphi') = \Phi_{\mathbf{p}}(\mathfrak{y}, \vec{T}, \vec{\theta}) \in \mathcal{M}_{\ell+\ell'+\ell''}(\text{source})$ and $u' : \Sigma' \to X$, where Σ' is the source curve of $\mathfrak{Y}$ as in Notation 4.15. We assume that $(\mathfrak{Y}, u', \varphi') \cup \vec{w}'$ is ϵ-close to $\mathbf{p} \cup \vec{w} \cup \vec{w}_{\text{can}}$. □

Definition 4.19. Suppose we are in Situation 4.18. We say that $(\mathfrak{Y}, u', \varphi') \cup \vec{w}'$ satisfies the *transversal constraint* if the following holds.

(1) If w'_i corresponds to $w_i \in \vec{w}$, then $u'(w'_i) \in \mathcal{D}_i$.
(2) Let $w'_j = \varphi'_{a'}(\tau'_j, t'_j) \in \vec{w}' \cap \Sigma'_{a'}$ be the marked point corresponding to the canonical marked point $w_{a,\text{can}} = \varphi_a(\tau_a, 0) \in \vec{w}_{\text{can}}$. Then we require:

$$f_{H,u',\Sigma'_{a'}}(\tau'_j) = f_{H,u,\Sigma_a}(\tau_a) = \frac{1}{2}\left(\lim_{\tau\to-\infty} f_{H,u,\Sigma_a}(\tau) + \lim_{\tau\to+\infty} f_{H,u,\Sigma_a}(\tau)\right). \tag{4.14}$$

Here f_{H,u,Σ_a} is the function (4.1) for the mainstream component Σ_a of Σ.
(3) In the situation of (2) we require also $t'_j = [0]$.

Remark 4.20. (1) The second equality of (4.14) is the definition of τ_a (see Lemma 4.7). So the actual condition is the first equality.
(2) Note that $\Sigma'_{a'}$ may have a sphere bubble (or may contain one of the marked points of $\mathfrak{Y}$) even in the case when Σ_a has no sphere bubble (or does not contain one of the marked points of $\vec{z}$). In fact, Σ_a may be glued with other mainstream component that has a sphere bubble when we obtain Σ' form Σ. Therefore, $\Sigma'_{a'}$ may not have a canonical marked point.
(3) Even in the case when $\Sigma'_{a'}$ has a canonical marked point, it may be different from w'_j. In fact $\Sigma'_{a'}$ may be obtained by gluing several mainstream components which have no sphere bubbles or points of $\vec{z}$.

Suppose we are in Situation 4.18. Let $z \in K^{\text{obst}}_{\text{v}}$. Recall from Definition 4.11 (5) that $K^{\text{obst}}_{\text{v}}$ contains a support of elements of $E_{\mathbf{p},\text{v}}(\mathfrak{y})$. Let $\mathfrak{v}_{\mathfrak{Y},\mathfrak{y};\text{v}} : K^{\text{obst}}_{\text{v}} \to \Sigma'$ be an embedding as in (4.11). By ϵ-closeness we have

$$d(u'(\mathfrak{v}_{\mathfrak{Y},\mathfrak{y};\text{v}}(z)), u(z)) < \epsilon.$$

We may choose $\epsilon > 0$ smaller than the injectivity radius of X. Therefore, there exists a unique minimal geodesic ℓ_z in X joining $u(z)$ with $u'(\mathfrak{v}_{\mathfrak{Y},\mathfrak{y};\text{v}}(z))$. The complex linear part of the parallel transport along ℓ_z defines a complex linear map

$$\text{Pal}_z : T_{u(z)}X \to T_{u'(\mathfrak{v}_{\mathfrak{Y},\mathfrak{y};\text{v}}(z))}X. \tag{4.15}$$

Definition 4.21. Suppose we are in Situation 4.18. Using the map in (4.15), we have a complex linear embedding

$$I_{\mathbf{p},\text{v};\Sigma',u',\varphi'} : E_{\mathbf{p},\text{v}}(\mathfrak{y}) \to C^\infty(\Sigma'; (u')^*TX \otimes \Lambda^{0,1}).$$

In Definition 4.25 we will define an obstruction space $E((\mathfrak{Y}\cup\bigcup_{c\in\mathcal{B}}\vec{w}'_c, u', \varphi'); \mathbf{q}; \mathcal{B})$ (see also Definition 4.25 for the notations used in this notation) for any $\mathbf{q}\in\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ as a sum of images of the maps in Definition 4.21 for suitable choices of $\mathbf{p}$'s. To carry out this argument we first observe the following.

Lemma 4.22. *Suppose we are in Situation 4.18. Then for any $\mathbf{p}\in\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ there exist $\epsilon_{\mathbf{p}}>0$ and a closed neighborhood $W(\mathbf{p})$ of $\mathbf{p}$ in $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ such that for any $\mathbf{q}\in W(\mathbf{p})$ there exists $\vec{w}^{\mathbf{q}}_{\mathbf{p}}$* uniquely *with the following properties:*

(1) *$\mathbf{q}\cup\vec{w}^{\mathbf{q}}_{\mathbf{p}}$ is $\epsilon_{\mathbf{p}}$-close to $\mathbf{p}\cup w_{\mathbf{p}}\cup\vec{w}_{\mathrm{can}}$.*
(2) *$\mathbf{q}\cup\vec{w}^{\mathbf{q}}_{\mathbf{p}}$ satisfies the transversal constraint.* (Definition 4.19.)
(3) *The linearization operator $D_{\mathbf{q}}\overline{\partial}_{J,H}$ at $\mathbf{q}$ in (4.7) is surjective* $\mod \oplus_{\mathrm{v}}$ $\mathrm{Im}\ I_{\mathbf{p},\mathrm{v};\mathbf{q}}$, *where $I_{\mathbf{p},\mathrm{v};\mathbf{q}}$ is the map in Definition 4.21 for the case $\mathbf{q}=(\Sigma',u',\varphi')$.*

Proof. If $w_{\mathbf{p},i}\in\vec{w}_{\mathbf{p}}$, the unique existence of $w^{\mathbf{q}}_{\mathbf{p},i}$ satisfying Definition 4.19 (1) can be proved in the same way as in [20, Lemma 9.9].

In case $w^{\mathbf{q}}_{\mathbf{p},i}$ corresponds to a canonical marked point, the unique existence of $w^{\mathbf{q}}_{\mathbf{p},i}$ satisfying Definition 4.19 (2)(3) is a consequence of the following two facts:

(i) $u_{\mathbf{q}}$ is C^1 close to $u_{\mathbf{p}}$.
(ii) The first derivative of the function f_{H,u,Σ_a} in (4.1) is strictly positive at τ_i. Here $\varphi_a(\tau_i,0)$ is the canonical marked point on $\mathbf{p}$ which corresponds to $w^{\mathbf{q}}_{\mathbf{p},i}$.

Furthermore by taking $W(\mathbf{p})$ small enough, the property (3) is satisfied because the surjectivity is an open condition. □

Then we make the following choices.

Choice 4.23. We fix ℓ, α_-, α_+.

- We take a finite set
$$\mathcal{A}_\ell(\alpha_-,\alpha_+)=\{\mathbf{p}_c\mid c\in\mathcal{C}_\ell(\alpha_-,\alpha_+)\}\subset\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+).$$
Here $\mathcal{C}_\ell(\alpha_-,\alpha_+)$ is an index set which will be taken as in Remark 4.24.
- For each $c\in\mathcal{C}_\ell(\alpha_-,\alpha_+)$ we take obstruction bundle data $\mathfrak{E}_{\mathbf{p}_c}$ centered at $\mathbf{p}_c$.
- For each $c\in\mathcal{C}_\ell(\alpha_-,\alpha_+)$ we take a closed neighborhood $W(\mathbf{p}_c)$ of $\mathbf{p}_c$ in $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ with the following property. For any $\mathbf{q}\in W(\mathbf{p}_c)$ there exists $\vec{w}^{\mathbf{q}}_{\mathbf{p}_c}$ such that $\mathbf{q}\cup\vec{w}^{\mathbf{q}}_{\mathbf{p}_c}$ is ϵ_c-close to $\mathbf{p}_c\cup w_{\mathbf{p}_c}\cup\vec{w}_{\mathrm{can}}$. Here $\epsilon_c>0$ depends on c, which will be determined later. Moreover, the linearization operator $D_{\mathbf{q}}\overline{\partial}_{J,H}$ in (4.7) is surjective $\mod \oplus_{\mathrm{v}}\mathrm{Im}\ I_{\mathbf{p}_c,\mathrm{v};\mathbf{q}}$ where $\mathrm{Im}\ I_{\mathbf{p}_c,\mathrm{v};\mathbf{q}}$ is the map in Definition 4.21 for the case $\mathbf{p}=\mathbf{p}_c$, $\mathbf{q}=(\Sigma',u',\varphi')$.
- We require
$$\bigcup_{c\in\mathcal{C}_\ell(\alpha_-,\alpha_+)}\mathrm{Int}\ W(\mathbf{p}_c)=\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+). \tag{4.16}$$

Remark 4.24. The logical order to make such choices is as follows. First for each $\mathbf{p} \in \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ we take obstruction bundle data $\mathfrak{E}_{\mathbf{p}}$ by Lemma 4.13. We take $\epsilon_{\mathbf{p}} > 0$ and a closed neighborhood $W(\mathbf{p})$ as in Lemma 4.22. Then we have

$$\bigcup_{\mathbf{p}} \mathrm{Int}\ W(\mathbf{p}) = \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+).$$

Finally, by compactness of the moduli space, we can take a finite set $\mathcal{C}_\ell(\alpha_-, \alpha_+)$ such that $c \in \mathcal{C}_\ell(\alpha_-, \alpha_+)$ satisfies the properties in Choice 4.23.

Definition 4.25. (1) For each $\mathbf{q} \in \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ we put

$$\mathcal{E}(\mathbf{q}) = \{c \in \mathcal{C}_\ell(\alpha_-, \alpha_+) \mid \mathbf{q} \in W(\mathbf{p}_c)\}.$$

(2) Let $\mathcal{B} \subset \mathcal{E}(\mathbf{q})$ be a nonempty subset.
(3) We consider $(\mathfrak{Y} \cup \bigcup_{c\in\mathfrak{B}} \vec{w}'_c, u', \varphi')$ such that for each c, $(\mathfrak{Y} \cup \vec{w}'_c, u', \varphi')$ is ϵ-close to $\mathbf{q} \cup \vec{w}^{\mathbf{q}}_c$. If $\epsilon > 0$ is small, then $(\mathfrak{Y} \cup \vec{w}'_c, u', \varphi')$ is ϵ-close to $\mathbf{p}_c \cup \vec{w}_c$ and we can define the map

$$I_{\mathbf{p}_c,\mathrm{v};\Sigma',u',\varphi'} : E_{\mathbf{p}_c,\mathrm{v}}(\mathfrak{y}_{\mathbf{p}_c}) \to C^\infty(\Sigma'; (u')^*TX \otimes \Lambda^{0,1})$$

in Definition 4.21 for each irreducible component v of $\mathbf{p}_c$. Here $(\mathfrak{Y} \cup \vec{w}'_c, \varphi') = \Phi_{\mathbf{p}_c}(\mathfrak{y}_{\mathbf{p}_c}, \vec{T}_{\mathbf{p}_c}, \vec{\theta}_{\mathbf{p}_c})$ and Σ' is the source curve of $\mathfrak{Y}$ as in Notation 4.15.[10] We now put

$$E((\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c, u', \varphi'); \mathbf{q}; \mathcal{B}) = \bigoplus_{c\in\mathcal{B}} \bigoplus_{\mathrm{v}} \mathrm{Im}\ I_{\mathbf{p}_c,\mathrm{v};\Sigma',u',\varphi'}. \tag{4.17}$$

We can perturb $E_{\mathbf{p}_c,\mathrm{v}}(\mathfrak{y}_{\mathbf{p}_c})$ slightly so that the right hand side of (4.17) is a direct sum. (See [20, Lemma 11.7], [13, Lemma 18.8].)

4.4. Kuranishi chart

In this subsection we construct a Kuranishi chart for any $\mathbf{q} \in \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$. We refer [23, Definition 3.1] for the definition of Kuranishi chart.

Definition 4.26. *Stabilization data centered at* $\mathbf{q} \in \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ are

$$\left(\vec{w}, \{\mathcal{V}(\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can},\mathrm{v}})\}_{\mathrm{v}}, \{(\psi_{\mathrm{v}}, \phi_{\mathrm{v}})\}_{\mathrm{v}}, \{\mathcal{D}_i\}_{w_i\in\vec{w}}\right)$$

in Definition 4.11 (1)(2)(3) and (7), which are sub-data of the obstruction bundle data at $\mathbf{q}$.

Situation 4.27. Suppose we are in the situation of Definition 4.25. We also take stabilization data at $\mathbf{q}$. Let $(\mathfrak{Y} \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}'_c, u', \varphi')$ be as in Definition 4.25 (3) and $\vec{w}'_{\mathbf{q}}$ be additional marked points on $\mathfrak{Y}$ such that $(\mathfrak{Y} \cup \vec{w}'_{\mathbf{q}}, \varphi')$ is ϵ-close to $\mathbf{q} \cup \vec{w}_{\mathbf{q}} \cup \vec{w}_{\mathbf{q},\mathrm{can}}$. Here $\vec{w}_{\mathbf{q}}$ is the additional marked points on $\mathfrak{Y}$ taken as a part of the stabilization data centered at $\mathbf{q}$ and $\vec{w}_{\mathbf{q},\mathrm{can}}$ are canonical marked points we put on the mainstream component without sphere bubble or marked points. □

Definition 4.28. In Situation 4.27 we consider the following conditions on an object $(\mathfrak{Y} \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, u', \varphi')$:

[10] As we noticed in Notation 4.15, $\mathfrak{y}_{\mathbf{p}_c}$ etc depend on the choice of $\mathbf{p}_c$. Here we put the suffix $c \in \mathcal{B}$ in the notations to remember the dependence.

(1) If Σ_a is the mainstream component and φ'_a is a parametrization of this mainstream component (which is a part of given φ'), the following equation is satisfied on $\mathbb{R} \times S^1$.

$$\begin{aligned} &\frac{\partial(u' \circ \varphi'_a)}{\partial \tau} + J\left(\frac{\partial(u' \circ \varphi'_a)}{\partial t} - \mathfrak{X}_{H_t} \circ u' \circ \varphi'_a\right) \\ &\equiv 0 \mod E((\mathfrak{Y} \cup \bigcup_{c \in \mathcal{B}} \vec{w}'_c, u', \varphi'); \mathbf{q}; \mathcal{B}). \end{aligned} \tag{4.18}$$

(2) If v is a bubble component, the following equation is satisfied on $\Sigma'_{\rm v}$.

$$\overline{\partial} u' \equiv 0 \mod E((\mathfrak{Y} \cup \bigcup_{c \in \mathcal{B}} \vec{w}'_c, u', \varphi'); \mathbf{q}; \mathcal{B}). \tag{4.19}$$

(3) For each $c \in \mathcal{E}(\mathbf{q})$ the additional marked points $\vec{w}'_c$ satisfy the transversal constraint in Definition 4.19 with respect to the obstruction bundle data $\mathfrak{E}_{\mathbf{p}_c}$ centered at $\mathbf{p}_c$. (Namely, for each $c \in \mathcal{E}(\mathbf{q})$ $(\mathfrak{Y}, u', \varphi') \cup \vec{w}'_c$ satisfies the transversal constraint in Definition 4.19.)
(4) The additional marked points $\vec{w}'_{\mathbf{q}}$ satisfy the transversal constraint in Definition 4.19 with respect to the stabilization data centered at $\mathbf{q}$ in Situation 4.27. (Namely, $(\mathfrak{Y}, u', \varphi') \cup \vec{w}'_{\mathbf{q}}$ also satisfies the transversal constraint in Definition 4.19.)[11]
(5) $(\mathfrak{Y} \cup \bigcup_{c \in \mathcal{E}(\mathbf{q})} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, u', \varphi')$ is ϵ_1-close to $\mathbf{q} \cup \bigcup_{c \in \mathcal{E}(\mathbf{q})} \vec{w}^{\mathbf{q}}_c \cup \vec{w}_{\mathbf{q}}$.

We define an orbifold ([34])

$$V(\mathbf{q}, \epsilon_1, \mathcal{B})$$

to be the set of isomorphism classes of $(\mathfrak{Y} \cup \bigcup_{c \in \mathcal{E}(\mathbf{q})} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, u', \varphi')$ satisfying the conditions (1)–(5) above. Here $(\mathfrak{Y} \cup \bigcup_{c \in \mathcal{E}(\mathbf{q})} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, u', \varphi')$ is said to be *isomorphic* to $(\mathfrak{Y}'' \cup \bigcup_{c \in \mathcal{E}(\mathbf{q})} \vec{w}''_c \cup \vec{w}''_{\mathbf{q}}, u'', \varphi'')$ if there exists a biholomorphic map $v : \Sigma' \to \Sigma''$ with the following properties.

(a) $u'' = u' \circ v$ holds outside the set of the transit points.
(b) If Σ'_a is a mainstream component of Σ' and $v(\Sigma'_a) = \Sigma''_{a'}$, then we have $(v \circ \varphi'_a)(\tau, t) = \varphi''_{a'}(\tau + \tau_a, t)$ on $\mathbb{R} \times S^1$ where $\tau_a \in \mathbb{R}$ is independent of (τ, t).
(c) $v(z'_i) = z''_i$ and $v(w'_i) = w''_i$.

Lemma 4.29. *If $\epsilon_1 > 0$ and $\epsilon_c > 0$ are small enough, then $V(\mathbf{q}, \epsilon_1, \mathcal{B})$ is a smooth manifold with boundary. Its dimension is*

$$\dim \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+) + \sum_{c \in \mathcal{B}} \sum_{{\rm v} \in {\rm Irr}(\mathbf{p}_c)} {\rm rank}\, E_{\mathbf{p}_c, {\rm v}}(\mathfrak{y}_{\mathbf{p}_c}). \tag{4.20}$$

Here $\mathrm{Irr}(\mathbf{p}_c)$ *is the set of irreducible components of* $\mathbf{p}_c$.

If $\mathbf{q}$ has $k+1$ mainstream components, then the element $[\mathbf{q} \cup \bigcup_{c \in \mathcal{E}(\mathbf{q})} \vec{w}^{\mathbf{q}}_c \cup \vec{w}_{\mathbf{q}}]$ of $V(\mathbf{q}, \epsilon_1, \mathcal{B})$ is in a codimension k corner.

[11]We take the stabilization data for $\mathbf{q}$ in Situation 4.27. This is enough to define the transversal constraint.

Proof. We first consider the set of isomorphism classes of $(\mathfrak{Y} \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, u', \varphi')$ satisfying Definition 4.28 (1)(2)(5) and denote it by $\widehat{V}(\mathbf{q}, \epsilon_1, \mathcal{B})$.[12]

We can prove that $\widehat{V}(\mathbf{q}, \epsilon_1, \mathcal{B})$ is a smooth manifold with boundary and corner in the same way as in [16, Chapter 8]. We use the map

$$\Phi_{\mathbf{q}} : \prod \mathcal{V}((\mathfrak{x}_{\mathbf{q}})_{\rm v} \cup \vec{w}_{\mathbf{q},\rm v} \cup \vec{w}_{\mathbf{q},{\rm can},\rm v}) \times D(k; \vec{T}_0) \times \prod_{j=1}^{m} \left(((T_{0,j}, \infty] \times S^1)/\sim\right) \\ \to \mathcal{M}_{\ell+\ell'+\ell''}(\text{source}) \tag{4.21}$$

to work out the gluing analysis in [16, Chapters 5,6]. (The map (4.21) is the same map as (4.8) except we use the stabilization data at $\mathbf{q}$.) (4.21) parametrizes the source curve (plus marked points) of elements of $\widehat{V}(\mathbf{q}, \epsilon_1, \mathcal{B})$. For each fixed source curve we can perform the gluing construction as in [16, Chapters 5,6][13] to find that $\widehat{V}(\mathbf{q}, \epsilon_1, \mathcal{B})$ is a smooth manifold strata-wise. To obtain a smooth structure on the union of the strata, we use the exponential decay estimate which we can prove in the same way as in [16, Chapter 8].[14] The way to use this exponential decay estimate is the same as in [20, Sections 9 and 10], [16, Chapter 8].

We note that the only difference for the gluing analysis in the current situation is the presence of the Hamiltonian term $\mathfrak{X}_{H_t} \circ u' \circ \varphi'_{\rm v}$. This term is also nonzero on the neck region where we glue two solutions. However, the Hamiltonian term is small in the exponential order on the neck region. We can prove it easily by looking at the coordinate change from $S^1 \times [0, \infty)$ to $D^2 \setminus \{0\}$. Namely the derivatives of this map decays exponentially as the second component of the domain goes to infinity. (See for example [13, Lemma 30.24] for the detail.) So it does not affect the argument here.

Once we proved that $\widehat{V}(\mathbf{q}, \epsilon_1, \mathcal{B})$ is a smooth manifold, we can prove that $V(\mathbf{q}, \epsilon_1, \mathcal{B})$ is a smooth manifold by the implicit function theorem. In fact, Definition 4.19 (1) cuts out a smooth submanifold transversally because of the implicit function theorem [23, Lemma 25.32]. (See also [13, Section 20], especially Lemma 20.7.) We can use the facts (i)(ii) appearing in the proof of Lemma 4.22 to show that Definition 4.19 (2)(3) cut out a smooth submanifold transversally. We have thus proved that $V(\mathbf{q}, \epsilon_1, \mathcal{B})$ is a smooth manifold.[15]

We assume that the source curve of $\mathbf{q}$ has exactly k mainstream components. Note that the space $D(k; \vec{T}_0)$ in (4.8) is a manifold with boundary. The point corresponding to the source curve of $[\mathbf{q} \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}^{\mathbf{q}}_c \cup \vec{w}_{\mathbf{q}}]$ corresponds

[12] A similar moduli space appeared in [13, Definition 18.15] and was called the thickened moduli space. The Hamiltonian term $\mathfrak{X}_{H_t} \circ u' \circ \varphi'_{\rm v}$ did not appear there.

[13] See [13, Part 3] (simple case), [13, Section 19] (the general case) for more detailed explanation if necessary.

[14] See [13, Theorem 13.2] (simple case) or [13, Theorem 19.5] (general case) for more details.

[15] In [13, Sections 19, 20, 21] we first cut out by the transversality constraint strata-wise and then show that those strata-wise smooth structure gives the smooth structure on the whole space. So the order of the proof there is slightly different from one we mention above. There is no mathematical issue on this point and we can do either way. The order is changed only for the convenience of the exposition.

to the codimension k boundary point of $D(k;\vec{T}_0)$. In fact, it corresponds to the point $(T,\infty,\dots,\infty)$ on the compactification of the image of the map $(T_0,\dots,T_k)\mapsto(T_0,T_1-T_0,\dots,T_k-T_{k-1})$. See the discussion right before Fig. 3. Therefore, $[\mathbf{q}\cup\bigcup_{c\in\mathcal{E}(\mathbf{q})}\vec{w}_c^{\mathbf{q}}\cup\vec{w}_{\mathbf{q}}]$ is on the codimension k boundary of $V(\mathbf{q},\epsilon_1,\mathcal{B})$.

The dimension formula follows from Lemma 2.15. □

We note that the group $\mathrm{Aut}^+(\mathbf{q})$ acts on $V(\mathbf{q},\epsilon_1,\mathcal{B})$ since the stabilization data are assumed to be preserved. In particular, $\mathrm{Aut}(\mathbf{q})$ acts on it. We also note that by the condition in Definition 4.11 (6) this action is effective. Therefore, the quotient space $V(\mathbf{q},\epsilon_1,\mathfrak{B})/\mathrm{Aut}(\mathbf{q})$ is an effective orbifold, which we denote by $U(\mathbf{q},\epsilon_1,\mathcal{B})$.

We define a vector bundle

$$E(\mathbf{q},\epsilon_1,\mathcal{B})\to U(\mathbf{q},\epsilon_1,\mathcal{B})$$

whose fiber at $(\mathfrak{Y}\cup\bigcup_{c\in\mathcal{E}(\mathbf{q})}\vec{w}'_c\cup\vec{w}'_{\mathbf{q}},u',\varphi')$ is $E((\mathfrak{Y}\cup\bigcup_{c\in\mathcal{B}}\vec{w}'_c,u',\varphi');\mathbf{q};\mathcal{B})$. We define its section $s_{(\mathbf{q},\epsilon_1,\mathcal{B})}$ by

$$s_{(\mathbf{q},\epsilon_1,\mathcal{B})}(\mathfrak{Y}\cup\bigcup_{c\in\mathcal{E}(\mathbf{q})}\vec{w}'_c\cup\vec{w}'_{\mathbf{q}},u',\varphi')$$
$$=\begin{cases}\overline{\partial}u' & \text{on a bubble component }\Sigma_{\mathrm{v}},\\ \frac{\partial(u'\circ\varphi'_{\mathrm{v}})}{\partial\tau}+J\left(\frac{\partial(u'\circ\varphi'_{\mathrm{v}})}{\partial t}-\mathfrak{X}_{H_t}\circ u'\circ\varphi'_{\mathrm{v}}\right) & \text{on a mainstream component }\Sigma_{\mathrm{v}}.\end{cases}$$

Note that the right hand side is an element of $E(\mathbf{q},\epsilon_1,\mathcal{B})$ by Definition 4.28.

By definition if $s_{(\mathbf{q},\epsilon_1,\mathcal{B})}(\mathfrak{Y}\cup\bigcup_{c\in\mathcal{E}(\mathbf{q})}\vec{w}'_c\cup\vec{w}'_{\mathbf{q}},u',\varphi')=0$, then $(\mathfrak{Y},u',\varphi')$ represents an element of $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$. We thus obtain a map

$$\psi_{(\mathbf{q},\epsilon_1,\mathcal{B})}:s^{-1}_{(\mathbf{q},\epsilon_1,\mathcal{B})}(0)\to\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+).$$

Proposition 4.30. *If $\epsilon_1>0$ is small, then $(U(\mathbf{q},\epsilon_1,\mathcal{B}),E(\mathbf{q},\epsilon_1,\mathcal{B}),s_{(\mathbf{q},\epsilon_1,\mathcal{B})},\psi_{(\mathbf{q},\epsilon_1,\mathcal{B})})$ is a Kuranishi chart of $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ at $\mathbf{q}$.*

Proof. Taking into account of the point mentioned in the proof of Lemma 4.29, the proof is the same as in [16, Chapter 8]. □

Lemma 4.31. *We assume that $\mathbf{q}\in S_k(\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+))$. Then $S_k(V(\mathbf{q},\epsilon_1,\mathcal{B}))$ is the set of $(\mathfrak{Y}\cup\bigcup_{c\in\mathcal{E}(\mathbf{q})}\vec{w}'_c\cup\vec{w}'_{\mathbf{q}},u',\varphi')\in V(\mathbf{q},\epsilon_1,\mathcal{B},u)$ such that $\mathfrak{Y}$ has at least $k+1$ mainstream components.*

Proof. As explained in the proof of Lemma 4.29, the codimension k corner of $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ corresponds to the codimension k corner of the $D(k;\vec{T}_0)$ factor of the left hand side of (4.8). Lemma 4.31 immediately follows from this fact. □

We also observe the following fact.

Lemma 4.32. *If $\mathbf{q}\in S_k(\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+))$ and $c\in\mathcal{E}(\mathbf{q})$, then we have $\mathbf{p}_c\in S_k(\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+))$.*

Proof. This follows from the following fact. If $\mathbf{q}$ is ϵ-close to $\mathbf{p}$, then the number of mainstream components of $\mathbf{q}$ is not greater than the number of mainstream components of $\mathbf{p}$. □

4.5. Coordinate change

Next we discuss coordinate changes of the Kuranishi charts. We refer [23, Definition 3.6] for the definition of coordinate changes.

Lemma 4.33. *For each $\mathbf{q}_1 \in \mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ there exists $\epsilon_1 > 0$ such that the following holds. Suppose $\mathbf{q}_2 \in \mathrm{Im}(\psi_{(\mathbf{q}_1,\epsilon_1,\mathcal{B})})$, then*

1. *$\mathcal{E}(\mathbf{q}_2) \subseteq \mathcal{E}(\mathbf{q}_1)$.*
2. *Let $\mathcal{B}_2 \subseteq \mathcal{E}(\mathbf{q}_2)$, $\mathcal{B}_1 \subseteq \mathcal{E}(\mathbf{q}_1)$ and $\mathcal{B}_2 \subseteq \mathcal{B}_1$. Then there exists $\epsilon_2 > 0$ such that there exists a coordinate change from*
$$(U(\mathbf{q}_2,\epsilon_2,\mathcal{B}_2), E(\mathbf{q}_2,\epsilon_2,\mathcal{B}_2), s_{(\mathbf{q}_2,\epsilon_2,\mathcal{B}_2)}, \psi_{(\mathbf{q},\epsilon_2,\mathcal{B}_2)})$$
to
$$(U(\mathbf{q}_1,\epsilon_1,\mathcal{B}_1), E(\mathbf{q}_1,\epsilon_1,\mathcal{B}_1), s_{(\mathbf{q}_1,\epsilon_1,\mathcal{B}_1)}, \psi_{(\mathbf{q},\epsilon_1,\mathcal{B}_1)}).$$

Proof. The proof is the same as [20, Sections 9 and 10] [16, Section 8.3]. □

Lemma 4.34. *We may choose $\epsilon_1 > 0$ and $\epsilon_2 > 0$ in Lemma 4.33 such that the following holds.*

1. *When we replace $\mathbf{q}_1$, ϵ_1 by $\mathbf{q}_2$, ϵ_2 in Lemma 4.33, the same conclusion as Lemma 4.33 holds.*
2. *If $\mathbf{q}_3 \in \mathrm{Im}(\psi_{(\mathbf{q}_2,\epsilon_2,\mathcal{B}_2)})$, then there exists $\epsilon_3 > 0$ such that we have the conclusion of Lemma 4.33 with $\mathbf{q}_1$, ϵ_1, $\mathbf{q}_2$ and ϵ_2 replaced by $\mathbf{q}_2$, ϵ_2, $\mathbf{q}_3$ and ϵ_3, respectively.*
   We also have the conclusion of Lemma 4.33 with $\mathbf{q}_1$, ϵ_1, $\mathbf{q}_2$ and ϵ_2 replaced by $\mathbf{q}_1$, ϵ_1, $\mathbf{q}_3$ and ϵ_3, respectively.
3. *Let (i,j) be one of $(i,j) = (1,2),(2,3),(1,3)$ and $\Phi_{\mathbf{q}_i,\mathbf{q}_j}$ be the coordinate change from*
$$(U(\mathbf{q}_j,\epsilon_j,\mathcal{B}_j), E(\mathbf{q}_j,\epsilon_j,\mathcal{B}_j), s_{(\mathbf{q}_j,\epsilon_j,\mathcal{B}_j)}, \psi_{(\mathbf{q},_j\epsilon_j,\mathcal{B}_j)})$$
to
$$(U(\mathbf{q}_i,\epsilon_i,\mathcal{B}_i), E(\mathbf{q}_i,\epsilon_i,\mathcal{B}_i), s_{(\mathbf{q}_i,\epsilon_i,\mathcal{B}_i)}, \psi_{(\mathbf{q},_i\epsilon_i,\mathcal{B}_i)})$$
obtained by Lemma 4.33. Then we have
$$\Phi_{\mathbf{q}_1,\mathbf{q}_3} = \Phi_{\mathbf{q}_1,\mathbf{q}_2} \circ \Phi_{\mathbf{q}_2,\mathbf{q}_3}.$$

Proof. The proof is the same as in [20, Section 7]. □

We can use Lemmas 4.30, 4.33 and 4.34 to construct a required Kuranishi structure on $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ in exactly the same way as in [20].

The isomorphism on the orientation bundle stated in Theorem 4.1 (1) can be proved as follows. Since the marked points are parametrized by complex coordinates, hence even dimensional and canonically oriented, it is enough to consider the case that $\ell = 0$. The fiber of the orientation bundle of $\mathcal{M}(X,H;\alpha_-,\alpha_+)$ at $\mathbf{p} \in \mathcal{M}(X,H;\alpha_-,\alpha_+)$ is defined by

$$o_{\mathcal{M}(X,H;\alpha_-,\alpha_+)}|_{\mathbf{p}} \otimes \mathbb{R}_{\alpha_-,\alpha_+} = \det T_{\gamma_+}R_{\alpha_+} \otimes \det D_{\mathbf{p}}\overline{\partial}_{J,H} \otimes \det T_{\gamma_-}R_{\alpha_-}, \tag{4.22}$$

where $D_{\mathbf{p}}\overline{\partial}_{J,H}$ is the linearized operator at $\mathbf{p}$ in (4.7) with fixed asymptotics $\gamma_\pm \in \overline{R}_{\alpha_\pm}$ and $\mathbb{R}_{\alpha_-,\alpha_+}$ stands for the translation action on $\widetilde{\mathcal{M}}(X,H;\alpha_-,\alpha_+)$.

Since the linearization operator (4.6) for bubble components is deformable to a complex linear Fredholm operator and the matching condition (♡) is taken in complex vector spaces, we concentrate on the linearization operator (4.5) for mainstream components with fixed asymptotics $\gamma_\pm$ here.

For $D_{\mathbf{p}}\overline{\partial}_{J,H}$, we have

$$\det D_{\mathbf{p}}\overline{\partial}_{J,H} \otimes \det TR_{\alpha_-} \otimes \det P(\gamma_-; \mathfrak{t}_{w_-}) \cong \det P(\gamma_+; t_{w_+}). \tag{4.23}$$

Fixing[16] an orientation on $\mathbb{R}$ once and for all, we can drop the factor $\mathbb{R}_{\alpha_-,\alpha_+}$ from the left hand side of (4.22). (This factor is important for (16.8) in [23, Condition X] as we will see below.) Combining (4.22) and (4.23), we obtain

$$\det T_{\gamma_+} R_{\alpha_+} \otimes \det P(\gamma_+; t_{w_+}) \cong o_{\mathcal{M}(X,H;\alpha_-,\alpha_+)}|_{\mathbf{p}} \otimes \det P(\gamma_-; \mathfrak{t}_{w_-}). \tag{4.24}$$

Recalling the definition of o_{R_α} from Definition 2.6, it is the desired orientation isomorphism.

Next we prove (16.8) in [23, Condition X]. Suppose that $\mathbf{p} \in \mathcal{M}(X, H; \alpha_-, \alpha_+)$ decomposes into $(\mathbf{p}_1, \mathbf{p}_2) \in \mathcal{M}(X, H; \alpha_-, \alpha_+) \times_{R_\alpha} \mathcal{M}(X, H; \alpha_-, \alpha)$. Then we find that

$$\det D_{\mathbf{p}}\overline{\partial}_{J,H} \cong \det D_{\mathbf{p}_2}\overline{\partial}_{J,H} \otimes \det T_\gamma R_\alpha \otimes \det D_{\mathbf{p}_1}\overline{\partial}_{J,H}, \tag{4.25}$$

where γ is the positive asymptotic limit of $\mathbf{p}_1$ and the negative asymptotic limit of $\mathbf{p}_2$. Then we have the following

$$\begin{aligned}
&o_{\mathcal{M}(X,H;\alpha_-,\alpha_+)}|_{\mathbf{p}} \otimes \mathbb{R}_{\alpha_-,\alpha_+}\\
&\cong \det T_{\gamma_+} R_{\alpha_+} \otimes \det D_{\mathbf{p}}\overline{\partial}_{J,H} \otimes \det T_{\gamma_-} R_{\alpha_-}\\
&\cong \det T_{\gamma_+} R_{\alpha_+} \otimes \det D_{\mathbf{p}_2}\overline{\partial}_{J,H} \otimes T_\gamma R_\alpha \otimes \det D_{\mathbf{p}_1}\overline{\partial}_{J,H} \otimes \det T_{\gamma_-} R_{\alpha_-}\\
&\cong o_{\mathcal{M}(X,H;\alpha_-,\alpha_+)}|_{\mathbf{p}_2} \otimes \mathbb{R}_{\alpha,\alpha_+} \otimes \det D_{\mathbf{p}_1}\overline{\partial}_{J,H} \otimes \det T_{\gamma_-} R_{\alpha_-}\\
&\cong o_{\mathcal{M}(X,H;\alpha,\alpha_+)}|_{\mathbf{p}_2} \otimes \mathbb{R}_{\alpha,\alpha_+} \otimes (\det T_\gamma R_\alpha)^* \otimes o_{\mathcal{M}(X,H;\alpha_-,\alpha)}|_{\mathbf{p}_1} \otimes \mathbb{R}_{\alpha_-,\alpha}\\
&\cong (-1)^{\dim \mathcal{M}(X,H;\alpha,\alpha_+)} \mathbb{R}_{\alpha,\alpha_+} \otimes o_{\mathcal{M}(X,H;\alpha,\alpha_+) \times_{R_\alpha} \mathcal{M}(X,H;\alpha_-,\alpha)} \otimes \mathbb{R}_{\alpha_-,\alpha}\\
&\cong (-1)^{\dim \mathcal{M}(X,H;\alpha,\alpha_+)} \mathbb{R}_{\mathrm{out}} \otimes o_{\mathcal{M}(X,H;\alpha,\alpha_+) \times_{R_\alpha} \mathcal{M}(X,H;\alpha_-,\alpha)} \otimes \mathbb{R}_{\alpha_-,\alpha_+}.
\end{aligned} \tag{4.26}$$

Here $\mathbb{R}_{\mathrm{out}}$ is the outward normal direction of $\partial\mathcal{M}(X, H; \alpha_-, \alpha_+)$. The first, third and fourth isomorphisms follows from (4.22). The second follows from (4.25). The fifth is due to the definition of the fiber product orientation on K-spaces, see [11, Convention 8.2.1 (4)]. For the sixth isomorphism, see the proof of [11, Proposition 8.3.3]. Thus we have proved Theorem 4.1 (1).

To prove Theorem 4.1 (2) we can use Lemmas 4.31 and 4.32 to show that the Kuranishi structure constructed above induces a Kuranishi structure on the codimension $k-1$ normalized corner $\widehat{S}_{k-1}(\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+))$ which is the disjoint union of the spaces $\mathcal{M}_{(\ell_1,\ldots,\ell_k)}(X, H; \alpha_0, \alpha_1., \ldots, \alpha_k)$ for various $\alpha_- = \alpha_0, \alpha_1, \ldots, \alpha_{k-1}, \alpha_k = \alpha_+ \in \mathfrak{A}$. Thus we have Theorem 4.1 (2).

Theorem 4.1 (3) follows from Definition 4.11 (6) and Remark 4.12 (1). Theorem 4.1 (4) is a consequence of Lemma 2.15.

Therefore, the proof of Theorem 4.1 is complete. □

[16] See [11, Convention 8.3.1].

Remark 4.35. In [20] we introduced ambient 'set' and used it to prove cocycle condition for the coordinate changes between Kuranishi charts. We can adopt the method in our situation as follows.

We first define the set $\mathcal{X}_\ell(X, H; \alpha_-, \alpha_+)$ as the set of equivalence classes of $(\Sigma, (z_-, z_+, \vec{z}), u, \varphi)$ which satisfies the conditions in Definition 3.3 except (4)(5)(6)(9). In other words, we do not require the condition that the equations (pseudo-holomorphic curve equation or Floer's equation) are satisfied or the stability. The definition of two such objects being equivalent is the same as $\sim_2$ in Definition 3.7. By definition $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ is a subset of $\mathcal{X}_\ell(X, H; \alpha_-, \alpha_+)$. For any $\mathbf{p} \in \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ we define its ϵ-neighborhood in $\mathcal{X}_\ell(X, H; \alpha_-, \alpha_+)$ as the subset consisting of element which is ϵ-close to $\mathbf{p}$ in the sense of Definition 4.16. It then defines a partial topology of the pair $(\mathcal{X}_\ell(X, H; \alpha_-, \alpha_+), \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+))$ in the sense of [20, Definition 4.1]. (This is a consequence of Lemma 4.34.) Thus in the same way as in [20, Section 7] we obtain a Kuranishi structure.

5. Compatibility of Kuranishi structures

In this section we complete the proof of Theorem 2.9. Namely, we modify the Kuranishi structures in Theorem 4.1 for the Morse–Bott case so that Kuranishi structures are compatible on boundary and corners under the fiber product. The argument presented here has not been given in detail for the moduli space of solutions to Floer's Eq. (2.4) in the previous literature. For the case of the moduli spaces of pseudo-holomorphic disks it is written in detail in [21]. The method of [21] is different from that of this section. We take a different route to illustrate two different methods. Both methods in [21] and in this section can be applied to both situations.

5.1. Outer collar

The statements (1) (2) of Theorem 2.9 are a part of Theorem 4.1 and already proved in Sect. 4. To prove (3) we will introduce an enhanced space $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)^{\boxplus 1}$ of $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ by putting a collar 'outside' of $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ and modify the Kuranishi structure on the outer collar $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)^{\boxplus 1} \backslash \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ in the sense of [23, Chapter 17]. As a topological space we define the space $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)^{\boxplus 1}$ as follows.[17]

Definition 5.1. We consider $(\Sigma, (z_-, z_+, \vec{z}), u, \varphi; \vec{t})$ where

- $((\Sigma, (z_-, z_+, \vec{z}), u, \varphi) \in \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$,
- $\vec{t}$ assigns a number $t_z \in [-1, 0]$ to each transit point z of Σ.

We denote by $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)^{\boxplus 1}$ the set of isomorphism classes of such objects $(\Sigma, (z_-, z_+, \vec{z}), u, \varphi; \vec{t})$. We call it the *outer collaring* of $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$.

We say a sequence $(\Sigma^j, (z^j_-, z^j_+, \vec{z}^{\,j}), u^j, \varphi^j; \vec{t}^{\,j}) \in \mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)^{\boxplus 1}$ converges to $(\Sigma, (z_-, z_+, \vec{z}), u, \varphi; \vec{t})$ if the following holds.

[17] See [23, Lemma-Definition 17.29] for the definition of the outer collar for a general K-space.

(1) $(\Sigma^j,(z^j_-,z^j_+,\vec z^{\,j}),u^j,\varphi^j)$ converges to $(\Sigma,(z_-,z_+,\vec z),u,\varphi)$ in the sense of Definition 3.17.
(2) Let z be a transit point of Σ.
   (a) Suppose $j_n\to\infty$ and there exists a sequence of transit points $z_n\in\Sigma^{j_n}$ which converges to z in an obvious sense, then $t^{j_n}_{z_{j_n}}$ converges to t_z.
   (b) If there is no such a sequence then $t_z=0$.

It is easy to see that we can define a topology in this way.

We can show that $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)^{\boxplus 1}$ is compact and Hausdorff. The evaluation map also extends to the outer collar so that it does not depend on $\vec t$. Moreover, by inspecting the discussion in [23, Chapter 17] we can show that $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)^{\boxplus 1}$ is the underlying topological space of the outer collared space of $\mathcal{M}_\ell(X,H;\alpha_-,\alpha_+)$ with respect to the Kuranishi structure defined in the last section. (However, we do not use this fact in this paper.)

Now the main part of this construction is Proposition 5.5. To state it we prepare some notations. Let

$$\alpha_-=\alpha_0,\alpha_1,\dots,\alpha_{m-1},\alpha_m=\alpha_+\in\mathfrak{A}.$$

In Sect. 3 we denoted by $\mathcal{M}_{(\ell_1,\dots,\ell_m)}(X,H;\alpha_0,\alpha_1.,\dots,\alpha_m)$ the fiber product (3.9)

$$\begin{aligned}&\mathcal{M}_{\ell_1}(X,H;\alpha_0,\alpha_1)\,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\mathcal{M}_{\ell_2}(X,H;\alpha_1,\alpha_2)\,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\cdots\\ &{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\mathcal{M}_{\ell_{i+1}}(X,H;\alpha_i,\alpha_{i+1})\,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\cdots\times_{\mathrm{ev}_-}\mathcal{M}_{\ell_m}(X,H;\alpha_{m-1},\alpha_m).\end{aligned}\tag{5.1}$$

Hereafter we write (5.1) as

$$\mathcal{M}_{\vec\ell}(X,H;\vec\alpha).$$

We will construct a certain Kuranishi structure on $\mathcal{M}_{\vec\ell}(X,H;\vec\alpha)\times[-1,0]^{m-1}$.

Definition 5.2. We put $\underline{m-1}=\{1,\dots,m-1\}$. Let

$$A\sqcup B\sqcup C=\underline{m-1}$$

be a decomposition into a disjoint union. We put $B=\{j_1,\dots,j_b\}$. We define an embedding

$$\mathcal{I}_{A,B,C}:[-1,0]^b\to[-1,0]^{m-1}$$

by $\mathcal{I}_{A,B,C}(t_1,\dots,t_b)=(s_1,\dots,s_{m-1})$ where

$$s_i=\begin{cases}-1 & \text{if } i\in A,\\ t_k & \text{if } i=j_k,(k=1,\dots,b),\\ 0 & \text{if } i\in C.\end{cases}\tag{5.2}$$

We put

$$\mathrm{Part}_3(m-1)=\{(A,B,C)\mid A\sqcup B\sqcup C=\underline{m-1}\}.$$

We note that the set of images of $\mathcal{I}_{A,B,C}$ for various $(A,B,C) \in \mathrm{Part}_3(m-1)$ coincides with the set of cells of the standard cell decomposition of $[-1,0]^{m-1}$.

The compatibility condition we formulate below (Condition 5.4) describes the restriction of the Kuranishi structure $\widehat{\mathcal{U}}_{\vec{\ell}}(X,H;\vec{\alpha})$ on the product space $\mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha}) \times [-1,0]^{m-1}$ to the image of the embedding:

$$\mathrm{id} \times \mathcal{I}_{A,B,C} : \mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha}) \times [-1,0]^{b} \to \mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha}) \times [-1,0]^{m-1}. \tag{5.3}$$

We need more notations. We write elements of the set A as $A = \{i(A,1), \dots, i(A,a)\}$ with $i(A,1) < i(A,2) < \cdots < i(A,a-1) < i(A,a)$ and consider the fiber product

$$\begin{aligned}
&\mathcal{M}_{\vec{\ell}_{A,1}}(X,H;\alpha_0,\dots,\alpha_{i(A,1)}) \\
&\quad {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{A,2}}(X,H;\alpha_{i(A,1)},\dots,\alpha_{i(A,2)}) \, {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \cdots \\
&\quad {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{A,j+1}}(X,H;\alpha_{i(A,j)},\dots,\alpha_{i(A,j+1)}) \, {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \cdots \\
&\quad {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{A,a+1}}(X,H;\alpha_{i(A,a)},\dots,\alpha_{m}).
\end{aligned} \tag{5.4}$$

Here

$$\vec{\ell}_{A,j} = (\ell_{i(A,j-1)+1},\dots,\ell_{i(A,j)}), \tag{5.5}$$

and $i(A,0) = 0$ and $i(A,a+1) = m$ by convention. Actually the fiber product (5.4) is nothing but $\mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha})$. Therefore, we can use (5.4) to define a fiber product Kuranishi structure on $\mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha})$.

Notation 5.3. Let $(A,B,C) \in \mathrm{Part}_3(m-1)$ and $j = 1,\dots,a+1$.

(1) We put

$$\mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha})^{+} = \mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha}) \times [-1,0]^{m-1},$$

where $m+1$ is the number of components of $\vec{\alpha}$. Note that $\mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha})^{\boxplus 1}$ in Definition 5.1 is a union of $\mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha})^{+}$ for various $\vec{\alpha}$.

(2) For $a,b \in \mathbb{Z}$ we put $[a,b]_{\mathbb{Z}} = [a,b] \cap \mathbb{Z}$ and $(a,b)_{\mathbb{Z}} = [a,b]_{\mathbb{Z}} \backslash \{a,b\}$.

(3) We decompose C into $C'_j(A) = [i(A,j-1), i(A,j)]_{\mathbb{Z}} \cap C$ and put $C_j(A) = \{i - i(A,j-1) \mid i \in C'_j(A)\}$, $c_j(A) = \# C_j(A)$. (Recall $i(A,0) = 0$ and $i(A,a+1) = m$ as above.)

(4) We also put

$$\vec{\alpha}_{A,j} = (\alpha_{i(A,j-1)},\dots,\alpha_{i(A,j)}).$$

We put $\alpha_0 = \alpha_-$ and $\alpha_m = \alpha_+$ by convention. We include them as elements of $\vec{\alpha}_{A,0}$ and $\vec{\alpha}_{A,a+1}$, respectively. Note that

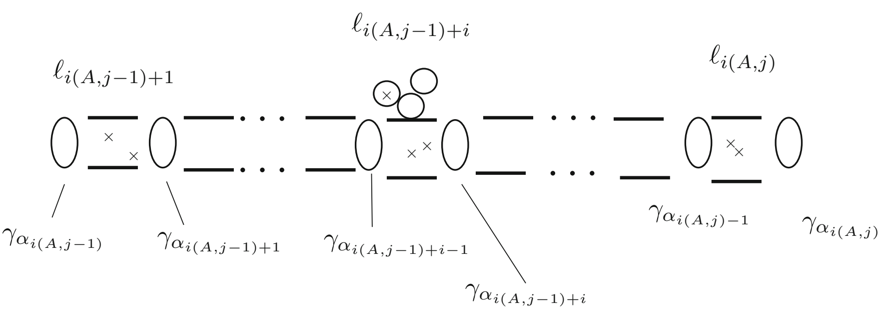
$$\mathcal{M}_{\vec{\ell}_{A,j}}(X,H;\alpha_{i(A,j-1)},\dots,\alpha_{i(A,j)}) = \mathcal{M}_{\vec{\ell}_{A,j}}(X,H;\vec{\alpha}_{A,j}). \tag{5.6}$$

See Fig. 5.

(5) We remove $\{\alpha_i \mid i \in C'_j(A)\}$ from $\vec{\alpha}_{A,j}$ to obtain $\vec{\alpha}_{A,j,C}$.

(6) We put $m_j(A) = i(A,j) - i(A,j-1)$ and

$$m_j(A,C) = \#(B \cap (i(A,j-1), i(A,j))_{\mathbb{Z}}) + 1.$$



$$\vec{\alpha}_{A,j} = (\alpha_{i(A,j-1)}, \dots, \alpha_{i(A,j)}), \qquad \vec{\ell}_{A,j} = (\ell_{i(A,j-1)}, \dots, \ell_{i(A,j)})$$

FIGURE 5. $\mathcal{M}_{\vec{\ell}_{A,j}}(X, H; \alpha_{i(A,j-1)}, \dots, \alpha_{i(A,j)})$

Then we have

$$\sum_{j=0}^{a+1}(m_j(A,C) - 1) = \#B = b. \tag{5.7}$$

(7) We define $\vec{\ell}_{A,j,C}$ as follows. Let $\vec{\alpha}_{A,j,C} = \{\alpha_{i(A,j-1)+k_s} \mid s = 0, \dots, m_j(A,C)\}$. Here $k_0 = 0 < k_1 < \cdots < k_{m_j(A,C)} = i(A,j) - i(A,j-1)$. Note if $i \in (i(A,j-1) + k_s, i(A,j-1) + k_{s+1})_{\mathbb{Z}}$, then $i \in C'_j(A)$. We put

$$\ell_{A,j,C,s} = \ell_{i(A,j-1)+k_{s-1}+1} + \cdots + \ell_{i(A,j-1)+k_s}$$

for $s = 1, \dots, m_j(A,C)$ and $\vec{\ell}_{A,j,C} = (\ell_{A,j,C,1}, \dots, \ell_{A,j,C,m_j(A,C)})$. See Fig. 6.

Now the compatibility condition we require is described as follows.

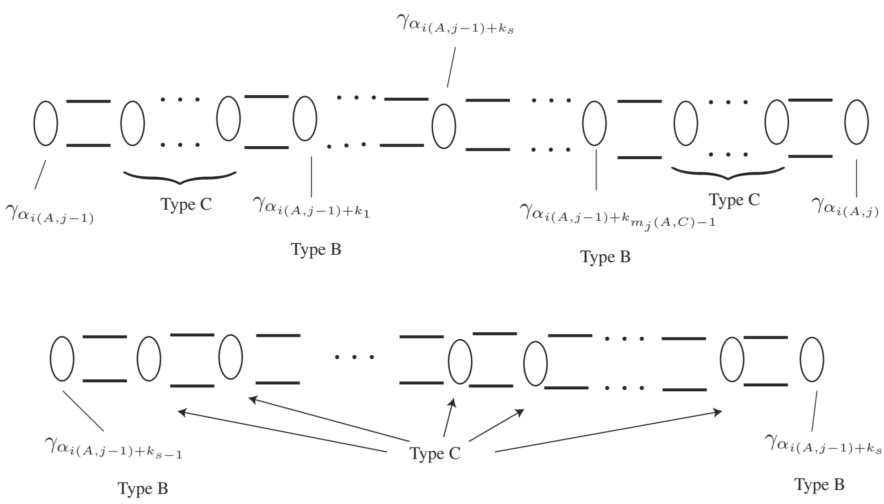
Condition 5.4. The restriction of the Kuranishi structure $\widehat{\mathcal{U}}_{\vec{\ell}}(X, H; \vec{\alpha})$ on the space $\mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha})^+$ to the image of the embedding (5.3) coincides with the fiber product of the restrictions of the Kuranishi structures $\widehat{\mathcal{U}}_{\vec{\ell}_{A,j,C}}(X, H; \vec{\alpha}_{A,j,C})$ to $\mathcal{M}_{\vec{\ell}_{A,j}}(X, H; \vec{\alpha}_{A,j}) \times [-1,0]^{m_j(A,C)-1}$.

We note that $\widehat{\mathcal{U}}_{\vec{\ell}_{A,j,C}}(X, H; \vec{\alpha}_{A,j,C})$ is a Kuranishi structure on the direct product space $\mathcal{M}_{\vec{\ell}_{A,j,C}}(X, H; \vec{\alpha}_{A,j,C})^+ = \mathcal{M}_{\vec{\ell}_{A,j,C}}(X, H; \vec{\alpha}_{A,j,C}) \times [-1,0]^{m_j(A,C)-1}$ ($j = 1, \dots, a+1$) and $\mathcal{M}_{\vec{\ell}_{A,j}}(X, H; \vec{\alpha}_{A,j})$ is a component of the normalized corner[18]

$$\widehat{S}_{c_j(A)}(\mathcal{M}_{\vec{\ell}_{A,j,C}}(X, H; \vec{\alpha}_{A,j,C})).$$

Therefore, we can restrict the Kuranishi structure $\widehat{\mathcal{U}}_{\vec{\ell}_{A,j,C}}(X, H; \vec{\alpha}_{A,j,C})$ to a Kuranishi structure on $\mathcal{M}_{\vec{\ell}_{A,j}}(X, H; \vec{\alpha}_{A,j}) \times [-1,0]^{m_j(A,C)-1}$.

[18] See [23, Definition 24.18] for the normalized corner.



$$\vec{\alpha}_{A,j,C} = (\alpha_{i(A,j-1)}, \alpha_{i(A,j-1)+k_1}, \dots, \alpha_{i(A,j-1)+k_{m_j(A,C)-1}}, \alpha_{i(A,j)})$$

FIGURE 6. $\vec{\alpha}_{A,j,C}$ and $\vec{\ell}_{A,j,C}$

We take the fiber product of them for various j using (5.4) and obtain a Kuranishi structure on $\mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha}) \times [-1, 0]^b$. (Note we use (5.7) here.)

Condition 5.4 requires that this Kuranishi structure coincides with the restriction of $\widehat{\mathcal{U}}_{\vec{\ell}}(X, H; \vec{\alpha})$ to the image of (5.3). Let us elaborate on Condition 5.4. Formula (5.2) shows that at the image of (5.3) we have $s_i = -1$ for $i \in A$ and $s_i = 0$ for $i \in C$.

We are taking a fiber product over $R_{\alpha_{i(A,j)}}$. Therefore, we take the fiber product corresponding to the singular points for which $s_i = -1$. This is related to the compatibility of the Kuranishi structures at the boundary and corners. (Theorem 2.9 (3) and [23, Condition 16.1 (X)].)

Let us consider the part $s_i = 0$. For simplicity of notation, we explain the case $m = 2$ and $C = \{1\}$. We consider $\mathcal{M}(X, H; \alpha_-, \alpha) \times_{R_\alpha} \mathcal{M}(X, H; \alpha, \alpha_+) \times \{0\}$. Condition 5.4 in this case means that the Kuranishi structure there is the restriction of the Kuranishi structure on $\mathcal{M}(X, H; \alpha_-, \alpha_+)$. This condition is used to glue $\mathcal{M}(X, H; \alpha_-, \alpha) \times_{R_\alpha} \mathcal{M}(X, H; \alpha, \alpha_+) \times [-1, 0]$ with $\mathcal{M}(X, H; \alpha_-, \alpha_+)$ there.

Proposition 5.5. *There exists a K-system*

$$\{(\mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha})^+, \ \widehat{\mathcal{U}}_{\vec{\ell}}(X, H; \vec{\alpha}))\}$$

for various $\vec{\ell}, \vec{\alpha}$ with the following properties.

(1) *They satisfy Condition 5.4.*

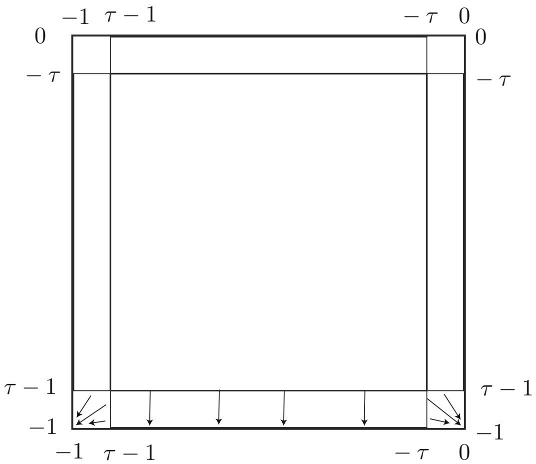


FIGURE 7. $\mathfrak{C}$-collared-ness and the retraction $I(S,M,L) \to [-1+\tau,-\tau]^{\#M}$

(2) *Let* $\mathfrak{C}$ *be*[19] *the union of the components of* $\partial\mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha})^+$ *which are in* $\mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha}) \times \partial([-1,0]^{m-1})$*. Then* $\widehat{\mathcal{U}}_{\vec{\ell}}(X,H;\vec{\alpha})$ *is* $\mathfrak{C}$*-collared in the sense of Remark 5.6 below.*

(3) *For the case* $\vec{\alpha} = (\alpha_-,\alpha_+)$*,* $\widehat{\mathcal{U}}_{\ell}(X,H;(\alpha_-,\alpha_+))$ *coincides with the Kuranishi structure we produced in Theorem 4.1.*

We remark that $\mathcal{M}_{\ell}(X,H;(\alpha_-,\alpha_+))^+ = \mathcal{M}_{\ell}(X,H;(\alpha_-,\alpha_+))$. So the statement (3) above makes sense.

Remark 5.6. (1) The definition of $\mathfrak{C}$-collared-ness of a K-space can be seen in [23, Definition 18.9] in the abstract setting. However, in our current situation we can describe the $\mathfrak{C}$-collared-ness of the Kuranishi structure more explicitly as follows:

For a decomposition of the set $\underline{m-1}$ into three disjoint union $(S,M,L) \in \mathrm{Part}_3(m-1)$, we define $I(S,M,L)$ to be the set of $(t_1,\dots,t_{m-1}) \in [-1,0]^{m-1}$ such that $t_i \in [-1,-1+\tau]$ if $i \in S$, $t_i \in [-\tau,0]$ if $i \in L$, and $t_i \in [-1+\tau,-\tau]$ if $i \in M$. We consider the retraction $I(S,M,L) \to [-1+\tau,-\tau]^{\#M}$ by forgetting t_i for $i \notin M$. We embed $[-1+\tau,-\tau]^{\#M}$ into $I(S,M,L)$ by putting $t_i = -1$ if $i \in S$, $t_i = 0$ if $i \in L$. Let $\pi : \partial\mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha})^+ \to [-1,0]^{m-1}$ be the obvious projection. Now we require that the restriction of our Kuranishi structure to $\pi^{-1}(I(S,M,L))$ is obtained by a pull back of one on the image of $\pi^{-1}([-1+\tau,-\tau]^{\#M})$. See Fig. 7.

(2) Indeed we use the $\mathfrak{C}$-collared-ness in the proof of Theorem 2.9 below.

[19] In [23, Situation 18.1] the symbol $\mathfrak{C}$ is used to indicate a certain specific set of components among boundary components and to define the notion of '$\mathfrak{C}$-partial outer collars' ([23, Definition 18.9]). In the current situation this $\mathfrak{C}$ stands for the boundary components arising from the $[-1,0]^{m-1}$ factors.

Proof of Theorem 2.9 assuming Proposition 5.5. We note that there is a retraction:

$$\mathcal{R} : \mathcal{M}_\ell(X, H; (\alpha_-, \alpha_+))^{\boxplus 1} \to \mathcal{M}_\ell(X, H; (\alpha_-, \alpha_+)). \tag{5.8}$$

This is a map which sends $(\Sigma, (z_-, z_+, \vec{z}), u, \varphi, \vec{t})$ to $(\Sigma, (z_-, z_+, \vec{z}), u, \varphi)$. It is easy to see that the inverse image $\mathcal{R}^{-1}(z)$ is $[-1, 0]^m$ if and only if $z \in \overset{\circ}{S}_m(\mathcal{M}_\ell(X, H; (\alpha_-, \alpha_+)))$ Here $\overset{\circ}{S}_m(\mathcal{M}_\ell(X, H; (\alpha_-, \alpha_+)))$ is the set of points in codimension m corners of $\mathcal{M}_\ell(X, H; (\alpha_-, \alpha_+))$ which do not lie in higher codimension corners. In fact, $\overset{\circ}{S}_m(\mathcal{M}_\ell(X, H; (\alpha_-, \alpha_+)))$ consists of $(\Sigma, (z_-, z_+, \vec{z}), u, \varphi)$ such that Σ has exactly m transit points.

By definition the original Kuranishi structure on $\mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha})^+$ is the direct product of the Kuranishi structure on $\mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha})$ (which is the restriction of the Kuranishi structure on $\mathcal{M}_{\vec{\ell}}(X, H; \alpha_-, \alpha_+)$ given in Theorem 4.1) and the trivial Kuranishi structure on $[-1, 0]^{m-1}$. (See [23, Lemma-Definition 17.38].)

We replace the direct product Kuranishi structure on $\mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha}) \times [-1, 0]^{m-1}$ by $\widehat{\mathcal{U}}_{\vec{\ell}}(X, H; \vec{\alpha})$ given in Proposition 5.5.

We first check that those Kuranishi structures $\widehat{\mathcal{U}}_{\vec{\ell}}(X, H; \vec{\alpha})$ can be glued to give a Kuranishi structure on $\mathcal{M}_\ell(X, H; (\alpha_-, \alpha_+))^{\boxplus 1}$.

Because of Proposition 5.5 (2), it suffices to check that they are compatible for various $\mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha}) \times [-1, 0]^{m-1}$ at their intersection points. (Here the $\mathfrak{C}$-collared-ness is used as we mentioned in Remark 5.6 (2).)

The compatibility of two different members of the set $\{\widehat{\mathcal{U}}_{\vec{\ell}}(X, H; \vec{\alpha})\}$ at the overlapped part is a consequence of Condition 5.4. Indeed it follows from the case when $A = \emptyset$ of Condition 5.4. (Note that they are glued at the part where a certain coordinate of the $[-1, 0]^{m-1}$ factor is 0.) The compatibility of $\widehat{\mathcal{U}}_{\vec{\ell}}(X, H; \vec{\alpha})$ with the Kuranishi structure on $\mathcal{M}_\ell(X, H; \alpha_-, \alpha_+)$ given by Theorem 4.1 follows from Condition 5.4 and Proposition 5.5 (3).

We thus obtain a Kuranishi structure on $\mathcal{M}_\ell(X, H; (\alpha_-, \alpha_+))^{\boxplus 1}$. It is immediate from construction that it satisfies Theorem 2.9 (4).

It remains to prove Theorem 2.9 (3). The main point we need to prove is the compatibility of our Kuranishi structures at the boundary and corners [23, Conditions 16.1 (X) (IX)]. This is the condition on the Kuranishi structure at the point where a certain coordinate of the $[-1, 0]^{m-1}$ factor is -1. This is a consequence of Condition 5.4. More precisely, the case $C = \emptyset$ of Condition 5.4 implies [23, Conditions 16.1 (X) (IX)].

The other defining conditions for our Kuranishi structures to form a linear K-system are easy to check. (The periodicity of linear K-system [23, Conditions 16.1 (VIII)] follows from Remark 4.3.) The proof of Theorem 2.9 is now complete. □

5.2. Proof of Proposition 5.5 I: Obstruction space with outer collar

This subsection and the next are occupied with the proof of Proposition 5.5. In this subsection we define an obstruction space of Kuranishi chart at each point in $\mathcal{M}_{\vec{\ell}}(X, H; \alpha)^+$ (Definition 5.17). Then we will construct a desired

Kuranishi structure and complete the proof of Proposition 5.5 in the next subsection.

Proof of Proposition 5.5. Let $(A,B,C) \in \mathrm{Part}_3(m-1)$, $\#B = b$ and $(D,E,F) \in \mathrm{Part}_3(b)$. We define $(A, B(D,E,F), C) \in \mathrm{Part}_3(m-1)$ as follows. We put $B = \{j_1,\dots,j_b\}$. Then $(A,B(D,E,F),C) = (A',B',C')$ such that

(1) $A' \supseteq A$, $C' \supseteq C$.
(2) $j_i \in A'$ if $i \in D$.
(3) $j_i \in B'$ if $i \in E$.
(4) $j_i \in C'$ if $i \in F$.

Note (1)–(4) above is equivalent to $\mathcal{I}_{A,B,C} \circ \mathcal{I}_{D,E,F} = \mathcal{I}_{A',B',C'}$.

Definition 5.7. We consider a system of closed subsets $\mathcal{V}(A,B,C)$ of $[-1,0]^{m-1}$ over $m = 1,2,\dots$ and $(A,B,C) \in \mathrm{Part}_3(m-1)$ with the following properties.

(1)
$$\bigcup_{(A,B,C)\in \mathrm{Part}_3(m-1)} \mathrm{Int}\,\mathcal{V}(A,B,C) = [-1,0]^{m-1}.$$

(2) If $(A,B(D,E,F),C) = (A',B',C')$, then
$$(\mathcal{I}_{A,B,C})^{-1}(\mathcal{V}(A',B',C')) = \mathcal{V}(D,E,F).$$

(3) If $\sigma \in \mathrm{Perm}(m-1)$, then
$$\sigma(\mathcal{V}(A,B,C)) = \mathcal{V}(\sigma A, \sigma B, \sigma C).$$
Here σ acts on $[-1,0]^{m-1}$ by permutation of the factors and to Part_3 $(m-1)$ in an obvious way.

(4) If $(A,B,C),(A',B',C') \in \mathrm{Part}_3(m-1)$ and $\mathcal{V}(A,B,C) \cap \mathcal{V}(A',B',C')$ is nonempty, then either $\mathrm{Im}(\mathcal{I}_{A,B,C})$ is a face of $\mathrm{Im}(\mathcal{I}_{A',B',C'})$ or $\mathrm{Im}(\mathcal{I}_{A',B',C'})$ is a face of $\mathrm{Im}(\mathcal{I}_{A,B,C})$.

(5) Let $\vec{t} = (t_1,\dots,t_{m-1}) \in \mathcal{V}(A,B,C)$. If $t_i = -1$ then $i \in A$. If $t_i = 0$ then $i \in C$.

(6) $\mathcal{V}(A,B,C)$ is a direct product in each sufficiently small neighborhood of the strata of $[-1,0]^{m-1}$ in the following sense. Let $\vec{t} \in S_\ell([-1,0]^{m-1})$ and take its neighborhood which is isometric to $\sigma(U \times (-\epsilon,0]^p \times [-1,-1+\epsilon)^q)$ for some $\sigma \in \mathrm{Perm}(m-1)$, $\epsilon > 0$ with $p+q = \ell$. (Here U is isometric to an open set of $\mathbb{R}^{m-1-\ell}$.) We then require
$$\mathcal{V}(A,B,C) \cap (\sigma(U \times (-\epsilon,0]^p \times [-1,-1+\epsilon)^q))$$
is isometric to the direct product $(\mathcal{V}(A,B,C) \cap \sigma(U \times \{0\} \times \{0\})) \times [0,\epsilon)^\ell$.

Lemma 5.8. *There exists a system of closed subsets $\mathcal{V}(A,B,C)$ satisfying (1)–(6) in Definition 5.7.*

Proof. The proof is by induction on m. There is nothing to prove in the case $m = 1$. In the case $m = 2$ we consider the interval $[-1,0] = [-1,0]^{2-1}$. $\mathrm{Part}_3(1)$ consists of exactly 3 elements such that the barycenter of the image of $\mathcal{I}_{A,B,C}$ are $-1,-1/2,0$, respectively. We take $\mathcal{V}(A,B,C) = [-1,-3/5]$, $[-4/5,-1/5]$, $[-2/5,0]$, respectively, for example.

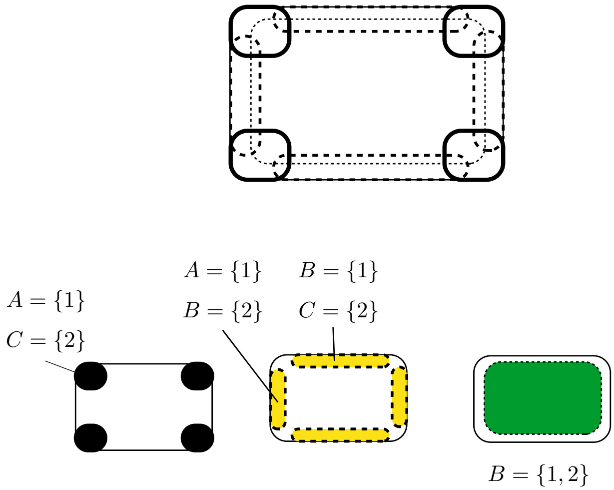


FIGURE 8. $\mathcal{V}(A, B, C)$

Suppose we defined $\mathcal{V}(A, B, C)$ for $(A, B, C) \in \mathrm{Part}_3(m' - 1)$ with $m' < m$. We consider $\mathcal{V}(A, B, C)$ for $(A, B, C) \in \mathrm{Part}_3(m - 1)$. The induction hypothesis and Definition 5.7 (2)(5) determine $\mathcal{V}(A, B, C) \cap \partial[-1, 0]^{m-1}$. For example, in the case $\mathcal{A}_0 = (A, B, C)$ where $B = \underline{m-1}$, $A = C = \emptyset$, we have $\mathcal{V}(\mathcal{A}_0) \cap \partial[-1, 0]^{m-1} = \emptyset$. We can define $\mathcal{V}(A, B, C)$ for the case $(A, B, C) \neq \mathcal{A}_0$ by taking a small neighborhood of $\mathcal{V}(A, B, C) \cap \partial[-1, 0]^{m-1}$. Properties (2)(4)(5) hold by construction and we can choose $\mathcal{V}(A, B, C)$ so that (3)(6) hold also. The union of the interiors of such $\mathcal{V}(A, B, C)$'s contains $\partial[-1, 0]^{m-1}$. Now we can choose $\mathcal{V}(\mathcal{A}_0)$ so that all of (1)–(6) are satisfied. See Fig. 8. $\square$

Example 5.9. Here is an example of $\mathcal{V}(A, B, C)$. Recall $b = \#B$. If $i \in A$, $-1 \leq t_i \leq (1 - 3^{b+1})/3^{b+1}$. If $i \in B$, $(2 - 3^{b+1})/3^{b+1} \leq t_i \leq -2/3^{b+1}$. If $i \in C$, $-1/3^{b+1} \leq t_i \leq 0$.

Now we take and fix a system $\{\mathcal{V}(A, B, C)\}$ of closed subsets in Definition 5.7. Let $(\mathbf{q}, \vec{t}) \in \mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha})^+$.

Definition 5.10. For each given point $\vec{t} \in [-1, 0]^{m-1}$, we put

$$\mathcal{B}(\vec{t}) = \{(A, B, C) \in \mathrm{Part}_3(m - 1) \mid \vec{t} \in \mathcal{V}(A, B, C)\}.$$

To each $(\mathbf{q}, \vec{t}) \in \mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha})^+$ and $\mathcal{A} \in \mathcal{B}(\vec{t})$, we are going to define a finite dimensional linear subspace

$$E_{(\mathbf{q}, \vec{t})}(\mathbf{q}; \mathcal{A}) \subset C^\infty(\Sigma_{\mathbf{q}}, u_{\mathbf{q}}^* TX \otimes \Lambda^{0,1}).$$

Their direct sum will be the obstruction space of the Kuranishi chart at $(\mathbf{q}, \vec{t})$. See Definition 5.17. The way we do so is similar to the argument in Sect. 4 but is slightly more complicated because of describing combinatorial patterns of corners of the moduli spaces in terms of outer collars.

We recall

$$\alpha_- = \alpha_0, \alpha_1, \ldots, \alpha_{m-1}, \alpha_m = \alpha_+ \in \mathfrak{A}$$

and put

$$\vec{\alpha} = (\alpha_0, \ldots, \alpha_m).$$

Let $B \subseteq \underline{m-1}$. We will construct a Kuranishi structure on $\mathcal{M}_\ell(X, H; \vec{\alpha})$. The way we do so is the same as the proof of Theorem 4.1. The construction in the proof of Theorem 4.1 involves various choices. We take different choices for different B. The choice in the case when $A = \emptyset$ and $B = \emptyset$ is exactly the same as one taken during the proof of Theorem 4.1. The detail follows. First we consider the case $A = \emptyset$.[20]

Choice 5.11. (1) We take a finite set.

$$\mathcal{A}_{\vec{\ell}}(\vec{\alpha}; B) = \{\mathbf{p}_c \mid c \in \mathfrak{C}_{\vec{\ell}}(\vec{\alpha}, B)\} \subset \mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha}).$$

Here $\mathfrak{C}_{\vec{\ell}}(\vec{\alpha}, B)$ is a certain index set which will be taken as in Condition 5.12.

(2) For each $\mathbf{p}_c \in \mathcal{A}_{\vec{\ell}}(\vec{\alpha}; B)$ we take its closed neighborhood $W(\mathbf{p}_c; \vec{\ell}, \vec{\alpha}, B)$ in $\mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha})$ which is sufficiently small so that Lemma 4.22 holds for $\mathbf{p} = \mathbf{p}_c$. We also take obstruction bundle data $\mathfrak{E}_{\mathbf{p}_c}(\vec{\ell}, \vec{\alpha}, B)$ centered at $\mathbf{p}_c$ for each $c \in \mathfrak{C}_{\vec{\ell}}(\vec{\alpha}, B)$.

Let $B = \{i(B,1), \ldots, i(B,b)\}$ with $1 \le i(B,1) < \cdots < i(B,b) \le m-1$ and we put

$$\begin{aligned}
\vec{\alpha}(B) &= (\alpha_-, \alpha_{i(B,1)}, \ldots, \alpha_{i(B,b)}, \alpha_+), \\
\ell_j(B) &= \ell_{i(B,j-1)+1} + \cdots + \ell_{i(B,j)}, \\
\vec{\ell}(B) &= (\ell_1(B), \ldots, \ell_{b+1}(B)).
\end{aligned}$$

Here we put $i(B,0) = 0$ and $i(B, b+1) = m$ by convention. We note

$$\mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha}) \subseteq \mathcal{M}_{\vec{\ell}(B)}(X, H; \vec{\alpha}(B)).$$

Condition 5.12. We require that the objects taken in Choice 5.11 have the following properties.

(1) If $\vec{\alpha} = (\alpha_-, \alpha_+)$, the choices are exactly the same as we took during the proof of Theorem 4.1. Namely the choices of $\mathcal{A}_\ell(\alpha_-, \alpha_+)$, $\mathfrak{C}_\ell(\alpha_-, \alpha_+)$ and $W(\mathbf{p}_c)$ are made as in Choice 4.23.

(2)

$$\mathcal{A}_{\vec{\ell}}(\vec{\alpha}; B) = \mathcal{A}_{\vec{\ell}(B)}(\vec{\alpha}(B); \underline{b}) \cap \mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha}). \tag{5.9}$$

[20] When $A \neq \emptyset$, we will apply Choice 5.11 for each moduli space $\mathcal{M}(X, H, \alpha_{i(A,j-1)}, \alpha_{i(A,j)})$ by restricting B to the subset $B'_j(A) := B \cap (i(A, j-1), i(A,j))_{\mathbb{Z}}$ with $j = 1, \ldots, a+1$.

(3) If $\mathbf{p}_c$ is an element of (5.9), we have

$$W(\mathbf{p}_c;\vec{\ell},\vec{\alpha},B) = W(\mathbf{p}_c;\vec{\ell}(B),\vec{\alpha}(B),\underline{b}) \cap \mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha}).$$

Moreover,

$$\mathfrak{E}_{\mathbf{p}_c}(\vec{\ell},\vec{\alpha},B) = \mathfrak{E}_{\mathbf{p}_c}(\vec{\ell}(B),\vec{\alpha}(B),\underline{b}).$$

(4) For each B

$$\bigcup_{\mathbf{p}_c \in \mathcal{A}_{\vec{\ell}}(\vec{\alpha},B)} \operatorname{Int} W(\mathbf{p}_c;\vec{\ell},\vec{\alpha},B) = \mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha}). \tag{5.10}$$

We need to require one more condition (Lemma 5.15 Condition $(*)$) which will be given later.

Now we describe the procedure of associating an obstruction space to an element $(\mathbf{q},\vec{t})$ when Choice 5.11 is given. We first review the procedure we have taken in Sect. 4. Let $\mathbf{q} \in \mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha})$.

(i) We put

$$G(\mathbf{q};\vec{\ell},\vec{\alpha},B) = \{\mathbf{p}_c \mid \mathbf{q} \in W(\mathbf{p}_c;\vec{\ell},\vec{\alpha},B)\}. \tag{5.11}$$

This is the same as in Definition 4.25 (1).

(ii) For each $\mathbf{p}_c \in G(\mathbf{q};\vec{\ell},\vec{\alpha},B)$ we take $\vec{w}^{\mathbf{q}}_{\mathbf{p}_c} \subset \Sigma_{\mathbf{q}}$ such that $\mathbf{q} \cup \vec{w}^{\mathbf{q}}_{\mathbf{p}_c}$ is ϵ_c-close to $\mathbf{p}_c \cup \vec{w}_{\mathbf{p}_c} \cup \vec{w}_{\rm can}$ and satisfies the transversal constraint. (Definition 4.19.) We note that such $\vec{w}^{\mathbf{q}}_{\mathbf{p}_c}$ uniquely exists if we take $W(\mathbf{p}_c;\vec{\ell},\vec{\alpha},B)$ sufficiently small. (Lemma 4.22.)

(iii) If $(\mathfrak{Y},u',\varphi') \cup \vec{w}'_c$ is ϵ-close to $\mathbf{q} \cup \vec{w}^{\mathbf{q}}_{\mathbf{p}_c}$, then by Definition 4.21 we have a complex linear embedding

$$I^{\vec{\ell},\vec{\alpha},B}_{\mathbf{p}_c,\mathrm{v};\Sigma',u',\varphi'} : E_{\mathbf{p}_c,\mathrm{v}}(\mathfrak{y}_{\mathbf{p}_c};\vec{\ell},\vec{\alpha},B) \to C^{\infty}(\Sigma';(u')^*TX \otimes \Lambda^{0,1}). \tag{5.12}$$

See Definition 4.25 (3). Here $(\mathfrak{Y} \cup \vec{w}'_c,\varphi') = \Phi_{\mathbf{p}_c}(\mathfrak{y}_{\mathbf{p}_c},\vec{T}_{\mathbf{p}_c},\vec{\theta}_{\mathbf{p}_c})$ as in Notation 4.15. We include $\vec{\ell},\vec{\alpha},B$ in the notation since the map depends on them.

Now we go back to our situation. We will sum up the images of the maps (5.12) not only for various $\mathbf{p}_c$ and v but also for various $(A,B,C) \in \mathcal{B}(\vec{t})$. We describe this process now.

Let $(\mathbf{q},\vec{t}) \in \mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha})^+$ and $\vec{\alpha} = (\alpha_0,\dots,\alpha_m)$ as before. Since $\mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha})$ is a fiber product with m-factors, $\mathbf{q}$ is decomposed into factors $\mathbf{q}_i$, $i=1,\dots,m$. Let $(A,B,C) \in \mathcal{B}(\vec{t})$. We put $A = \{i(A,1),\dots,i(A,a)\}$ with $i(A,1) < \dots < i(A,a)$. We define

$$\mathbf{q}_{(A,j)} = (\mathbf{q}_{i(A,j-1)+1},\dots,\mathbf{q}_{i(A,j)}) \in \mathcal{M}_{\vec{\ell}_{A,j}}(X,H;\vec{\alpha}_{A,j}). \tag{5.13}$$

Recall $\vec{\alpha}_{A,j} = (\alpha_{i(A,j-1)},\dots,\alpha_{i(A,j)})$ and $\vec{\ell}_{A,j} = (\ell_{i(A,j-1)+1},\dots,\ell_{i(A,j)})$. We put $m_j(A) = i(A,j) - i(A,j-1)$. We also put $B'_j(A) = B \cap (i(A,j-1),i(A,j))_{\mathbb{Z}}$ and

$$B_j(A) = \{i - i(A,j-1) \mid i \in B'_j(A)\}.$$

Note that we made choices for $\vec{\ell}_{A,j},\vec{\alpha}_{A,j}$ and $B_j(A)$ as in Choice 5.11.

Definition 5.13. For $(\mathbf{q},\vec{t}) \in \mathcal{M}_{\vec{\ell}}(X,H,\vec{\alpha})^+$ we define a set $\widetilde{\mathfrak{F}}(\mathbf{q},\vec{t})$ by

$$\widetilde{\mathfrak{F}}(\mathbf{q},\vec{t}) = \{((A,B,C),j) \mid (A,B,C) \in \mathcal{B}(\vec{t}),\ j = 1,\dots,a+1 = \#A+1\}.$$

We define an equivalence relation $\sim$ on it as follows. Let $((A(k),B(k),C(k)), j(k))$ be elements of this set for $k=1,2$. Then

$$((A(1),B(1),C(1)),j(1)) \sim ((A(2),B(2),C(2)),j(2))$$

if and only if the following (1) (2) hold. We put $A(k) = \{i(A(k),1),\dots,i(A(k), a(k))\}$ with $i(A(k),1) < \cdots < i(A(k),a(k))$.

(1) $i(A(1),j(1)-1) = i(A(2),j(2)-1)$. $i(A(1),j(1)) = i(A(2),j(2))$.
(2) $B(1) \cap (i(A(1),j(1)-1), i(A(1),j(1)))_{\mathbb{Z}} = B(2) \cap (i(A(2),j(2)-1), i(A(2),j(2)))_{\mathbb{Z}}$.

(Note that it automatically implies $\vec{\ell}_{A(1),j(1)} = \vec{\ell}_{A(2),j(2)}$.) The conditions (1)(2) imply that the map (5.14) below is independent of the $\sim$ equivalence class.

Now we put[21]

$$\mathfrak{F}(\mathbf{q},\vec{t}) = \widetilde{\mathfrak{F}}(\mathbf{q},\vec{t})/\sim.$$

For $\mathfrak{z} \in \mathfrak{F}(\mathbf{q},\vec{t})$ the three objects $\vec{\alpha}_{A,j}$, $B_j(A)$, $\vec{\ell}_{A,j}$ and $\mathbf{q}_{(A,j)}$ are determined in a way independent of the representatives. We write them as $\vec{\alpha}_{\mathfrak{z}}$, $B(\mathfrak{z})$, $\vec{\ell}_{\mathfrak{z}}$ and $\mathbf{q}_{\mathfrak{z}}$.

Let $\mathfrak{z} \in \mathfrak{F}(\mathbf{q},\vec{t})$ and $\mathbf{p}_c \in G(\mathbf{q}_{\mathfrak{z}};\vec{\ell}_{\mathfrak{z}},\vec{\alpha}_{\mathfrak{z}},B(\mathfrak{z}))$. Then we obtain the linear map (5.12) which is

$$I^{\vec{\ell}_{\mathfrak{z}},\vec{\alpha}_{\mathfrak{z}},B(\mathfrak{z})}_{\mathbf{p}_c,\mathrm{v};\Sigma_{\mathbf{q}_{\mathfrak{z}}},u_{\mathbf{q}_{\mathfrak{z}}},\varphi_{\mathbf{q}_{\mathfrak{z}}}} : E_{\mathbf{p}_c,\mathrm{v}}(\mathfrak{y}_{\mathbf{p}_c};\vec{\ell},\vec{\alpha},B) \to C^\infty(\Sigma_{\mathbf{q}_{\mathfrak{z}}};u_{\mathfrak{z}}^*TX \otimes \Lambda^{0,1}). \tag{5.14}$$

Here $(\Sigma_{\mathbf{q}_{\mathfrak{z}}} \cup \vec{z}_{\mathbf{q}_{\mathfrak{z}}} \cup \vec{w}^{\mathbf{q}_{\mathfrak{z}}}_{\mathbf{p}_c}) = \Phi_{\mathbf{p}_c}(\mathfrak{y}_{\mathbf{p}_c},\vec{T}_{\mathbf{p}_c},\vec{\theta}_{\mathbf{p}_c})$[22] and $\Sigma_{\mathbf{q}_{\mathfrak{z}}}$, $u_{\mathbf{q}_{\mathfrak{z}}}$, $\varphi_{\mathbf{q}_{\mathfrak{z}}}$ are the source curve, the map to X, and the parametrization of the mainstream, which are parts of $\mathbf{q}_{\mathfrak{z}}$, respectively. Note that the target space of the map (5.14) is a subset of $C^\infty(\Sigma_{\mathbf{q}};u^*TX \otimes \Lambda^{0,1})$ and is the sum of the set of smooth sections of the irreducible components.

Remark 5.14. The notion of stabilization data is similar to the obstruction bundle data (Definition 4.11), except we do not include obstruction spaces (Definition 4.11 (5)(6)).

Lemma 5.15. *We can achieve the choices laid out in Choice 5.11 so that Condition 5.12 and the following condition (∗) are satisfied.*

(*) *The sum of the images of the map (5.14) for various $\mathfrak{z} \in \mathfrak{F}(\mathbf{q},\vec{t})$, $\mathbf{p}_c \in G(\mathbf{q}_{\mathfrak{z}};\vec{\ell}_{\mathfrak{z}},\vec{\alpha}_{\mathfrak{z}},B(\mathfrak{z}))$ and irreducible components v of $\mathbf{p}_c$ is a direct sum in $C^\infty(\Sigma_{\mathbf{q}};u^*TX \otimes \Lambda^{0,1})$.*

[21] The sets $\mathcal{V}(A,B,C)$ are used to define $\mathcal{B}(\vec{t})$. Then $\mathcal{B}(\vec{t})$ is used to define $\widetilde{\mathfrak{F}}(\mathbf{q},\vec{t})$.
[22] As we noticed in Notation 4.15, $\mathfrak{y}_{\mathbf{p}_c}$ etc on the right hand side also depend on $\mathbf{q}_{\mathfrak{z}}$ in this case, but we omit $\mathbf{q}_{\mathfrak{z}}$ from the notations when no confusion can occur.

We will prove Lemma 5.15 in the next subsection. Assuming it for the moment, we continue the proof of Proposition 5.5.

Note that for each $\mathfrak{z} \in \mathfrak{F}(\mathbf{q}, \vec{t})$ and $\mathbf{p}_c \in G(\mathbf{q}_{\mathfrak{z}}; \vec{\ell}_{\mathfrak{z}}, \vec{\alpha}_{\mathfrak{z}}, B(\mathfrak{z}))$ we took additional marked points $\vec{w}_{\mathbf{p}_c}^{\mathbf{q}_{\mathfrak{z}}}$ on $\mathbf{q}_{\mathfrak{z}}$. These marked points can be regarded as marked points on $\mathbf{q}$. We also take stabilization data centered at $\mathbf{q}$. In particular, we take $\vec{w}_{\mathbf{q}}$.

Situation 5.16. • We have $(\mathfrak{Y}, u', \varphi')$ and marked points $\vec{w}'_{\mathbf{p}_c}$ on the source curve of $\mathfrak{Y}'$ for each $\mathfrak{z} \in \mathfrak{F}(\mathbf{q}, \vec{t})$ and $\mathbf{p}_c \in G(\mathbf{q}_{\mathfrak{z}}; \vec{\ell}_{\mathfrak{z}}, \vec{\alpha}_{\mathfrak{z}}, B(\mathfrak{z}))$. We also take $\vec{w}'_{\mathbf{q}}$.
• For each $\mathfrak{z} \in \mathfrak{F}(\mathbf{q}, \vec{t})$ and $\mathbf{p}_c \in G(\mathbf{q}_{\mathfrak{z}}; \vec{\ell}_{\mathfrak{z}}, \vec{\alpha}_{\mathfrak{z}}, B(\mathfrak{z}))$, we assume that $(\mathfrak{Y}, u', \varphi') \cup \vec{w}'_{\mathbf{p}_c}$ is ϵ_1-close to $\mathbf{q} \cup \vec{w}_{\mathbf{p}_c}^{\mathbf{q}_{\mathfrak{z}}}$.
• We assume that $(\mathfrak{Y}, u', \varphi') \cup \vec{w}'_{\mathbf{q}}$ is ϵ_1-close to $\mathbf{q} \cup \vec{w}_{\mathbf{q}}$.
• We assume that $\mathfrak{Y}$ is decomposed into m extended mainstream components $\mathfrak{Y}_i$ $(i = 1, \dots, m)$, and $(\mathfrak{Y}_i, u', \varphi') \cup (\vec{w}'_{\mathbf{q}} \cap \mathfrak{Y}_i)$ is ϵ_1-close to $\mathbf{q}_i \cup (\vec{w}_{\mathbf{q}} \cap \mathbf{q}_i)$. □

In Situation 5.16 we have

$$I^{\vec{\ell}_{\mathfrak{z}}, \vec{\alpha}_{\mathfrak{z}}, B(\mathfrak{z})}_{\mathbf{p}_c, \mathrm{v}; \Sigma', u', \varphi'} : E_{\mathbf{p}_c, \mathrm{v}}(\mathfrak{y}'_{\mathbf{p}_c}; \vec{\ell}, \vec{\alpha}, B) \to C^\infty(\Sigma'; (u')^*TX \otimes \Lambda^{0,1}) \tag{5.15}$$

in the same way as in (5.12). Here $\mathfrak{Y} \cup \vec{w}'_{\mathbf{p}_c} = \Phi_{\mathbf{p}_c}(\mathfrak{y}'_{\mathbf{p}_c}, \vec{T}'_{\mathbf{p}_c}, \theta'_{\mathbf{p}_c})$ so this map depends also on $\vec{w}'_{\mathbf{p}_c}$.

Definition 5.17. For each $(\mathbf{q}, \vec{t}) \in \mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha})^+$ we define a linear subspace $E((\mathfrak{Y}, u', \varphi') \cup \bigcup \vec{w}'_{\mathbf{p}_c}; \mathbf{q})$ of $C^\infty(\Sigma'; (u')^*TX \otimes \Lambda^{0,1})$ by the sum of all the images of the map (5.15) for various v, $\mathfrak{z} \in \mathfrak{F}(\mathbf{q}, \vec{t})$ and $\mathbf{p}_c \in G(\mathbf{q}_{\mathfrak{z}}; \vec{\ell}_{\mathfrak{z}}, \vec{\alpha}_{\mathfrak{z}}, B(\mathfrak{z}))$. We call it the *obstruction space* of Kuranishi chart at $(\mathbf{q}, \vec{t})$.

5.3. Proof of Proposition 5.5 II: Kuranishi chart and coordinate change

We now define Kuranishi charts of our Kuranishi structure.

Definition 5.18. Let $(\mathbf{q}, \vec{t}) \in \mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha})^+$. In Situation 5.16 we consider the following conditions on an object $(((\mathfrak{Y}, u', \varphi') \cup \bigcup_{\mathfrak{z},c} \vec{w}'_{\mathbf{p}_c} \cup \vec{w}'_{\mathbf{q}}), \vec{t}')$:

(1) If Σ_a is the mainstream component of $\mathfrak{Y}$ and φ'_a is a parametrization of this mainstream component (which is a part of given φ'), the following equation is satisfied on $\mathbb{R} \times S^1$.

$$\begin{aligned} &\frac{\partial (u' \circ \varphi'_a)}{\partial \tau} + J\left(\frac{\partial (u' \circ \varphi'_a)}{\partial t} - \mathfrak{X}_{H_t} \circ u' \circ \varphi'_a\right) \\ &\equiv 0 \mod E((\mathfrak{Y}, u', \varphi') \cup \bigcup \vec{w}'_{\mathbf{p}_c}; \mathbf{q}). \end{aligned} \tag{5.16}$$

(2) If Σ'_{v} is a bubble component of $\mathfrak{Y}$, the following equation is satisfied on Σ'_{v}.

$$\overline{\partial} u' \equiv 0 \mod E((\mathfrak{Y}, u', \varphi') \cup \bigcup \vec{w}'_{\mathbf{p}_c}; \mathbf{q}). \tag{5.17}$$

(3) For each $\mathfrak{z} \in \mathfrak{F}(\mathbf{q}, \vec{t})$ and $\mathbf{p}_c \in G(\mathbf{q}_{\mathfrak{z}}; \vec{\ell}_{\mathfrak{z}}, \vec{\alpha}_{\mathfrak{z}}, B(\mathfrak{z}))$ the additional marked points $\vec{w}'_{\mathbf{p}_c}$ satisfy the transversal constraint in Definition 4.19 with respect to $\mathbf{p}_c$.

(4) The additional marked points $\vec{w}'_{\mathbf{q}}$ satisfy the transversal constraint in Definition 4.19 with respect to $\mathbf{q}$.

(5) For each $\mathfrak{z}$ and $c \in \mathcal{E}(\mathbf{q})$, $(\mathfrak{Y}, u', \varphi') \cup \vec{w}'_{\mathbf{p}_c} \cup \vec{w}'_{\mathbf{q}}$ is ϵ_1-close to $\mathbf{q} \cup \vec{w}^{\mathbf{q}_{\mathfrak{z}}}_{\mathbf{p}_c} \cup \vec{w}_{\mathbf{q}}$.
(6) $|\vec{t}' - \vec{t}| < \epsilon_1$.

The set of isomorphism classes of $(((\mathfrak{Y}, u', \varphi') \cup \bigcup_{\mathfrak{z},c} \vec{w}'_{\mathbf{p}_c} \cup \vec{w}'_{\mathbf{q}}), \vec{t}')$ satisfying Conditions (1) - (6) above is denoted by

$$V((\mathbf{q}, \vec{t}), \epsilon_1)$$

where the isomorphism is defined in the same way as in Definition 4.28.

We can prove that $V((\mathbf{q}, \vec{t}), \epsilon_1)$ is a smooth manifold in the way similar to Lemma 4.29. It is easy to see that it is $\mathrm{Aut}(\mathbf{q})$ invariant and so we obtain an orbifold $U((\mathbf{q}, \vec{t}), \epsilon_1) = V((\mathbf{q}, \vec{t}), \epsilon_1)/\mathrm{Aut}(\mathbf{q})$. We define a vector bundle $E((\mathbf{q}, \vec{t}), \epsilon_1)$ on it by taking $E((\mathfrak{Y}, u', \varphi') \cup \bigcup \vec{w}'_{\mathbf{p}_c}; \mathbf{q})$ as the fiber. Then the left hand side of (5.16) and (5.17) define its smooth section, which we denote by $s_{(\mathbf{q},\vec{t}),\epsilon_1}$. An element of its zero set determines an element of $\mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha})^+$. We thus obtain $\psi_{(\mathbf{q},\vec{t}),\epsilon_1}$.

We can prove that $(U((\mathbf{q}, \vec{t}), \epsilon_1), E((\mathbf{q}, \vec{t}), \epsilon_1), s_{(\mathbf{q},\vec{t}),\epsilon_1}, \psi_{(\mathbf{q},\vec{t}),\epsilon_1})$ is a Kuranishi chart of $(\mathbf{q}, \vec{t})$ in the same way as in Proposition 4.30.

We can define a coordinate change among them and show the compatibility among them in the same way as in Lemmas 4.33 and 4.34. We finally adjust the size of the Kuranishi neighborhoods $\{V((\mathbf{q}, \vec{t}), \epsilon_1)\}_{(\mathbf{q},\vec{t})}$ to obtain a Kuranishi structure on $\mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha})^+$ in the same way again as in Lemmas 4.33 and 4.34.

Condition 5.4 is a consequence of the Property (5) in Definition 5.7 and the way we used $\mathcal{V}(A, B, C)$ to define $\mathcal{B}(\vec{t})$ and $\mathfrak{F}(\mathbf{q}, \vec{t})$. Let us elaborate on this point.

If $t_i = 0$, then $i \in C$. Therefore, we can apply Condition 5.12 (2)(3) to see that the obstruction spaces are restrictions on those which we defined on the moduli space obtained by performing the gluing at the corresponding transit points.

If $t_i = -1$, then $i \in A$. Therefore, we are taking fiber product Kuranishi structure at the corresponding transit points. (See (5.13) and Definition 5.13.)

Proposition 5.5 (2) is a consequence of Definition 5.7 (6) and the construction.

Proposition 5.5 (3) is a consequence of Condition 5.12 (1).

Therefore, to complete the proof of Proposition 5.5 and of Theorem 2.9 it remains to prove Lemma 5.15.

Proof of Lemma 5.15. The proof is by induction on m, where we recall $\#\vec{\alpha} = m - 1$. If $m = 1$ then Choice 5.11 is given during the proof of Theorem 4.1. We suppose that we made the choice for m' that is smaller than m and we will prove the case of m. We also assume that, as a part of the induction hypothesis, the conclusion of Lemma 5.15 holds if the number of components of $\vec{\alpha}$ is strictly smaller than $m + 1$.

We now consider $\vec{\alpha}$ with $(m + 1)$ components and $\vec{\ell}$. Let $\vec{t} \in [-1, 0]^{m-1}$ and $\mathbf{q} \in \mathcal{M}_{\vec{\ell}}(X, H; \vec{\alpha})$. We define a relation $<$ on the set $\mathcal{B}(\vec{t})$ in Definition 5.10 such that $(A', B', C') < (A, B, C)$ if and only if $\partial(\mathrm{Im}\, \mathcal{I}_{A,B,C}) \supseteq \mathrm{Im}\, \mathcal{I}_{A',B',C'}$.

By Definition 5.7 (4) the set $\mathcal{B}(\vec{t})$ is linearly ordered by this relation $<$. Therefore, there exists a maximal element, which we denote by (A_0, B_0, C_0).
Case 1: $B_0 \neq \underline{m-1}$.
Case 1-1: $A_0 \neq \emptyset$. Let $i \in A_0$. We note that if $(A, B, C) \in \mathcal{B}(\vec{t})$ then $\mathrm{Im}\,\mathcal{I}_{A,B,C}$ is a face of $\mathrm{Im}\,\mathcal{I}_{A_0,B_0,C_0}$. Since $\mathrm{Im}\,\mathcal{I}_{A_0,B_0,C_0} \subseteq \{t_i = -1\}$, we have $\mathrm{Im}\,\mathcal{I}_{A,B,C} \subseteq \{t_i = -1\}$. Therefore, $i \in A$ by Definition 5.7 (4).

We consider $\vec{\alpha}_1 = (\alpha_0, \dots, \alpha_i)$, $\vec{\alpha}_2 = (\alpha_i, \dots, \alpha_m)$ and $\vec{\ell}_1 = (\ell_0, \dots, \ell_i)$, $\vec{\ell}_2 = (\ell_i, \dots, \ell_m)$. We have

$$(\mathbf{q}, \vec{t}) \in \mathcal{M}_{\vec{\ell}_1}(X, H; \vec{\alpha}_1)^+ \times_{R_{\alpha_i}} \mathcal{M}_{\vec{\ell}_2}(X, H; \vec{\alpha}_2)^+.$$

We decompose $(\mathbf{q}, \vec{t})$ into $(\mathbf{q}_1, \vec{t}_1)$ and $(\mathbf{q}_2, \vec{t}_2)$. Since $i \in A$ for all the elements $(A, B, C) \in \mathcal{B}(\vec{t})$, we find easily that

$$\mathfrak{F}(\mathbf{q}, \vec{t}) \cong \mathfrak{F}(\mathbf{q}_1, \vec{t}_1) \sqcup \mathfrak{F}(\mathbf{q}_2, \vec{t}_2).$$

Therefore, by induction hypothesis we can show that Lemma 5.15 Condition $(*)$ holds at this element $(\mathbf{q}, \vec{t})$.
Case 1-2: $C_0 \neq \emptyset$. Let $i \in C_0$. In the same way as in Case 1-1 we can show $i \in C$ for all $(A, B, C) \in \mathcal{B}(\vec{t})$. Then we put $\vec{\alpha}' = (\alpha_0, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_m)$, $\vec{\ell}' = (\ell_1, \dots, \ell_{i-1}, \ell_i + \ell_{i+1}, \ell_{i+2}, \dots, \ell_m)$. We may identify $(\mathbf{q}, \vec{t})$ as an element $(\mathbf{q}', \vec{t}')$ of $\mathcal{M}_{\vec{\ell}'}(X, H; \vec{\alpha}')^+$. Then we have $\mathfrak{F}(\mathbf{q}', \vec{t}') \cong \mathfrak{F}(\mathbf{q}, \vec{t})$. Therefore, we can check Lemma 5.15 Condition $(*)$ by induction hypothesis.
Case 2: $B_0 = \underline{m-1}$. In this case $A_0 = C_0 = \emptyset$. Let (A_1, B_1, C_1) be the maximal element of $\mathcal{B}(\vec{t}) \backslash \{(A_0, B_0, C_0)\}$. Using the argument of Case 1, we can show that the sum of the images of $I^{\vec{\ell}_{\mathfrak{z}}, \vec{\alpha}_{\mathfrak{z}}, B(\mathfrak{z})}_{\mathbf{p}_c, \mathrm{v}; \Sigma_{\mathbf{q}_{\mathfrak{z}}}, u_{\mathbf{q}_{\mathfrak{z}}}, \varphi_{\mathbf{q}_{\mathfrak{z}}}}$ for $\mathfrak{z} \in \mathcal{B}(\vec{t}) \backslash \{(A_0, B_0, C_0)\}$, $\mathbf{p}_c \in G(\mathbf{q}; \vec{\ell}, \vec{\alpha}, B)$ and v (irreducible components of $\mathbf{p}_c$), is a direct sum.

Now we can make our choice for (A_0, B_0, C_0) so that Lemma 5.15 Condition $(*)$ also holds in this case. (See [23, Lemma 11.7] for example.)

The proof of Lemma 5.15 is now complete. $\square$

The proof of Proposition 5.5 is now complete. $\square$

6. Construction of morphism

6.1. Statement

Let H^1, H^2 be two functions $X \times S^1 \to \mathbb{R}$ which are Morse–Bott nondegenerate in the sense of Condition 2.1. We put

$$\widetilde{\mathrm{Per}}(H^j) = \coprod_{\alpha \in \mathfrak{A}^j} R^j_\alpha \tag{6.1}$$

as in (2.6) for $j = 1, 2$. We define a local system $o_{R^j_\alpha}$ on each R^j_α as in Definition 2.6. Let J_1, J_2 be two almost complex structures tamed by ω. Using them we obtain linear K-systems $\mathcal{F}_j := \mathcal{F}_X(H^j, J_j)$ $(j = 1, 2)$ by Theorem 2.9 whose spaces of connecting orbits are $\mathcal{M}(X, J_j, H^j; \alpha_-, \alpha_+)^{\boxplus 1}$. In this section we will construct a morphism $\mathfrak{N}_{21} : \mathcal{F}_1 \to \mathcal{F}_2$ of linear K-systems [23, Definition 16.19].

Situation 6.1. We consider a smooth function[23] $H^{21} : X \times \mathbb{R} \times S^1 \to \mathbb{R}$ such that:

(1) If $\tau < -1$, then $H^{21}(x,\tau,t) = H^1(x,t)$.
(2) If $\tau > 1$, then $H^{21}(x,\tau,t) = H^2(x,t)$.

We put $H^{21}_{\tau,t}(x) = H(x,\tau,t)$ and denote by $\mathfrak{X}_{H^{21}_{\tau,t}}$ the Hamiltonian vector field associated to $H_{\tau,t}$.

We also consider a one parameter family $\mathcal{J}^{21} = \{J^{21}_\tau\}$ of almost complex structures tamed by ω such that:

(i) If $\tau < -1$, then $J^{21}_\tau = J_1$.
(ii) If $\tau > 1$, then $J^{21}_\tau = J_2$.

When no confusion can occur, we simply write $\mathcal{J} = \mathcal{J}^{21}$ in this section. $\square$

Definition 6.2. Suppose we are in Situation 6.1. Let $\alpha_- \in \mathfrak{A}_1$ and $\alpha_+ \in \mathfrak{A}_2$. We consider the set of smooth maps $u : \mathbb{R} \times S^1 \to X$ with the following properties.

(1) It satisfies

$$\frac{\partial u}{\partial \tau} + J^{21}_\tau \left(\frac{\partial u}{\partial t} - \mathfrak{X}_{H^{21}_{\tau,t}} \circ u \right) = 0. \tag{6.2}$$

Here τ and t are the coordinates of $\mathbb{R}$ and $S^1 = \mathbb{R}/\mathbb{Z}$, respectively.

(2) There exist $\tilde{\gamma}^- = (\gamma^-, w^-) \in R^1_{\alpha_-}$ and $\tilde{\gamma}^+ = (\gamma^+, w^+) \in R^2_{\alpha_+}$ such that

$$\lim_{\tau \to \pm\infty} u(\tau,t) = \gamma^\pm(t) \tag{6.3}$$

and $w^- \# u \sim w^+$.

We denote by $\mathcal{N}^{\rm reg}(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)$ the totality of such maps u. We define $\mathrm{ev}_\pm : \mathcal{N}^{\rm reg}(X,\mathcal{J},H^{21};\alpha_-,\alpha_+) \to R^1_{\alpha_-}$, $R^2_{\alpha_+}$ by $\mathrm{ev}_\pm(u) = \tilde{\gamma}^\pm$.

Lemma 6.3. *Suppose $u : \mathbb{R} \times S^1 \to X$ satisfies (6.2). We assume*

$$\int_{\mathbb{R}\times S^1} \left\| \frac{\partial u}{\partial \tau} \right\|^2 d\tau dt < \infty.$$

Then there exist $\tilde{\gamma}^- = (\gamma^-, w^-) \in R^1_{\alpha_-}$ and $\tilde{\gamma}^+ = (\gamma^+, w^+) \in R^2_{\alpha_+}$ such that (6.3) is satisfied.

The proof is similar to the proof of Proposition 2.4 and is omitted.

Theorem 6.4. *Suppose we are in the situation of Definition 6.2.*

(1) *The space $\mathcal{N}^{\rm reg}(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)$ has a compactification*
$$\mathcal{N}(X,\mathcal{J},H^{21};\alpha_-,\alpha_+).$$
(2) *The compact space $\mathcal{N}(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)$ has a Kuranishi structure with corners. The map* ev *is extended to it as a strongly smooth map.*
(3) *There exists a morphism of linear K-systems $\mathfrak{N}_{21} : \mathcal{F}_1 \to \mathcal{F}_2$ whose interpolation space is $\mathcal{N}(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)^{\boxplus 1}$ given in Definition 6.10 (2).*

[23]The reason we put 2, 1 in this order in the notations H^{21} and $\mathfrak{N}_{21}$ is to be consistent with the order of compositions.

(4) *The Kuranishi structure on* $\mathcal{N}(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)^{\boxplus 1}$ *in (3) coincides with the given one in (2) on* $\mathcal{N}(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+) \subset \mathcal{N}(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)^{\boxplus 1}$.

Proof. The proof of Theorem 6.4 occupies the rest of this section.

6.2. Proof of Theorem 6.4 (1)(2): Kuranishi structure

We begin with defining the compactification $\mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$. Let $(\Sigma, (z_-, z_+, \vec{z}))$ be a genus zero semistable curve with $\ell + 2$ marked points. We define the notion of mainstream as in Definition 3.2. Let Σ_a and $\Sigma_{a'}$ be two mainstream components.

Definition 6.5. We say $a < a'$ if the connected component of $\Sigma \backslash \{z_{a',-}\}$ containing z_- contains Σ_a.[24]

We observe that one of $a < a'$, $a' < a$ or $a = a'$ holds for any pair of mainstream components $(\Sigma_a, \Sigma_{a'})$.

Definition 6.6. The set $\widehat{\mathcal{N}}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$ consists of triples

$$((\Sigma, (z_-, z_+, \vec{z}), a_0), u, \varphi)$$

satisfying the following conditions: Here $\ell = \#\vec{z}$.

(1) $(\Sigma, (z_-, z_+, \vec{z}))$ is a genus zero semi-stable curve with ℓ+2 marked points.
(2) φ is a parametrization of the mainstream.
(3) Σ_{a_0} is one of the mainstream components. We call it the *main component*.
(4) For each extended mainstream component $\widehat{\Sigma}_a$, the map u induces $u_a : \widehat{\Sigma}_a \backslash \{z_{a,-}, z_{a,+}\} \to X$ which is a continuous map.[25]
(5) If Σ_a is a mainstream component and $\varphi_a : \mathbb{R} \times S^1 \to \Sigma_a$ is as above, then the composition $u_a \circ \varphi_a$ satisfies the equation

$$\frac{\partial(u_a \circ \varphi_a)}{\partial \tau} + J_{a,\tau}\left(\frac{\partial(u_a \circ \varphi_a)}{\partial t} - \mathfrak{X}_{H^a_{\tau,t}} \circ (u_a \circ \varphi_a)\right) = 0, \tag{6.4}$$

where

$$H^a_{\tau,t} = \begin{cases} H^1_t & \text{if } a < a_0, \\ H^{21}_{\tau,t} & \text{if } a = a_0, \\ H^2_t & \text{if } a > a_0, \end{cases}$$

and

$$J_{a,\tau} = \begin{cases} J_1 & \text{if } a < a_0, \\ J^{21}_\tau & \text{if } a = a_0, \\ J_2 & \text{if } a > a_0. \end{cases}$$

[24] $z_{a',-}$ and $z_{a',+}$ are transit points of $\Sigma_{a'}$ which are defined in Definition 3.2.
[25] In other words u is a continuous map from the complement of the set of the transit points.

(6)

$$\int_{\mathbb{R}\times S^1} \left\| \frac{\partial(u\circ\varphi_a)}{\partial\tau} \right\|^2 d\mathrm{d}\tau\mathrm{d}t < \infty.$$ [26]

(7) Suppose Σ_{v} is a bubble component in $\widehat{\Sigma}_a$. Let $\varphi_a(\tau,t)$ be the root of the tree of sphere bubbles containing Σ_{v}. Then u is J-holomorphic on Σ_{v} where

$$J = \begin{cases} J_1 & \text{if } a < a_0, \\ J^{21}_\tau & \text{if } a = a_0, \\ J_2 & \text{if } a > a_0. \end{cases}$$

(8) If Σ_{a_1} and Σ_{a_2} are mainstream components and $z_{a_1,+} = z_{a_2,-}$, then

$$\lim_{\tau\to+\infty}(u_{a_1}\circ\varphi_{a_1})(\tau,t) = \lim_{\tau\to-\infty}(u_{a_2}\circ\varphi_{a_2})(\tau,t)$$

holds for each $t\in S^1$. ((6) and Lemma 6.3 imply that the left and right hand sides both converge.)

(9) If Σ_a, $\Sigma_{a'}$ are mainstream components and $z_{a,-}=z_-$, $z_{a',+}=z_+$, then there exist $(\gamma_\pm,w_\pm)\in R_{\alpha_\pm}$ such that

$$\lim_{\tau\to-\infty}(u_a\circ\varphi_a)(\tau,t) = \gamma_-(t),$$
$$\lim_{\tau\to+\infty}(u_{a'}\circ\varphi_{a'})(\tau,t) = \gamma_+(t).$$

Moreover,

$$[u_*[\Sigma]]\#w_- = w_+,$$

where $\#$ is the obvious concatenation.

(10) We assume that $((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ is stable in the sense of Definition 6.9 below.

To define stability we first define the group of automorphisms.

Definition 6.7. Assume that $((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ satisfies (1)–(9) above. The *extended automorphism group* $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ of $((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ consists of maps $v:\Sigma\to\Sigma$ with the following properties:

(1) $v(z_-) = z_-$ and $v(z_+) = z_+$. In particular, v preserves each of the mainstream component Σ_a of Σ. Moreover, v fixes each of the transit points.

(2) $u = u\circ v$ holds outside the set of the transit points.

(3) If Σ_a is a mainstream component of Σ, there exists $\tau_a\in\mathbb{R}$ such that

$$(v\circ\varphi_a)(\tau,t) = \varphi_a(\tau+\tau_a,t) \tag{6.5}$$

on $\mathbb{R}\times S^1$.

(4) We require $\tau_{a_0} = 0$.

(5) There exists $\sigma\in\mathrm{Perm}(\ell)$ such that $v(z_i) = z_{\sigma(i)}$.

[26] Condition (6) follows from the rest of the conditions in Definition 6.6 using, e.g., [30, (2.14)]. The same remark holds for Definitions 8.3, 10.4, 10.7, 10.12, 10.14, 10.16.

The *automorphism group* denoted by $\mathrm{Aut}((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ of $((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ consists of the elements of $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ such that σ in Item (5) above is the identity.

Remark 6.8. Definition 6.7 is mostly the same as Definition 3.4. The most important difference is Item (4) where we assume $\tau_{a_0}=0$. Note the equation (6.4) is *not* invariant under the translation of the $\mathbb{R}$ direction on the main component.

Definition 6.9. We say $((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ is *stable* if $\mathrm{Aut}((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ is a finite group. (This is equivalent to the finiteness of $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$.)

Definition 6.10. (1) On the set $\widehat{\mathcal{N}}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)$ we define two equivalence relations $\sim_1$, $\sim_2$. The definition of $\sim_1$ is the same as Definition 3.7. The definition of $\sim_2$ is the same as Definition 3.7 except we require $\tau_{a_0}=0$, in addition. We put

$$\widetilde{\mathcal{N}}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)=\widehat{\mathcal{N}}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)/\sim_1,$$
$$\mathcal{N}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)=\widehat{\mathcal{N}}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)/\sim_2 .$$

In the case $\ell=0$ we write $\mathcal{N}(X,H;\alpha_-,\alpha_+)$ etc.

(2) We define $\mathcal{N}(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)^{\boxplus 1}$ in the same way as in Definition 5.1. Namely, it is the set of equivalence classes of objects $(((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi),\vec{t})$ where $((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)\in\mathcal{N}(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)$ and $\vec{t}$ assigns numbers $t_p\in[-1,0]$ to each transit points.[27]

Definition 6.11. We put X = one point and $H^{21}\equiv 0$. Then we obtain the space $\mathcal{N}_\ell(\text{one point},\mathcal{J},0;\alpha_0,\alpha_0)$. Here α_0 is the unique point in $\mathrm{Per}(0)$. We denote this space by $\mathcal{N}_\ell(\text{source})$.

We remark that $\mathcal{N}_\ell(\text{source})$ is similar to but is different from $\mathcal{M}_\ell(\text{source})$. In fact, $\mathcal{N}_\ell(\text{source})$ includes the data that specify which mainstream component is the main component and also the isomorphism between two elements of $\mathcal{N}_\ell(\text{source})$ is required to be strictly compatible with the parametrization of the main component.

Example 6.12. $\mathcal{N}_0(\text{source})$ is one point. $\mathcal{N}_1(\text{source})$ is $S^1\times[0,1]$. In fact, if there is only one mainstream component and the marked point is $\varphi(\tau,t)$, then the coordinates (τ,t) determine an element of $\mathcal{N}_1(\text{source})$. We compactify it by including the case when there are two mainstream components. In such a case the marked point can not be on the main component. So the S^1 factor of the coordinates of the marked points determine an element of $\mathcal{N}_1(\text{source})$. There are two cases: $a<a_0$ or $a>a_0$. (Here a is the mainstream component which is not the main component.) Thus $\mathcal{N}_1(\text{source})$ is a union of $\mathbb{R}\times S^1$ and two copies of S^1.

We can define a topology on $\mathcal{N}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)$ in the same way as in Definition 3.17 and can prove the following:

[27] This space is the outer collaring of $\mathcal{N}(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)$ in the sense of [23, Definition 17.29].

Lemma 6.13. *The space $\mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$ is compact and Hausdorff.*

Next, we define a Kuranishi structure on the compactification $\mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$. We define the notion of symmetric stabilization of an element $[((\Sigma, (z_-, z_+, \vec{z}), a_0), u, \varphi)]$ of $\mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$ in exactly the same way as in Definition 4.4.

We define the notion of *canonical marked point* $w_{a,\mathrm{can}}$ of a mainstream component Σ_a of Σ such that there is no marked or singular points on Σ_a other than transit point, as follows. If $a \neq a_0$, the definition of $w_{a,\mathrm{can}}$ is exactly the same as Definition 4.8. For $a = a_0$, we do not define canonical marked points.

Remark 6.14. Note that $\mathcal{M}_0(\text{source}) = \emptyset$. This is the reason we need to introduce the canonical marked points. On the other hand, $\mathcal{N}_0(\text{source})$ consists of one point and is not an empty set. Namely the unique element in it is represented by a stable object. This is the reason we do not need to introduce the canonical marked points on the main component.

Let $\vec{w}_{\mathrm{can}}$ be the totality of all the canonical marked points. In the same way as in Lemma 4.9, we can prove that $((\Sigma, (z_-, z_+, \vec{z}), a_0) \cup \vec{w} \cup \vec{w}_{\mathrm{can}}, \varphi)$ is stable[28], where $[((\Sigma, (z_-, z_+, \vec{z}), a_0), u, \varphi)] \in \mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$ and $\vec{w}$ is a symmetric stabilization.

We then define the notion of obstruction bundle data $\mathfrak{E}_{\mathbf{p}}$ for each element $\mathbf{p} \in \mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$ in the same way as in Definition 4.11. Its existence can be proved easily. (See for example [13, Lemma 17.11].)

We next explain how we use the obstruction bundle data $\mathfrak{E}_{\mathbf{p}}$ to define an obstruction space for each object close to $\mathbf{p} \in \mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$.

Let $\mathbf{p} = [((\Sigma, (z_-, z_+, \vec{z}), a_0), u, \varphi)]$. We assume that Σ has exactly k transit points. Taking the condition that the main component is equipped with the parametrization φ_{a_0} into account, we obtain a map

$$\begin{aligned}\Phi_{\mathbf{p}} : \prod_{\mathrm{v}} \mathcal{V}(\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can,v}}) \times (T_0, \infty]^k \times \prod_{j=1}^{m} \left(((T_{0,j}, \infty] \times S^1)/\sim\right) \\ \to \mathcal{N}_{\ell+\ell'+\ell''}(\text{source})\end{aligned} \tag{6.6}$$

in the same way as in (4.8). Here $\mathcal{V}(\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can,v}})$ is an open subset of $\overset{\circ}{\mathcal{M}}{}^{\mathrm{cl}}_*$ or $\overset{\circ}{\mathcal{M}}_*(\text{source})$. See Definition 4.11 (2) (a)(b).

The factor $(T_0, \infty]^k$ parametrizes the way how we smooth the singular points that are the transit points. In (4.8) the similar factor is $D(k; \vec{T}_0)$. The difference is that in the situation of (4.8) the isomorphism $v : \Sigma \to \Sigma$ includes the translation, which is a map v such that $v \circ \varphi_a(\tau, t) = \varphi_a(\tau + \tau_0, t)$. This shifts $T_i \in (T_{0,j}, \infty]$ by τ_0. (Here τ_0 is independent of a.) On the other hand, the isomorphism v here is required to commute strictly with φ_{a_0}. So there is not such a shift. Taking this point into account, the map (6.6) is defined in the same way as in (4.8).

[28] We say $((\Sigma, (z_-, z_+, \vec{z}), a_0) \cup \vec{w} \cup \vec{w}_{\mathrm{can}}, \varphi)$ is *stable* if the automorphism group $\mathrm{Aut}^+((\Sigma, (z_-, z_+, \vec{z}), a_0), \varphi)$ defined as in Definition 6.7 by removing u (the condition (2)) is finite.

Now let

$$(\mathfrak{Y} \cup \vec{w}', \varphi', a_0') = \Phi_{\mathbf{p}}(\mathfrak{y}, \vec{T}, \vec{\theta}) \in \mathcal{N}_{\ell+\ell'+\ell''}(\text{source}) \tag{6.7}$$

where $\mathfrak{y} = (\mathfrak{y}_{\rm v})$. (v is in the set of irreducible components of $\mathbf{p}$.) Here $\vec{w}'$ is the set of the additional marked points corresponding to $\vec{w}$ and $\vec{w}_{\rm can}$. The notation $\mathfrak{Y}$ includes the marked points corresponding to $\vec{z}$ and $z_\pm$. a_0' is the datum to specify the main component and φ' is the parametrization of the mainstream. $(\vec{T}, \vec{\theta}) \in (T_0, \infty]^k \times \prod_{j=1}^m \left(((T_{0,j}, \infty] \times S^1)/\sim\right)$.

Let $u' : \Sigma' \setminus \{\text{transit points}\} \to X$. We assume that $(\mathfrak{Y}, u', \varphi', a_0')$ satisfies Definition 6.6 (1)(2)(3)(4)(8)(9). We can then define the notion that $(\mathfrak{Y} \cup \vec{w}', u', \varphi', a_0')$ is ϵ-close to $\mathbf{p} \cup \vec{w} \cup \vec{w}_{\rm can}$ in the same way as in Definition 4.16. (Note Definition 4.16 (2) is the pseudo-holomorphicity (or pseudo-holomorphicity with Hamiltonian term) at the neck region. We use the almost complex structure specified in Definition 6.6 (5).)

Definition 6.15. We define the *transversal constraint* for $(\mathfrak{Y} \cup \vec{w}', u', \varphi', a_0')$ as follows. Let w_i' be one of the points of $\vec{w}'$. If w_i' corresponds to a point in $\vec{w}$, we require Definition 4.19 (1). If w_i' corresponds to $w_{a,{\rm can}}$ with $a \neq a_0'$, we require Definition 4.19 (2)(3).

Suppose $(\mathfrak{Y} \cup \vec{w}', u', \varphi', a_0')$ is ϵ-close to $\mathbf{p} \cup \vec{w} \cup \vec{w}_{\rm can}$. In the same way as in (4.21), we define a map

$$I_{\mathbf{p},{\rm v};\Sigma',u',\varphi',a_0'} : E_{\mathbf{p},{\rm v}}(\mathfrak{y}) \to C^\infty(\Sigma'; (u')^*TX \otimes \Lambda^{0,1}) \tag{6.8}$$

where $\mathfrak{y}$ is as in (6.7) and v is its irreducible component, and $E_{\mathbf{p},{\rm v}}(\mathfrak{y})$ is defined as a part of the obstruction bundle data $\mathfrak{E}_{\mathbf{p}}$ as in (4.3) (see Definition 4.11 (5)).

Now in the same way as in Choice 4.23, we proceed as follows. We first observe that for any $\mathbf{p} \in \mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$ there exist $\epsilon_{\mathbf{p}} > 0$ and a closed small neighborhood $W(\mathbf{p})$ of $\mathbf{p}$ such that if $\mathbf{q} \in W(\mathbf{p})$ there exists $\vec{w}^{\mathbf{q}}_{\mathbf{p}}$ uniquely with the following properties:

(1) $\mathbf{q} \cup \vec{w}^{\mathbf{q}}_{\mathbf{p}}$ is $\epsilon_{\mathbf{p}}$-close to $\mathbf{p} \cup w_{\mathbf{p}} \cup \vec{w}_{\rm can}$.
(2) $\mathbf{q} \cup \vec{w}^{\mathbf{q}}_{\mathbf{p}}$ satisfies the transversal constraint.
(3) The linearization operator $D_{\mathbf{q}}\overline{\partial}_{J,H^{21}}$ at $\mathbf{q}$ as in (4.7) is surjective mod $\oplus_{\rm v}$ Im $I_{\mathbf{p},{\rm v};\mathbf{q}}$, where $I_{\mathbf{p},{\rm v};\mathbf{q}}$ is the map in (6.8) for $\mathbf{q} = (\Sigma', u', \varphi')$.

The proof of this fact is the same as that of Lemma 4.22. Then we have

$$\bigcup_{\mathbf{p}} \text{Int } W(\mathbf{p}) = \mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+).$$

Therefore, by compactness of the moduli space we can take a finite subset indexed by a finite set $\mathfrak{C}_\ell(H^{21}, \mathcal{J}; \alpha_-, \alpha_+)$

$$\mathcal{A}_\ell(H^{21}, \mathcal{J}; \alpha_-, \alpha_+) = \{\mathbf{p}_c \mid c \in \mathfrak{C}_\ell(H^{21}, \mathcal{J}; \alpha_-, \alpha_+)\} \subset \mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$$

such that for each $c \in \mathfrak{C}_\ell(H^{21}, \mathcal{J}; \alpha_-, \alpha_+)$ we take obstruction bundle data $\mathfrak{E}_{\mathbf{p}_c}$ centered at $\mathbf{p}_c$, and a closed neighborhood $W(\mathbf{p}_c)$ of $\mathbf{p}_c$ in $\mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$ with the following property. For each element $\mathbf{q} \in W(\mathbf{p}_c)$ there exists $\vec{w}^{\mathbf{q}}_{\mathbf{p}_c}$ such that

(1) $\mathbf{q} \cup \vec{w}^{\mathbf{q}}_{\mathbf{p}_c}$ is ϵ_c-close to $\mathbf{p}_c \cup w_{\mathbf{p}_c} \cup \vec{w}_{\rm can}$.

(2) $\mathbf{q} \cup \vec{w}^{\mathbf{q}}_{\mathbf{p}_c}$ satisfies the transversal constraint.
(3)
$$\bigcup_{c \in \mathcal{C}_\ell(H^{21},\mathcal{J};\alpha_-,\alpha_+)} \mathrm{Int}\, W(\mathbf{p}_c) = \mathcal{N}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+). \tag{6.9}$$

Definition 6.16. (1) For each $\mathbf{q} \in \mathcal{N}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)$ we put
$$\mathcal{E}(\mathbf{q}) = \{c \in \mathcal{C}_\ell(H^{21},\mathcal{J};\alpha_-,\alpha_+) \mid \mathbf{q} \in W(\mathbf{p}_c)\}.$$

(2) Let $\mathcal{B} \subset \mathcal{E}(\mathbf{q})$ be a nonempty subset.
(3) We consider $(\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c, u', \varphi', a'_0)$ such that for each c, $(\mathfrak{Y} \cup \vec{w}'_c, u', \varphi', a'_0)$ is ϵ-close to $\mathbf{q} \cup \vec{w}^{\mathbf{q}}_c$. If $\epsilon > 0$ is small, then $(\mathfrak{Y} \cup \vec{w}'_c, u', \varphi', a'_0)$ is ϵ-close to $\mathbf{p}_c \cup \vec{w}_c$ and we can define the map
$$I_{\mathbf{p}_c,\mathrm{v};\Sigma',u',\varphi',a'_0} : E_{\mathbf{p}_c,\mathrm{v}}(\mathfrak{y}_{\mathbf{p}_c}) \to C^\infty(\Sigma';(u')^*TX \otimes \Lambda^{0,1})$$
as in (6.8) for each irreducible component v of $\mathbf{p}_c$. Here $(\mathfrak{Y} \cup \vec{w}'_c, \varphi', a'_0) = \Phi_{\mathbf{p}_c}(\mathfrak{y}_{\mathbf{p}_c}, \vec{T}_{\mathbf{p}_c}, \vec{\theta}_{\mathbf{p}_c})$ as in Notation 4.15. We now put
$$E((\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c, u', \varphi', a'_0);\mathbf{q};\mathcal{B}) = \bigoplus_{c\in\mathcal{B}} \bigoplus_{\mathrm{v}} \mathrm{Im}\, I_{\mathbf{p}_c,\mathrm{v};\Sigma',u',\varphi',a'_0}. \tag{6.10}$$

We can define the notion of the *stabilization data* centered at $\mathbf{q} \in \mathcal{N}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)$ in the same way as in Definition 4.26.

Using those data we fixed, we will define a Kuranishi chart of $\mathbf{q}$ as follows.

Definition 6.17. We consider the following conditions on an object $(\mathfrak{Y} \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, u', \varphi', a'_0)$:

(1) If Σ'_a is the mainstream component and φ'_a is a parametrization of this mainstream component, the following equation is satisfied on $\mathbb{R} \times S^1$.
$$\begin{aligned}&\frac{\partial(u'\circ\varphi'_a)}{\partial\tau} + J_1\left(\frac{\partial(u'\circ\varphi'_a)}{\partial t} - \mathfrak{X}_{H^1_t}\circ u'\circ\varphi'_a\right)\\ &\equiv 0 \mod E((\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c, u', \varphi', a'_0);\mathbf{q};\mathcal{B})\end{aligned} \tag{6.11}$$
for $a < a'_0$,
$$\begin{aligned}&\frac{\partial(u'\circ\varphi'_a)}{\partial\tau} + J_2\left(\frac{\partial(u'\circ\varphi'_a)}{\partial t} - \mathfrak{X}_{H^2_t}\circ u'\circ\varphi'_a\right)\\ &\equiv 0 \mod E((\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c, u', \varphi', a'_0);\mathbf{q};\mathcal{B})\end{aligned} \tag{6.12}$$
for $a > a'_0$,
$$\begin{aligned}&\frac{\partial(u'\circ\varphi'_a)}{\partial\tau} + J^{21}_\tau\left(\frac{\partial(u'\circ\varphi'_a)}{\partial t} - \mathfrak{X}_{H^{21}_{(\tau,t)}}\circ u'\circ\varphi'_a\right)\\ &\equiv 0 \mod E((\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c, u', \varphi', a'_0);\mathbf{q};\mathcal{B})\end{aligned} \tag{6.13}$$
for $a = a'_0$.

(2) If $\Sigma'_{\rm v}$ is a bubble component, the following equation is satisfied on $\Sigma'_{\rm v}$.

$$\overline{\partial}_J u' \equiv 0 \mod E((\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c, u', \varphi', a'_0); \mathbf{q}; \mathcal{B}). \tag{6.14}$$

Here the almost complex structure J is as follows. Let $\widehat{\Sigma}'_a$ be the extended mainstream component containing $\Sigma'_{\rm v}$. If $a < a'_0$, then $J = J_1$. If $a > a'_0$, then $J = J_2$. If $a = a'_0$ and $\varphi_{a'_0}(\tau, t)$ is the root of the tree of sphere components containing $\Sigma'_{\rm v}$, then $J = J^{21}_\tau$.

(3) For each $c \in \mathcal{E}(\mathbf{q})$ the additional marked points $\vec{w}'_c$ satisfy the transversal constraint with respect to $\mathbf{p}_c$.

(4) The additional marked points $\vec{w}'_{\mathbf{q}}$ satisfy the transversal constraint with respect to $\mathbf{q}$.

(5) $(\mathfrak{Y} \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, u', \varphi', a'_0)$ is ϵ_1-close to $\mathbf{q} \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}^{\mathbf{q}}_c \cup \vec{w}_{\mathbf{q}}$.

The set of isomorphism classes of $(\mathfrak{Y} \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, u', \varphi', a'_0)$ satisfying the conditions (1)–(5) above is denoted by

$$V(\mathbf{q}, \epsilon_1, \mathcal{B}),$$

where $(\mathfrak{Y}' \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, u', \varphi', a'_0)$ is said to be *isomorphic* to $(\mathfrak{Y}'' \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}''_c \cup \vec{w}''_{\mathbf{q}}, u'', \varphi'', a''_0)$ if there exists a biholomorphic map $v : \Sigma' \to \Sigma''$ such that

(a) $u'' = u' \circ v$ holds outside the set of the transit points.

(b) If Σ'_a is a mainstream component of Σ' and $v(\Sigma'_a) = \Sigma''_{a'}$, then we have $(v \circ \varphi'_a)(\tau + \tau_a, t) = \varphi''_{a'}(\tau, t)$ for some τ_a.

(c) We assume that if $a = a'_0$ and $v(\Sigma'_a) = \Sigma''_{a'}$, then $a' = a''_0$. Moreover, $\tau_{a'_0} = 0$.

(d) $v(z'_i) = z''_i$ and $v(w'_i) = w''_i$.

In the same way as in Lemma 4.29, we can prove that $V(\mathbf{q}, \epsilon_1, \mathcal{B})$ is a smooth manifold with boundary and corner if $\epsilon_1 > 0$ and $\epsilon_c > 0$ are small enough.

We note that the group $\mathrm{Aut}^+(\mathbf{q})$ acts on $V(\mathbf{q}, \epsilon_1, \mathcal{B})$ since the stabilization data are assumed to be preserved by it. In particular, $\mathrm{Aut}(\mathbf{q})$ acts on it. By the condition in Definition 4.11 (6) this action is effective. Therefore, the quotient space $V(\mathbf{q}, \epsilon_1, \mathcal{B})/\mathrm{Aut}(\mathbf{q})$ is an effective orbifold, which we denote by

$$U(\mathbf{q}, \epsilon_1, \mathcal{B}).$$

We define a vector bundle on $U(\mathbf{q}, \epsilon_1, \mathcal{B})$ such that its fiber at $(\mathfrak{Y}, u', \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, \varphi', a'_0)$ is $E((\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c, u', \varphi', a'_0); \mathbf{q}; \mathcal{B})$. We denote this vector bundle by

$$E(\mathbf{q}, \epsilon_1, \mathcal{B}).$$

We can define its section $s_{(\mathbf{q},\epsilon_1,\mathcal{B})}$ by using the left hand side of (6.11)–(6.14). An element of its zero set represents an element of $\mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$. Thus we obtain:

$$\psi_{(\mathbf{q},\epsilon_1,\mathcal{B})} : s^{-1}_{(\mathbf{q},\epsilon_1,\mathcal{B})}(0) \to \mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+).$$

We thus obtain a Kuranishi chart

$$(U(\mathbf{q},\epsilon_1,\mathcal{B}),E(\mathbf{q},\epsilon_1,\mathcal{B}),s_{(\mathbf{q},\epsilon_1,\mathcal{B})},\psi_{(\mathbf{q},\epsilon_1,\mathcal{B})}).$$

In the way similar to the proof of Lemmas 4.33, 4.34, and also using the exponential decay estimates in the same way as in [16], we can define coordinate change among them. We thus obtain a Kuranishi structure on $\mathcal{N}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)$. We have proved Theorem 6.4 (1)(2).

6.3. Proof of Theorem 6.4 (3)(4): Kuranishi structure with outer collar

The strategy of the proof of Theorem 6.4 (3)(4) is similar to that in Sect. 5. Namely we first take the outer collaring $\mathcal{N}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)^{\boxplus 1}$ as in Sect. 5 and modify them on $S_k(\mathcal{N}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+))\times[-1,0]^k$. (Note that the union of $S_k(\mathcal{N}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+))\times[-1,0]^k$ for various k is $\mathcal{N}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+)^{\boxplus 1}$.) The details are as follows.

Let $\mathfrak{A}_r$ be the index set of the critical submanifolds of H^r $(r=1,2)$. Let

$$\alpha_-=\alpha_{1,0},\alpha_{1,1},\ldots,\alpha_{1,m_1-1},\alpha_{1,m_1}\in\mathfrak{A}_1,$$
$$\alpha_{2,1},\ldots,\alpha_{2,m_2},\alpha_{2,m_2+1}=\alpha_+\in\mathfrak{A}_2.$$

We consider the fiber product

$$\begin{aligned}&\mathcal{M}_{\ell_{1,1}}(X,J_1,H^1;\alpha_{1,0},\alpha_{1,1})\,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\cdots\\&\quad\cdots{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\mathcal{M}_{\ell_{1,m_1}}(X,J_1,H^1;\alpha_{1,m_1-1},\alpha_{1,m_1})\\&\quad{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\mathcal{N}_{\ell'}(X,\mathcal{J},H^{21};\alpha_{1,m_1},\alpha_{2,1})\\&\quad{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\mathcal{M}_{\ell_{2,2}}(X,J_2,H^2;\alpha_{2,1},\alpha_{2,2})\,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\cdots\\&\quad\cdots{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\mathcal{M}_{\ell_{2,m_2+1}}(X,J_2,H^2;\alpha_{2,m_2},\alpha_{2,m_2+1})\end{aligned}\tag{6.15}$$

which we denoted by $\mathcal{N}_{\vec\ell_1,\ell',\vec\ell_2}(X,\mathcal{J},H^{21};\vec\alpha_1,\vec\alpha_2)$. We observe that

$$\begin{aligned}&S_m(\mathcal{N}_\ell(X,\mathcal{J},H^{21};\alpha_-,\alpha_+))\\&\quad=\bigcup_{\substack{\vec\ell_1,\ell',\vec\ell_2\\|\vec\ell_1|+\ell'+|\vec\ell_2|=\ell}}\ \bigcup_{\substack{\vec\alpha_1,\vec\alpha_2\\m_1+m_2=m,\ \alpha_{1,0}=\alpha_-,\alpha_{2,m_2+1}=\alpha_+}}\mathcal{N}_{\vec\ell_1,\ell',\vec\ell_2}(X,\mathcal{J},H^{21};\vec\alpha_1,\vec\alpha_2).\end{aligned}$$

Note the sum is taken for $m_1=\#\vec\alpha_1-1\ge 0$ and $m_2=\#\vec\alpha_2-1\ge 0$. We will construct a Kuranishi structure for each of

$$\mathcal{N}_{\vec\ell_1,\ell',\vec\ell_2}(X,\mathcal{J},H^{21};\vec\alpha_1,\vec\alpha_2)^+=\mathcal{N}_{\vec\ell_1,\ell',\vec\ell_2}(X,\mathcal{J},H^{21};\vec\alpha_1,\vec\alpha_2)\times[-1,0]^{m_1}\times[-1,0]^{m_2}.$$

Let $A_r\sqcup B_r\sqcup C_r=\underline{m_r}$ for $r=1,2$. It induces $\mathcal{I}_{A_r,B_r,C_r}:[-1,0]^{b_r}\to[-1,0]^{m_r}$ by (5.2).

We will formulate the compatibility condition below (Condition 6.20), which describes the restriction of the Kuranishi structure $\widehat{\mathcal{U}}_{\vec\ell_1,\ell',\vec\ell_2}(X,\mathcal{J},H^{21};\vec\alpha_1,\vec\alpha_2)$ of the product space $\mathcal{N}_{\vec\ell_1,\ell',\vec\ell_2}(X,\mathcal{J},H^{21};\vec\alpha_1,\vec\alpha_2)^+$ to the image of the embedding:

$$\begin{aligned}&\mathrm{id}\times\mathcal{I}_{A_1,B_1,C_1}\times\mathcal{I}_{A_2,B_2,C_2}:\mathcal{N}_{\vec\ell_1,\ell',\vec\ell_2}(X,\mathcal{J},H^{21};\vec\alpha_1,\vec\alpha_2)\times[-1,0]^{b_1+b_2}\\&\quad\to\mathcal{N}_{\vec\ell_1,\ell',\vec\ell_2}(X,\mathcal{J},H^{21};\vec\alpha_1,\vec\alpha_2)^+.\end{aligned}\tag{6.16}$$

We put $A_r = \{i(A_r,1),\dots,i(A_r,a_r)\}$ with $i(A_r,1) < i(A_r,2) < \cdots < i(A_r,a_r-1) < i(A_r,a_r)$ and consider the fiber product

$$\begin{aligned}
&\mathcal{M}_{\vec\ell_{1,A_1,1}}(X,J_1,H^1;\alpha_{1,0},\dots,\alpha_{1,i(A_1,1)})\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec\ell_{1,A_1,2}}(X,J_1,H^1;\alpha_{1,i(A_1,1)},\dots,\alpha_{i(A_1,2)})\,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\cdots\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec\ell_{1,A_1,j+1}}(X,J_1,H^1;\alpha_{i(A_1,j)},\dots,\alpha_{i(A_1,j+1)})\,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\cdots\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec\ell_{1,A_1,a_1}}(X,J_1,H^1;\alpha_{1,i(A_1,a_1-1)},\dots,\alpha_{1,i(A_1,a_1)})\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{N}_{\vec\ell_{1,A_1,a_1+1},\ell',\vec\ell_{2,A_2,1}}(X,\mathcal{J},H^{21};\\
&\quad (\alpha_{1,i(A_1,a_1)},\dots,\alpha_{1,m_1}),(\alpha_{2,1},\dots,\alpha_{2,i(A_2,1)}))\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec\ell_{2,A_2,2}}(X,J_2,H^2;\alpha_{2,i(A_2,1)},\dots,\alpha_{i(A_2,2)})\,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\cdots\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec\ell_{2,A_2,j+1}}(X,J_2,H^2;\alpha_{i(A_2,j)},\dots,\alpha_{i(A_2,j+1)})\,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-}\cdots\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec\ell_{2,A_2,a_2+1}}(X,J_2,H^2;\alpha_{2,i(A_2,a_2)},\dots,\alpha_{2,m_2+1}).
\end{aligned}\tag{6.17}$$

Here $\vec\ell_{r,A_r,j} = (\ell_{r,i(A_r,j-1)+1},\dots,\ell_{r,i(A_r,j)})$.

Remark 6.18. Here and hereafter $i(A_1,0)=0$ and $i(A_2,a_2+1)=m_2+1$ by convention.

The fiber product (6.17) is nothing but $\mathcal{M}_{\vec\ell_1,\ell',\vec\ell_2}(X,\mathcal{J},H^{21};\vec\alpha_1,\vec\alpha_2)$. Therefore, we can use (6.17) to define a fiber product Kuranishi structure on the space $\mathcal{N}_{\vec\ell_1,\ell',\vec\ell_2}(X,\mathcal{J},H^{21};\vec\alpha_1,\vec\alpha_2)$.

We need more notation. Although the notation is rather heavy, its geometric meaning is simple. Namely we will consider the moduli space $\mathcal{M}_{\vec\ell_{r,A_r,j,C_r}}(X,J_r,H^r;\vec\alpha_{r,A_r,j,C_r})$ etc., which is obtained by allowing to smooth the singularities at the transit points corresponding to the indices belonging to C_r.

Notation 6.19. Let $j=1,\dots,a_r+1$ $(r=1,2)$.

(1) We decompose C_r into $C'_j(A_r) = [i(A_r,j-1),i(A_r,j)]_{\mathbb{Z}} \cap C_r$ and put $C_j(A_r) = \{i - i(A_r,j-1) \mid i \in C'_j(A_r)\}$, $c_j(r,A_r) = \# C_j(A_r)$. Here for $r=1$
$$0 = i(A_1,0) < i(A_1,1) < \cdots < i(A_1,a_1) \le m_1$$
with $i(A_1,0)=0, i(A_1,a_1+1)=m_1$ as convention. For $r=2$,
$$i(A_2,0) = 1 \le i(A_2,1) < \cdots < i(A_2,a_2) < m_2+1$$
with $i(A_2,0)=1, i(A_2,a_2+1)=m_2+1$ as convention.

(2) We put $\vec\alpha_{r,A_r,j} = (\alpha_{r,i(A_r,j-1)},\dots,\alpha_{r,i(A_r,j)})$. Note that[29]
$$\begin{aligned}
&\mathcal{M}_{\vec\ell_{A_r,j}}(X,J_r,H^r;\alpha_{r,i(A_r,j-1)},\dots,\alpha_{r,i(A_r,j)})\\
&\quad = \mathcal{M}_{\vec\ell_{A_r,j}}(X,J_r,H^r;\vec\alpha_{r,A_r,j}).
\end{aligned}\tag{6.18}$$

(3) We remove $\{\alpha_{r,i} \mid i \in C'_j(A_r)\}$ from $\vec\alpha_{r,A_r,j}$ to obtain $\vec\alpha_{r,A_r,j,C_r}$.

[29] In case $i(A_1,a_1) = i(A_1,a_1+1)$ or $i(A_2,0) = i(A_2,1)$, we do not consider $\mathcal{M}_{\vec\ell_{r,j}}(X,J_r,H^r;\vec\alpha_{r,A_r,j})$.

(4) Noticing $i(A_1, a_1 + 1) = m_1$ for $r = 1$, we apply (2)(3) to obtain $\vec{\alpha}_{1,A_1,a_1+1}$, $\vec{\alpha}_{1,A_1,a_1+1,C_1}$. Also noticing $i(A_2, 0) = 1$ for $r = 2$, we apply (2)(3) to obtain $\vec{\alpha}_{2,A_2,1}$, $\vec{\alpha}_{2,A_2,1,C_2}$.

Note that $\vec{\alpha}_{A_1,0}$, $\vec{\alpha}_{1,A_1,0,C_1}$ and $\vec{\alpha}_{2,A_2,a_2}$, $\vec{\alpha}_{2,A_2,a_2,C_2}$ are defined according to Remark 6.18.

(5) We put $m_j(r, A_r) = i(A_r, j) - i(A_r, j - 1)$,

$$m_j(r, A_r, C_r) = \#(B_r \cap (i(A_r, j - 1), i(A_r, j))_{\mathbb{Z}}) + 1.$$

Therefore,

$$\sum_{j=1}^{a_r+1} (m_j(r, A_r, C_r) - 1) = \#B_r = b_r. \tag{6.19}$$

(6) We define $\vec{\ell}_{r,A_r,j,C_r}$ as follows. Let

$$\vec{\alpha}_{r,A_r,j,C_r} = \{\alpha_{r,i(A_r,j-1)+k_s} \mid s = 0, \ldots, m_j(r, A_r, C_r)\}.$$

Here $0 \le k_0 < k_1 < \cdots < k_{m_j(r,A_r,C_r)}$.[30]

Note if $i \in (i(A_r, j-1)+k_s, i(A_r, j-1)+k_{s+1})_{\mathbb{Z}}$, then $i \in C'_j(A_r)$. We put

$$\ell_{r,A_r,j,C_r,s} = \ell_{r,i(A_r,j-1)+k_{s-1}+1} + \cdots + \ell_{r,i(A_r,j-1)+k_s}$$

and

$$\vec{\ell}_{r,A_r,j,C_r} = (\ell_{r,A_r,j,C_r,1}, \ldots, \ell_{r,A_r,j,C_r,m_j(r,A_r,C_r)})$$

Finally we put

$$\begin{aligned} \ell'' = \ell_{1,i(A_1,a_1)+k_{m_{a_1}(1,A_1,C_1)}+1} + \cdots + \ell_{1,m_1} + \\ + \ell' + \ell_{2,2} + \cdots + \ell_{2,\min(i(A_2,1),i(B_2,1))}. \end{aligned} \tag{6.20}$$

See Fig. 9.

It is easy to check

$$\sum_{r,s,j} \ell_{r,A_r,j,C_r,s} + \ell'' = \sum_{r,i} \ell_{r,i} + \ell'. \tag{6.21}$$

We note that in Proposition 5.5, we determined a Kuranishi structure on

$$\begin{aligned} &\mathcal{M}_{\vec{\ell}_{r,A_r,j,C_r}}(X, J_r, H^r; \vec{\alpha}_{r,A_r,j,C_r})^+ \\ &= \mathcal{M}_{\vec{\ell}_{r,A_r,j,C_r}}(X, J_r, H^r; \vec{\alpha}_{r,A_r,j,C_r}) \times [-1, 0]^{m_j(r,A_r,C_r)-1}. \end{aligned} \tag{6.22}$$

We denote it by $\widehat{\mathcal{U}}_{\vec{\ell}_{r,A_r,j,C_r}}(X, J_r, H^r; \vec{\alpha}_{r,A_r,j,C_r})$. By construction we have

$$\mathcal{M}_{\vec{\ell}_{r,A_r,j}}(X, J_r, H^r; \vec{\alpha}_{r,A_r,j}) \subseteq \widehat{S}_{c_j(r,A_r)}(\mathcal{M}_{\vec{\ell}_{r,A_r,j,C_r}}(X, J_r, H^r; \vec{\alpha}_{r,A_r,j,C_r})).$$

By restriction, $\widehat{\mathcal{U}}_{\vec{\ell}_{r,A_r,j,C_r}}(X, J_r, H^r; \vec{\alpha}_{r,A_r,j,C_r})$ determines a Kuranishi structure of (6.17) times $[-1, 0]^*$ except one of the factors

$$\mathcal{N}_{\vec{\ell}_{1,A_1,a_1+1},\ell',\vec{\ell}_{2,A_2,1}}(X, \mathcal{J}, H^{21}; \vec{\alpha}_{1,A_1,a_1+1}, \vec{\alpha}_{2,A_2,1}). \tag{6.23}$$

[30] Note that k_s depends on r, A_r, B_r, C_r. We also note that $k_0 = 0$ holds unless $r = 2, j = 1$ and $i(A_2, 1) = 1$. Also $k_{m_j(r,A_r,C_r)} = i(A_r, j) - i(A_r, j - 1)$ holds unless $r = 1, j = a_1 + 1$ and $i(A_1, a_1) = m_1$.

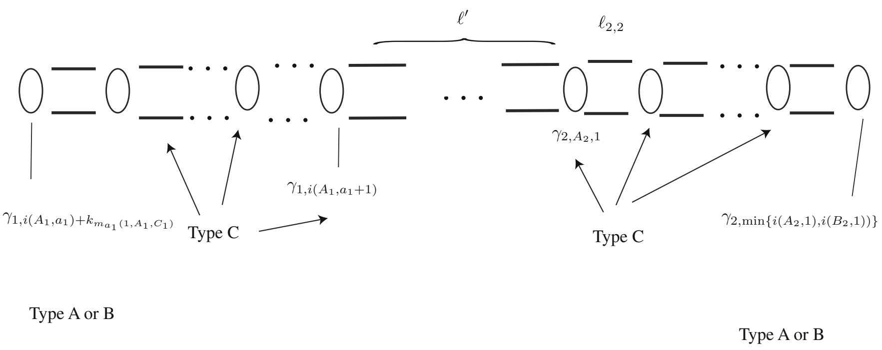


FIGURE 9. The case $m_1 \in C_1$ and $1 \in C_2$

By construction, we can easily show that (6.23) is a component of

$$\widehat{S}_{c_{a_1+1}(1,A_1)+c_1(2,A_2)}(\mathcal{N}_{\vec{\ell}_{1,A_1,a_1+1,C_1},\ell'',\vec{\ell}_{2,A_2,1,C_2}}(X,\mathcal{J},H^{21};\vec{\alpha}_{1,A_1,a_1+1,C_1},\vec{\alpha}_{2,A_2,1,C_2})).$$

(See Notation 6.19 (1) for the notations.) Note the sum of the exponent $[-1,0]$ of appearing in (6.22) plus $m_{a_1+1}(1,A_1,C_1)-1+m_1(2,A_2,C_2)-1$ is b_1+b_2. This is a consequence of (6.19).

Now the compatibility condition we require is described as follows.

Condition 6.20. We require the K-system

$$\{(\mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J},H^{21};\vec{\alpha}_1,\vec{\alpha}_2)^+,\ \widehat{\mathcal{U}}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J},H^{21};\vec{\alpha}_1,\vec{\alpha}_2)^+)\}$$

satisfies the following.

Its restriction to the image of the embedding (6.16) is the fiber product of the following factors. (Here we use the fiber product description (6.17).)

(1) The restriction of the Kuranishi structure $\widehat{\mathcal{U}}_{\vec{\ell}_{r,A_r,j},C_r}(X,J_r,H^r;\vec{\alpha}_{r,A_r,j,C_r})$ to $\mathcal{M}_{\vec{\ell}_{r,A_r,j}}(X,J_r,H^r;\vec{\alpha}_{r,A_r,j})\times[-1,0]^{m_j(r,A_r,C_r)-1}$.

(2) The restriction of the Kuranishi structure

$$\widehat{\mathcal{U}}_{\vec{\ell}_{1,A_1,a_1+1,C_1},\ell'',\vec{\ell}_{2,A_2,1,C_2}}(X,\mathcal{J},H^{21};\vec{\alpha}_{1,A_1,a_1+1,C_1},\vec{\alpha}_{2,A_2,1,C_2})$$

to (6.23) $\times[-1,0]^{m_{a_1+1}(1,A_1,C_1)-1}\times[-1,0]^{m_1(2,A_2,C_2)-1}$.

Proposition 6.21. *There exists a K-system*

$$\{(\mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J},H^{21};\vec{\alpha}_1,\vec{\alpha}_2)^+,\ \widehat{\mathcal{U}}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J},H^{21};\vec{\alpha}_1,\vec{\alpha}_2))\}$$

for various $\vec{\ell}_1,\ell',\vec{\ell}_2,\vec{\alpha}_1,\vec{\alpha}_2$ with the following properties.

(1) *They satisfy Condition 6.20.*

(2) *Let $\mathfrak{C}$ be the union of the components of $\mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J},H^{21};\vec{\alpha}_1,\vec{\alpha}_2)^+$ which are in $\mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J},H^{21};\vec{\alpha}_1,\vec{\alpha}_2)\times\partial([-1,0]^{m-1})$. Then the Kuranishi structure $\widehat{\mathcal{U}}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J},H^{21};\vec{\alpha}_1,\vec{\alpha}_2)$ is $\mathfrak{C}$-collared in the sense of Remark 5.6 (by replacing $\mathcal{M}_{\vec{\ell}}(X,H;\vec{\alpha})^+$ by $\mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J},H^{21};\vec{\alpha}_1,\vec{\alpha}_2)^+$).*

(3) *For the case* $\vec{\alpha}_1 = (\alpha_-)$ *and* $\vec{\alpha}_2 = (\alpha_+)$, $\widehat{\mathcal{U}}_{\emptyset,\ell,\emptyset}(X, \mathcal{J}, H^{21}; \vec{\alpha}_1, \vec{\alpha}_2)$ *coincides with the Kuranishi structure we constructed during the proof of Theorem* 6.4*s (1)(2).*

Proof. The proof is entirely the same as the proof of Proposition 5.5. □

Now we replace the Kuranishi structures of $S_m(\mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)) \times [-1, 0]^m$ with ones in Proposition 6.21 to obtain the Kuranishi structures in Theorem 6.4 (3)(4). The proof of Theorem 6.4 is complete. □

Remark 6.22. We may choose H^{21} so that

$$H^{21}(x, \tau, t) = (1 - \chi(\tau))H^1(x, t) + \chi(\tau)H^2(x, t),$$

where $\tau : \mathbb{R} \to [0, 1]$ is an increasing function which is 0 when $\tau < -1$ and is 1 when $\tau > 1$. In this case the energy loss of the morphism obtained in Theorem 6.4 is estimated from above by

$$\int_{t \in S^1} \sup_{x \in X} |H^1(x, t) - H^2(x, t)| \mathrm{d}t.$$

This is a well established result. See for example, [6, Section 2, 3°] [35, Lemma 4.1], [22, Lemma 9.3] for a proof of this inequality.

7. Construction of homotopy

Situation 7.1. (1) Let H^1, H^2 be two Hamiltonians $X \times S^1 \to \mathbb{R}$ which are Morse–Bott non-degenerate in the sense of Condition 2.1. We consider a family of Hamiltonians parametrized by an interval $[0, 1]$.

(2) Suppose $H^{21,[0,1]} : X \times \mathbb{R} \times S^1 \times [0, 1] \to \mathbb{R}$ is a smooth function and $\mathcal{J}^{[0,1]} = \{J_{\tau,s} \mid \tau \in \mathbb{R}, s \in [0, 1]\}$ is an $\mathbb{R} \times [0, 1]$ parametrized smooth family of tame almost complex structures on X.

(3) For each $s \in [0, 1]$ we assume the pair $(H^{21,s}, \mathcal{J}^s (= \mathcal{J}^{21,s}))$ is as in Situation 6.1.

(4) We assume that the families $H^{21,[0,1]}$ and $\mathcal{J}^{[0,1]}$ are collared in the following sense: We consider the retraction $\mathcal{R} : [0, 1] \to [\tau, 1 - \tau]$ such that $\mathcal{R}(s) = \tau$ if $s \in [0, \tau]$, $\mathcal{R}(s) = 1 - \tau$ if $s \in [1 - \tau, 1]$ and $\mathcal{R}(s) = s$ otherwise. Then $H^{21,s} = H^{21,\mathcal{R}(s)}$ and $\mathcal{J}^s = \mathcal{J}^{\mathcal{R}(s)}$.

□

We note that such families always exist. More precisely, we have the following.

Lemma 7.2. *Suppose we are given* H^{21,s_0} *and* $\mathcal{J}^{s_0}$ *for* $s_0 = 0, 1$. *Then there exist* $H^{21,[0,1]}$ *and* $\mathcal{J}^{[0,1]}$ *as in Situation* 7.1*, whose restrictions to* s_0 *coincide with* H^{21,s_0} *and* $\mathcal{J}^{s_0}$ *for* $s_0 = 0, 1$*, respectively.*

Proof. This is an immediate consequence of the facts that the set of tame almost complex structures is contractible and the set of smooth functions is contractible. □

Definition 7.3. Let $R_{\alpha_-} \in \widetilde{\mathrm{Per}}(H^1)$ and $R_{\alpha_+} \in \widetilde{\mathrm{Per}}(H^2)$. We put

$$\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+) = \bigcup_{s\in[0,1]} \mathcal{N}_\ell(X, \mathcal{J}^s, H^{21,s}; \alpha_-, \alpha_+) \times \{s\}.$$

Here $\mathcal{N}_\ell(X, \mathcal{J}^s, H^{21,s}; \alpha_-, \alpha_+)$ is defined in Definition 6.10.

We can define a topology of $\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+)$ in the way similar to one in Definition 3.17 and show that it is Hausdorff and compact.

We consider the boundary components of $\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+)$ consisting of the disjoint union

$$\left(\mathcal{N}_\ell(X, \mathcal{J}^0, H^{21,0}; \alpha_-, \alpha_+) \times \{0\}\right) \sqcup \left(\mathcal{N}_\ell(X, \mathcal{J}^1, H^{21,1}; \alpha_-, \alpha_+) \times \{1\}\right).$$

We call it the *vertical boundary* and denote it by $\mathfrak{C}^v$. We consider the projection $\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+) \to [0,1]$. In the definition of $\mathfrak{C}$-collared-ness (see Remark 5.6) we replace $[-1,0]^{m-1}$ by $[0,1]$ to define $\mathfrak{C}^v$-collared-ness. Then $\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+) \to [0,1]$ is $\mathfrak{C}^v$-collared, by Situation 7.1 (4). (This is the collared-ness in the $[0,1]$ direction.)

The complement of the vertical boundary is written as $\mathfrak{C}^h$ and we call it the *horizontal boundary*. We denote by $\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+)^{\mathfrak{C}^h \boxplus 1}$ the space

$$\bigcup_{s\in[0,1]} \mathcal{N}_\ell(X, \mathcal{J}^s, H^{21,s}; \alpha_-, \alpha_+)^{\boxplus 1} \times \{s\}.$$

This space coincides with the 'partial outer collaring' of $\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+)$ in the horizontal direction, which is introduced in [23, Chapter 19]. (However, we do not use [23, Chapter 19] in this article. The symbol $\mathfrak{C}^h \boxplus 1$ here can be regarded just as a notation.)

Theorem 7.4. *We can define a Kuranishi structure on*

$$\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+)^{\mathfrak{C}^h \boxplus 1}$$

so that it will be an interpolation spaces of a $[0,1]$*-parametrized family of morphisms between linear K-systems associated to* H^1 *and* H^2 *obtained by Theorem* 2.9 *in the sense of* [23, Condition 16.21].

We need certain additional properties on our morphism for applications. We will state them below.

Proposition 7.5. *We may choose our Kuranishi structure on*

$$\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+)^{\mathfrak{C}^h \boxplus 1}$$

so that it is $\mathfrak{C}^v$*-collared.*

Proof of Theorem 7.4 and Proposition 7.5. We are given morphisms associated to $s_0 = 0, 1$. We made various choices in Sect. 6. Below we will show that we can define a $[0,1]$ parametrized morphism so that its boundary becomes the union of two morphisms associated to $s_0 = 0, 1$.

Again the proof is by two steps. In the first step we construct a Kuranishi structure on $\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+)$ which may not be compatible with the fiber product description at the horizontal boundary. (We construct our Kuranishi structure on $\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+)$ so that it is compatible with the given Kuranishi structure at the vertical boundary $\{0,1\} = \partial[0,1]$.) We then modify it on the collar so that it becomes a $[0,1]$-parametrized interpolation space. The details are as follows.

Remark 7.6. Since the proof is a repetition of the construction in the previous sections, the readers may skip it and go directly to Sect. 8. We provide the details of the proof here for the sake of completeness.

We first define the notion of ϵ-closeness. Let

$$\widehat{\mathbf{q}} = (\mathbf{q}, s_{\mathbf{q}}) \in \mathcal{N}_\ell(X, \mathcal{J}^s, H^{21,s}; \alpha_0, \alpha_+) \times \{s_{\mathbf{q}}\}$$

and we take stabilization data at $\mathbf{q}$. Hereafter we omit the symbol $\widehat{\ }$, and write

$$\mathbf{q} = (\mathbf{q}, s_{\mathbf{q}})$$

by an abuse of notation. Then we consider

$$(\mathfrak{Y} \cup \vec{w}', \varphi', a_0) = \Phi_{\mathbf{q}}(\mathfrak{y}, \vec{T}, \vec{\theta}) \in \mathcal{N}_{\ell+\ell'+\ell''}(\text{source})$$

as in (6.7). We consider $(\mathfrak{Y} \cup \vec{w}', u', \varphi', a_0', s')$ where u' is a map from the curve $\mathfrak{Y}$ to X and $s' \in [0,1]$. We say that it is ϵ-*close* to $\mathbf{q} \cup \vec{w} \cup \vec{w}_{\rm can}$ if Definition 4.16 (1)–(4) hold and $|s_{\mathbf{q}} - s'| < \epsilon$. In (2) we use the Hamiltonian $H^{21,s'}$ and the family of almost complex structures $\mathcal{J}_{s'}$ as in Definition 6.6 (5). In (3) we also use $H^{21,s'}$ to define the redefined connecting orbit map.

Next, for a point $\mathbf{p} \in \mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$ we suppose to be given obstruction bundle data $\mathfrak{E}_{\mathbf{p}}$ and $(\mathfrak{Y} \cup \vec{w}', u', \varphi', a_0', s')$ is $\epsilon_{\mathbf{p}}$-close to $\mathbf{p} \cup \vec{w} \cup \vec{w}_{\rm can}$ for a sufficiently small $\epsilon_{\mathbf{p}} > 0$. Then we can define a complex linear map

$$I_{\mathbf{p},\mathrm{v};\Sigma',u',\varphi',a_0'} : E_{\mathbf{p},\mathrm{v}}(\mathfrak{y}_{\mathbf{p}}) \to C^\infty(\Sigma'; (u')^*TX \otimes \Lambda^{0,1}) \tag{7.1}$$

in the same way as in the parametrized version of Definition 4.21.

We recall from the paragraph around (6.9) that we took a finite set indexed by the set $\mathcal{C}_\ell(H^{21}, \mathcal{J}; \alpha_-, \alpha_+)$

$$\mathcal{A}_\ell(H^{21}, \mathcal{J}; \alpha_-, \alpha_+) = \{\mathbf{p}_c \mid c \in \mathcal{C}_\ell(H^{21}, \mathcal{J}; \alpha_-, \alpha_+)\} \subset \mathcal{N}_\ell(X, \mathcal{J}, H^{21}; \alpha_-, \alpha_+)$$

and for each $c \in \mathcal{C}_\ell(H^{21}, \mathcal{J}; \alpha_-, \alpha_+)$ we took obstruction bundle data $\mathfrak{E}_{\mathbf{p}_c}$ centered at $\mathbf{p}_c$. It corresponds to the case $s_0 = 0, 1$ in the current circumstances. We write $\mathcal{A}_\ell(H^{21,s_0}, \mathcal{J}_{s_0}; \alpha_-, \alpha_+)$, $\mathcal{C}_\ell(H^{21,s_0}, \mathcal{J}_{s_0}; \alpha_-, \alpha_+)$, $\mathfrak{E}_{\mathbf{p}_c,s_0}$ to specify s_0. We also took $W(\mathbf{p}_c)$ such that (6.9) is satisfied. We denote them by $\overline{W}(\mathbf{p}_c, s_0)$ here. Note it is a neighborhood of $\mathbf{p}_c$ in $\mathcal{N}_\ell(X, \mathcal{J}^{s_0}, H^{21,s_0}; \alpha_-, \alpha_+)$.

We take a closed neighborhood $W(\mathbf{p}_c)$ of $\mathbf{p}_c$ in $\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+)$ such that

$$W(\mathbf{p}_c) \cap \mathcal{N}_\ell(X, \mathcal{J}^{s_0}, H^{21,s_0}; \alpha_-, \alpha_+) = \overline{W}(\mathbf{p}_c, s_0).$$

Moreover, using Situation 7.1 (4) we may assume

$$\begin{aligned} W(\mathbf{p}_c) &= \overline{W}(\mathbf{p}_c, s_0) \times [0, \epsilon) \qquad \text{for } s_0 = 0, \\ W(\mathbf{p}_c) &= \overline{W}(\mathbf{p}_c, s_0) \times (1-\epsilon, 1] \qquad \text{for } s_0 = 1. \end{aligned} \tag{7.2}$$

Now we similarly take a finite set

$$\begin{aligned}&\mathcal{A}_\ell(H^{21,[0,1]},\mathcal{J}^{[0,1]};\alpha_-,\alpha_+)=\{\mathbf{p}_c\mid c\in\mathcal{C}_\ell(H^{21,[0,1]},\mathcal{J}^{[0,1]};\alpha_-,\alpha_+)\}\\&\quad\subset\mathcal{N}_\ell(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\alpha_-,\alpha_+)\backslash\partial_{\mathfrak{C}^v}(\mathcal{N}_\ell(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\alpha_-,\alpha_+))\end{aligned}$$

and a closed neighborhood $W(\mathbf{p}_c)$ of $\mathbf{p}_c$ in $\mathcal{N}_\ell(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\alpha_-,\alpha_+)$ for each $c\in\mathcal{C}_\ell(H^{21,[0,1]},\mathcal{J}^{[0,1]};\alpha_-,\alpha_+)$ such that the following conditions are satisfied.

Condition 7.7. (1)

$$\begin{aligned}&\mathcal{N}_\ell(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\alpha_-,\alpha_+)\\&\quad=\bigcup_{c\in\mathcal{C}_\ell(H^{21,[0,1]},\mathcal{J}^{[0,1]};\alpha_-,\alpha_+)}\operatorname{Int}W(\mathbf{p}_c)\cup\\&\qquad\bigcup_{s_0\in\{0,1\}}\ \bigcup_{c\in\mathcal{C}_\ell(H^{21,s_0},\mathcal{J}^{s_0};\alpha_-,\alpha_+)}\operatorname{Int}\overline{W}(\mathbf{p}_c,s_0).\end{aligned}$$

(2) If $c\in\mathcal{C}_\ell(H^{21,[0,1]},\mathcal{J}^{[0,1]};\alpha_-,\alpha_+)$, then

$$W(\mathbf{p}_c)\cap\partial_{\mathfrak{C}^v}(\mathcal{N}_\ell(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\alpha_-,\alpha_+))=\emptyset.$$

(3) Any element of $W(\mathbf{p}_c)$ together with certain marked points is ϵ_c-close to $\mathbf{p}_c\cup\vec{w}\cup\vec{w}_{\rm can}$.

Let $\mathbf{q}\in\mathcal{N}_\ell(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\alpha_-,\alpha_+)$. We put

$$\begin{aligned}\mathcal{E}(\mathbf{q})=\{&\mathbf{p}_c\in\mathcal{A}_\ell(H^{21,[0,1]},\mathcal{J}^{[0,1]};\alpha_-,\alpha_+)\\&\cup\bigcup_{s_0=0,1}\mathcal{A}_\ell(H^{21,s_0},\mathcal{J}_{s_0};\alpha_-,\alpha_+)\mid\mathbf{q}\in\overline{W}(\mathbf{p}_c,s_0)\}.\end{aligned}$$

By taking $\mathcal{A}_\ell(H^{21,[0,1]},\mathcal{J}^{[0,1]};\alpha_-,\alpha_+)$ suitably, we may take $\epsilon_c>0$ in Condition 7.7 (3) small enough as we wish. Therefore, for each $\mathbf{p}_c\in\mathcal{E}(\mathbf{q})$ we can uniquely find $\vec{w}_c^{\mathbf{q}}$ for any $\mathbf{q}\in W(\mathbf{p}_c)$ such that $\mathbf{q}\cup\vec{w}_c^{\mathbf{q}}$ is ϵ_c'-close to $\mathbf{p}_c\cup\vec{w}_{\mathbf{p}_c}\cup\vec{w}_{\mathbf{p}_c,{\rm can}}$ and $\vec{w}_c^{\mathbf{q}}$ satisfies the transversal constraint, and, moreover, the linearization operator $D_{\mathbf{q}}\overline{\partial}_{J,H}$ at $\mathbf{q}$ in (4.7) is surjective $\mod\oplus_{\rm v}\operatorname{Im}I_{\mathbf{p}_c,{\rm v};\mathbf{q}}$, where $I_{\mathbf{p}_c,{\rm v};\mathbf{q}}$ is the map in (7.1) for $\mathbf{p}=\mathbf{p}_c$, $\mathbf{q}=(\Sigma',u',\varphi')$.

Now let us consider $(\mathfrak{Y}\cup\bigcup_{\mathbf{p}_c\in\mathcal{E}(\mathbf{q})}\vec{w}_c',u',\varphi',a_0',s')$ such that $(\mathfrak{Y}\cup\vec{w}_c',u',\varphi',a_0',s')$ is ϵ-close to $\mathbf{q}\cup\vec{w}_c^{\mathbf{q}}$ for each $\mathbf{p}_c\in\mathcal{E}(\mathbf{q})$. We may choose $\epsilon>0$ and $\epsilon_c>0$ small so that we obtain a complex linear map

$$I_{\mathbf{p}_c,{\rm v};\Sigma',u',\varphi',a_0',s'}:E_{\mathbf{p}_c,{\rm v}}(\mathfrak{y}_{\mathbf{p}_c})\to C^\infty(\Sigma';(u')^*TX\otimes\Lambda^{0,1})$$

as in (7.1). We now put

$$E((\mathfrak{Y}\cup\bigcup_{c\in\mathcal{B}}\vec{w}_c',u',\varphi',a_0',s');\mathbf{q};\mathcal{B})=\bigoplus_{c\in\mathcal{B}}\bigoplus_{\rm v}\operatorname{Im}I_{\mathbf{p}_c,{\rm v};\Sigma',u',\varphi',a_0',s'},\quad(7.3)$$

where $\mathcal{B}$ is a subset of $\{c\mid\mathbf{p}_c\in\mathcal{E}(\mathbf{q})\}$. By perturbing the bundle part of the obstruction bundle data $\mathfrak{E}_{\mathbf{p}_c}$ for $\mathbf{p}_c\in\mathcal{A}_\ell(H^{21,[0,1]},\mathcal{J}^{[0,1]};\alpha_-,\alpha_+)$ slightly we may assume that the right hand side of (7.3) is a direct sum.[31] (We can do

[31] See [20, Subsection 11.4].

so without changing the obstruction bundle data which was already fixed at $s_0 = 0, 1$.)

Now we define $V(\mathbf{q}, \epsilon_1, \mathcal{B})$ to be the set of the isomorphism classes of $(\mathfrak{Y} \cup \vec{w}'_c, u', \varphi', a'_0, s')$ satisfying Definition 6.17 (1)–(5) and $|s' - s_{\mathbf{q}}| < \epsilon_1$. Note that we use s' to parametrize a Hamiltonian and a family of almost complex structures appearing in Definition 6.17 (1),(2) and $s_{\mathbf{q}}$ is the $[0, 1]$ component of $\mathbf{q}$.

In the same way as in Lemma 4.29, we can show that $V(\mathbf{q}, \epsilon_1, \mathcal{B})$ is a smooth manifold with corners by choosing various constants sufficiently small. Then in the same way as the proof of Theorem 6.4 (2), we can find other objects so that $V(\mathbf{q}, \epsilon_1, \mathcal{B})$ together with them is a Kuranishi chart of $\mathbf{q}$. We can also show the existence of coordinate changes in the same way as in Lemmas 4.33 and 4.34. We shrink the Kuranishi neighborhood and discuss in the same way as in Lemmas 4.33, 4.34 and [16, Chapter 8] to obtain a Kuranishi structure on $\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+)$. Using Condition 7.7 (2) and (7.2) we can show that this Kuranishi structure is $\mathfrak{C}^v$-collared and its restriction to $\partial[0, 1] = \{0, 1\}$ coincides with the given one. See the paragraphs after Definition 7.3 for the $\mathfrak{C}^v$-collared-ness.

We next define the Kuranishi structure on $\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+)^{\boxplus 1}$ extending the one on $\mathcal{N}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+)$ in the way as in the previous sections as follows.

We consider the following fiber product. (Here and in the next condition we use Notation 6.19.)

$$\begin{aligned}
&\mathcal{M}_{\vec{\ell}_{1,A_1,1}}(X, J_1, H^1; \vec{\alpha}_{1,A_1,1})\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{1,A_1,2}}(X, J_1, H^1; \vec{\alpha}_{1,A_1,2}) {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{1,A_1,j}}(X, J_1, H^1; \vec{\alpha}_{1,A_1,j}) {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{1,A_1,a_1+1}}(X, J_1, H^1; \vec{\alpha}_{1,A_1,a_1+1})\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{N}_{\vec{\ell}_{1,A_1,a_1+1},\ell',\vec{\ell}_{2,A_2,1}}(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \vec{\alpha}_{1,A_1,a_1+1}, \vec{\alpha}_{2,A_2,1})\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{2,A_2,2}}(X, J_2, H^2; \vec{\alpha}_{2,A_2,2}) {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{2,A_2,j}}(X, J_2, H^2; \vec{\alpha}_{2,A_2,j}) {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{2,A_2,a_2+1}}(X, J_2, H^2; \vec{\alpha}_{2,A_2,a_2+1}).
\end{aligned} \tag{7.4}$$

(Recall $\vec{\alpha}_{1,A_1,a_1+1} = (\alpha_{1,i(A_1,a_1)}, \ldots, \alpha_{1,m_1})$ and $\vec{\alpha}_{2,A_2,a_2+1} = (\alpha_{1,i(A_2,a_2)}, \ldots, \alpha_{2,m_2+1})$ and other notations in Notation 6.19.) The fiber product (7.4) is the same as in (6.17) except one of the factors in (6.17) is replaced by

$$\mathcal{N}_{\vec{\ell}_{1,A_1,a_1+1},\ell',\vec{\ell}_{2,A_2,1}}(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; (\vec{\alpha}_{1,A_1,a_1+1}, \vec{\alpha}_{2,A_2,1})). \tag{7.5}$$

We use the embedding

$$\begin{aligned}
&\mathrm{id} \times \mathcal{I}_{A_1,B_1,C_1} \times \mathcal{I}_{A_2,B_2,C_2} : \mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \vec{\alpha}_1, \vec{\alpha}_2) \times [-1, 0]^{b_1+b_2}\\
&\quad \to \mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \vec{\alpha}_1, \vec{\alpha}_2)^+
\end{aligned} \tag{7.6}$$

which is defined in the same way as in (6.16).

We use the fact that (7.5) is a component of

$$\widehat{S}_{c_{a_1+1}(1,A_1)+c_1(2,A_2)}\ (\mathcal{N}_{\vec{\ell}_{1,A_1,a_1+1,C_1},\ell'',\vec{\ell}_{2,A_2,1,C_2}}(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\vec{\alpha}_{1,A_1,a_1+1},\vec{\alpha}_{2,A_2,1})).$$

As we mentioned already, the fiber product factor of (7.4) other than (7.5) is

$$\mathcal{M}_{\vec{\ell}_{r,A_r,j}}(X,J_r,H^r;\vec{\alpha}_{r,A_r,j}), \tag{7.7}$$

which is a component of

$$\widehat{S}_{c_j(r,A_r)}(\mathcal{M}_{\vec{\ell}_{r,A_1,j},C_r}(X,J_r,H^r;\vec{\alpha}_{r,A_r,j,C_r})).$$

We put

$$\begin{aligned}&\mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\vec{\alpha}_1,\vec{\alpha}_2)^+\\&\quad=\mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\vec{\alpha}_1,\vec{\alpha}_2)\times[-1,0]^{m-1},\end{aligned}$$

where $m=\#\vec{\alpha}_1+\#\vec{\alpha}_2-2$.

Condition 7.8. We require the K-system

$\{(\mathcal{M}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\vec{\alpha}_1,\vec{\alpha}_2)^+,\ \widehat{\mathcal{U}}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\vec{\alpha}_1,\vec{\alpha}_2))\}$ satisfies the following.

(1) The restriction of $\widehat{\mathcal{U}}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\vec{\alpha}_1,\vec{\alpha}_2)$ to the image of the embedding (7.6) is the fiber product of the following Kuranishi strucrures. (Here we use the fiber product description (7.4).)

 (a) The restrictions of the Kuranishi structure

$$\widehat{\mathcal{U}}_{\vec{\ell}_{A_r,j,C_r}}(X,J_r,H^r;\vec{\alpha}_{r,A_r,j,C_r})$$

 on $\mathcal{M}_{\vec{\ell}_{r,A_r,j},C_r}(X,J_r,H^r;\vec{\alpha}_{r,A_r,j,C_r})^+$ to the space which is a direct product (7.7) $\times[-1,0]^{m_j(r,A_r,C_r)-1}$.

 (b) The restrictions of the Kuranishi structure

$$\widehat{\mathcal{U}}_{\vec{\ell}_{1,A_1,a_1+1,C_1},\ell'',\vec{\ell}_{2,A_2,1,C_2}}(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\vec{\alpha}_{1,A_1,a_1+1},\vec{\alpha}_{2,A_2,1})$$

 on $\mathcal{N}_{\vec{\ell}_{1,A_1,a_1+1,C_1},\ell'',\vec{\ell}_{2,A_2,1,C_2}}(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\vec{\alpha}_{1,A_1,a_1+1},\vec{\alpha}_{2,A_2,1})^+$ to the space which is a direct product

$$(7.5)\times[-1,0]^{m_{a_1}(1,A_1,C_1)}\times[-1,0]^{m_0(2,A_2,C_2)}.$$

(2) The restriction of $\widehat{\mathcal{U}}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\vec{\alpha}_1,\vec{\alpha}_2)$ to the vertical boundary $\partial_{\mathfrak{C}^v}(\mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\vec{\alpha}_1,\vec{\alpha}_2)^{\mathfrak{C}^h\boxplus 1})$ is isomorphic to the Kuranishi structure $\bigcup_{s_0=0,1}\widehat{\mathcal{U}}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J}^{s_0},H^{21,s_0};\vec{\alpha}_1,\vec{\alpha}_2)$, which we produced in Proposition 6.21.

Proposition 7.9. *There exists a K-system*

$\{(\mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\vec{\alpha}_1,\vec{\alpha}_2)^+,\ \widehat{\mathcal{U}}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X,\mathcal{J}^{[0,1]},H^{21,[0,1]};\vec{\alpha}_1,\vec{\alpha}_2))\}$ *satisfying the following properties.*

(1) *They satisfy Condition 7.8.*

(2) *Let* $\mathfrak{C}$ *be the union of the components of* $\mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \vec{\alpha}_1, \vec{\alpha}_2)^+$ *which are in* $\mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \vec{\alpha}_1, \vec{\alpha}_2) \times \partial([-1,0]^{m-1})$. *Then the Kuranishi structure* $\widehat{\mathcal{U}}_{\vec{\ell}_1,\ell',\vec{\ell}_2}(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \vec{\alpha}_1, \vec{\alpha}_2)^+$ *is* $\mathfrak{C}$-*collared in the sense of Remark* 5.6.

(3) *For the case* $\vec{\alpha}_1 = (\alpha_-)$ *and* $\vec{\alpha}_2 = (\alpha_+)$, $\widehat{\mathcal{U}}_{\emptyset,\ell,\emptyset}(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \vec{\alpha}_1, \vec{\alpha}_2)^+$ *coincides with the Kuranishi structure we constructed in the first half of the proof of Theorem* 7.4 *and Proposition* 7.5.

Proof. The proof is the same as that of Proposition 5.5. □

Now using Proposition 7.9 we modify the Kuranishi structures on the spaces $\mathcal{N}_\ell(X, H^{21,[0,1]}, \mathcal{J}^{[0,1]}; \alpha_-, \alpha_+)^{\boxplus 1}$ and complete the proof of Theorem 7.4 and Proposition 7.5. □

Remark 7.10. In this section we have constructed a homotopy between two morphisms. We can continue this process to obtain a homotopy of homotopies etc. In this paper we do not need such a higher homotopy by the following reason: First we note that what we obtain in this paper is a linear K-system. In [23, Chapter 16] we introduced the notion of an inductive system of linear K-systems.[32] Sometimes such a structure is easier to obtain since we need to study only a finite number of moduli spaces at each stage. The proof we provide in this paper defines a system of *Kuranishi structures* on infinitely many moduli spaces at once. This is the reason we do not need to study homotopy of homotopies. On the other hand, to construct a system of *CF-perturbations* on such infinitely many moduli spaces, we need to stop at a certain energy level and use a homotopy inductive limit argument. Such an argument is given in [23, Chapter 19]. (We can quote the statements of [23] literary.) Homotopy of homotopies we used in the homotopy inductive limit argument of [23, Chapter 19] is a direct product of the homotopy obtained in Theorem 7.4 and an interval $[0, 1]$ in our case.

8. Composition of morphisms

8.1. Statement

Situation 8.1. (1) Let H^r $(r = 1, 2, 3)$ be periodic Hamiltonian functions which are Morse–Bott non-degenerate in the sense of Condition 2.1.

(2) For each pair of r, r' with $r \neq r' \in \{1, 2, 3\}$, let $H^{r'r} : X \times \mathbb{R} \times S^1 \to \mathbb{R}$ be a smooth function and $\mathcal{J}^{r'r} = \{J_{\tau;r'r} \mid \tau \in \mathbb{R}\}$ be an $\mathbb{R}$-parametrized smooth family of tame almost complex structures on X.

(3) We assume that $H^{r'r}$, $\mathcal{J}^{r'r}$ are as in Situation 6.1 with H^1,H^2, J_1, J_2, $\mathcal{J}$ replaced by H^r,$H^{r'}$, J_r, $J_{r'}$, $\mathcal{J}^{r'r}$, respectively.

(4) Let $\mathcal{F}_r := \mathcal{F}_X(H^r, J_r)$ be the linear K-system constructed by Theorem 2.9 from H^r and J_r. We made the choices during the constructions.

[32]It is a structure which gives a sequence of the partial version of the linear K-system for which the moduli spaces of energy $\le E_n$ $(n = 0, 1, \dots,)$ are used. It gives also a homotopy equivalence (modulo E_n) between the structure with the moduli spaces of energy $\le E_n$ and that with the moduli spaces of energy $\le E_{n+1}$.

(5) Let $\mathfrak{N}_{r'r} : \mathcal{F}_r \to \mathcal{F}_{r'}$ be the morphism defined by Theorem 6.4 using $H^{r'r}$ and $\mathcal{J}^{r'r}$ in place of H^{21} and $\mathcal{J}$. We made the choices during the construction.

$\square$

In this section, we prove the following.

Theorem 8.2. *In Situation 8.1 the composition $\mathfrak{N}_{32} \circ \mathfrak{N}_{21}$ is homotopic to $\mathfrak{N}_{31}$.*

Proof. By Lemma 7.2 and Theorem 7.4, the homotopy class of $\mathfrak{N}_{r'r}$ is independent of $\mathcal{J}^{r'r}$, $H^{r'r}$ or other choices we made during the construction. So it suffices to prove Theorem 8.2 for certain fixed choices of them. We take the choices as follows. We first take $\mathcal{J}^{21}$, H^{21} and $\mathcal{J}^{32}$, H^{32}. *We assume that they satisfy Situation 6.1 (1)(2)(i)(ii) with* ± 1 *replaced by* $\pm 1/4$. We then define $\mathcal{J}^{31}$, H^{31} as follows.

$$H^{31}_\tau = \begin{cases} H^{21}_{\tau+1/2} & \text{if } \tau \le 0, \\ H^{32}_{\tau-1/2} & \text{if } \tau \ge 0. \end{cases} \tag{8.1}$$

$$J^{31}_\tau = \begin{cases} J^{21}_{\tau+1/2} & \text{if } \tau \le 0, \\ J^{32}_{\tau-1/2} & \text{if } \tau \ge 0. \end{cases} \tag{8.2}$$

It is easy to see that they satisfy Situation 6.1.

We will construct a homotopy between $\mathfrak{N}_{32} \circ \mathfrak{N}_{21}$ and $\mathfrak{N}_{31}$ where $\mathfrak{N}_{32}$, $\mathfrak{N}_{21}$, $\mathfrak{N}_{31}$ are obtained by Theorem 6.4 using those choices.

The interpolation space of the homotopy is obtained by compactifying the solution space of the τ, t, T dependent Hamiltonian perturbed pseudo-holomorphic curve equation. We will use the following two parameter family of Hamiltonians. Let $T \ge 0$ and $\tau \in \mathbb{R}$. We put

$$H^{31,T}_\tau = \begin{cases} H^{21}_{\tau+1/2+T} & \text{if } \tau \le 0, \\ H^{32}_{\tau-1/2-T} & \text{if } \tau \ge 0, \end{cases} \tag{8.3}$$

$$J^{31}_{\tau,T} = \begin{cases} J^{21}_{\tau+1/2+T} & \text{if } \tau \le 0, \\ J^{32}_{\tau-1/2-T} & \text{if } \tau \ge 0. \end{cases} \tag{8.4}$$

We put $\mathcal{J}^{31,T} = \{J^{31}_{\tau+1/2+T}\}$ and consider

$$\bigcup_{T \ge 0} \mathcal{N}_\ell(X, \mathcal{J}^{31,T}, H^{31,T}; \alpha_-, \alpha_+) \times \{T\}, \tag{8.5}$$

where $\mathcal{N}_\ell(X, \mathcal{J}^{31,T}, H^{31,T}; \alpha_-, \alpha_+)$ is as in Definition 6.10. We can define a topology on it in the same way as in Definition 3.17. This space then becomes Hausdorff. However, it is not compact since the domain $[0, \infty)$ of T is not compact. We compactify it by adding certain space at $T = \infty$ as follows.

Definition 8.3. Let $\alpha_- \in \mathfrak{A}_1$ and $\alpha_+ \in \mathfrak{A}_3$. (Here $\mathfrak{A}_r$ be the index set of the critical submanifolds of the linear K-system $\mathcal{F}_r$.) The set $\widehat{\mathcal{N}}_\ell(X, \mathcal{J}^{31,\infty}, H^{31,\infty}; \alpha_-, \alpha_+)$ consists of $((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi)$ satisfying the following conditions:

(1) $(\Sigma, (z_-, z_+, \vec{z}))$ is a genus zero semi-stable curve with $\ell+2$ marked points.
(2) φ is a parametrization of the mainstream.
(3) Σ_{a_1}, Σ_{a_2} are two of the mainstream components. We call them the *first main component* and the *second main component*.
(4) For each extended main stream component $\widehat{\Sigma}_a$, the map u induces $u_a : \widehat{\Sigma}_a \backslash \{z_{a,-}, z_{a,+}\} \to X$ which is a continuous map.
(5) We define the relation $<$ on the set of mainstream components as in Definition 6.5. We require $a_1 < a_2$. If Σ_a is a mainstream component and $\varphi_a : \mathbb{R} \times S^1 \to \Sigma_a$ is as above, then the composition $u_a \circ \varphi_a$ satisfies the equation

$$\frac{\partial(u_a \circ \varphi_a)}{\partial \tau} + J_{a,\tau}\left(\frac{\partial(u_a \circ \varphi_a)}{\partial t} - \mathfrak{X}_{H^a_{\tau,t}} \circ (u_a \circ \varphi_a)\right) = 0, \tag{8.6}$$

where

$$H^a_{\tau,t} = \begin{cases} H^1_t & \text{if } a < a_1, \\ H^{21}_{\tau,t} & \text{if } a = a_1, \\ H^2_t & \text{if } a_1 < a < a_2, \\ H^{32}_{\tau,t} & \text{if } a = a_2, \\ H^3_t & \text{if } a > a_2, \end{cases}$$

and

$$J_{a,\tau} = \begin{cases} J_1 & \text{if } a < a_1, \\ J^{21}_\tau & \text{if } a = a_1, \\ J_2 & \text{if } a_1 < a < a_2, \\ J^{32}_\tau & \text{if } a = a_2, \\ J_3 & \text{if } a > a_2. \end{cases}$$

(6)

$$\int_{\mathbb{R}\times S^1} \left\| \frac{\partial(u \circ \varphi_a)}{\partial \tau} \right\|^2 \mathrm{d}\tau \mathrm{d}t < \infty.$$

(7) Suppose Σ_{v} is a bubble component in $\widehat{\Sigma}_a$. Let $\varphi_a(\tau, t)$ be the root of the tree of sphere bubbles containing Σ_{v}. Then u is J-holomorphic on Σ_{v} where

$$J = \begin{cases} J_1 & \text{if } a < a_1, \\ J^{21}_\tau & \text{if } a = a_1, \\ J_2 & \text{if } a_1 < a < a_2, \\ J^{32}_\tau & \text{if } a = a_2, \\ J_3 & \text{if } a > a_2. \end{cases}$$

(8) If Σ_a and $\Sigma_{a'}$ are mainstream components and $z_{a,+} = z_{a',-}$, then

$$\lim_{\tau\to+\infty} (u_a \circ \varphi_a)(\tau, t) = \lim_{\tau\to-\infty} (u_{a'} \circ \varphi_{a'})(\tau, t)$$

holds for each $t \in S^1$. ((6) and Lemma 6.3 imply that the left and right hand sides both converge.)

(9) If Σ_a, $\Sigma_{a'}$ are mainstream components and $z_{a,-} = z_-$, $z_{a',+} = z_+$, then there exist $(\gamma_\pm, w_\pm) \in R_{\alpha_\pm}$ such that

$$\lim_{\tau \to -\infty} (u_a \circ \varphi_a)(\tau, t) = \gamma_-(t)$$
$$\lim_{\tau \to +\infty} (u_{a'} \circ \varphi_{a'})(\tau, t) = \gamma_+(t).$$

Moreover,

$$[u_*[\Sigma]] \# w_- = w_+$$

where # is the obvious concatenation.

(10) We assume $((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi)$ is stable in the sense of Definition 8.4 below.

Assume that $((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi)$ satisfies (1)–(9) above. The *extended automorphism group* $\mathrm{Aut}^+((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi)$ of $((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2)), u, \varphi)$ consists of maps $v : \Sigma \to \Sigma$ satisfying (1)(2)(3) (5) of Definition 6.7 and $\tau_{a_1} = \tau_{a_2} = 0$. The *automorphism group* denoted by $\mathrm{Aut}((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi)$ of $((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi)$ consists of the elements of $\mathrm{Aut}^+((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi)$ such that σ in (5) of Definition 6.7 is the identity.

Definition 8.4. An element $((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi)$ in Definition 8.3 is said to be *stable* if the group $\mathrm{Aut}((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi)$ is a finite group.

We can define the equivalence relation $\sim_2$ on $\widehat{\mathcal{N}}_\ell(X, \mathcal{J}^{31,\infty}, H^{31,\infty}; \alpha_-, \alpha_+)$ in the same way as in Definition 3.7 except we require $\tau_{a_1} = \tau_{a_2} = 0$. We put

$$\mathcal{N}_\ell(X, \mathcal{J}^{31,\infty}, H^{31,\infty}; \alpha_-, \alpha_+) = \widehat{\mathcal{N}}_\ell(X, \mathcal{J}^{31,\infty}, H^{31,\infty}; \alpha_-, \alpha_+)/ \sim_2 . \tag{8.7}$$

Definition 8.5. The set $\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)$ is the union of (8.5) and $\mathcal{N}_\ell(X, \mathcal{J}^{31,\infty}, H^{31,\infty}; \alpha_-, \alpha_+)$.

We can define a topology on $\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)$ in the same way as in Definition 3.17 and show that it is Hausdorff and compact. Theorem 8.2 will follow from the next result. We define the topological space

$$\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)^{\boxplus 1}$$

in the same way as in Definition 5.1.

Theorem 8.6. (1) *There exists a Kuranishi structure on the compact space* $\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)$.

(2) *The Kuranishi structure in (1) extends to a Kuranishi structure on the space* $\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)^{\boxplus 1}$, *which becomes the interpolation space of a homotopy between* $\mathfrak{N}_{32} \circ \mathfrak{N}_{21}$ *and* $\mathfrak{N}_{31}$.

8.2. Proof of Theorem 8.6 (1): Kuranishi structure

Proof. The proof of Theorem 8.6 occupies the rest of this section. In this subsection we prove (1). The construction of the Kuranishi structure on the space $\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)$ is mostly the same as that of the first half of the proof of Theorem 7.4, where we constructed the Kuranishi structure on

$$\mathcal{M}_\ell(X, \mathcal{J}^{[0,1]}, H^{21,[0,1]}; \alpha_-, \alpha_+).$$

In fact, we are studying $[0,\infty)\times\mathbb{R}\times[0,1]$-parametrized family of Hamiltonians and almost complex structures and we are also given the choices which we need for the definition of the Kuranishi structure at $0 \in \partial[0,\infty)$.

The main difference is that we also include $T = \infty$. So we here concentrate on constructing a Kuranishi neighborhood of a point at $\mathcal{N}_\ell(X, \mathcal{J}^{31,\infty}, H^{31,\infty}; \alpha_-, \alpha_+)$.

Let $\mathbf{p} = ((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi)$ be a representative of an element of the moduli space $\mathcal{N}_\ell(X, \mathcal{J}^{31,\infty}, H^{31,\infty}; \alpha_-, \alpha_+)$. We assume that Σ has $k_1 + k_2 + k_3 + 2$ mainstream components. Namely there exit k_1 mainstream components Σ_a with $a < a_1$, k_2 mainstream components Σ_a with $a_1 < a < a_2$ and k_3 mainstream components Σ_a with $a_2 < a$. So there are $k_1 + k_2 + k_3 + 1$ transit points. We consider $\Sigma\backslash\Sigma_{a_1}\backslash\Sigma_{a_2}$ which has three connected components. Among those transit points k_1, k_2+1, k_3 lie on the closure of each of those connected components. We take $T_{1,1}, \dots, T_{1,k_1}$, $T_{2,0}, \dots, T_{2,k_2}$, $T_{3,1}, \dots, T_{3,k_3}$ which are parameters $\in (T_0, \infty]$ to smooth those transit points.

We consider the moduli space $\mathcal{N}_\ell(\text{source})$ as in Definition 6.11. We also define $\mathcal{N}_\ell(\text{source}, \infty)$ as follows. In Definition 8.3 we consider the case when X is one point and $H^1 = H^2 = H^3 = 0$. The space we obtain as $\mathcal{N}_\ell(X, \mathcal{J}^{31,\infty}, H^{31,\infty}; \alpha_-, \alpha_+)$ in that case is the space $\mathcal{N}_\ell(\text{source}, \infty)$ by definition. We put

$$\mathcal{N}_\ell(\text{source}, (T^0, \infty]) = \mathcal{N}_\ell(\text{source}, \infty) \cup \bigcup_{T > T_0} (\mathcal{N}_\ell(\text{source}) \times \{T\}). \quad (8.8)$$

We go back to the situation where we have $\mathbf{p} = ((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi) \in \mathcal{N}_\ell(X, \mathcal{J}^{31,\infty}, H^{31,\infty}; \alpha_-, \alpha_+)$. We take stabilization data at $\mathbf{p}$. Especially we fix a symmetric stabilization $\vec{w}$ of $\mathbf{p}$. We also take the canonical marked point $w_{a,\mathrm{can}}$ on each mainstream component Σ_a where there is at most two special points, (that are $z_{a,-}$ and $z_{z,+}$.) The canonical marked point $w_{a,\mathrm{can}}$ is defined as in Definition 4.8 if $a \neq a_1, a_2$. We do not define canonical marked points in case $a = a_1, a_2$. (See Remark 6.14.) The totality of the canonical marked points is denoted by $\vec{w}_{\mathrm{can}}$. We can prove that $(\Sigma, (z_-, z_+, \vec{z})) \cup \vec{w} \cup \vec{w}_{\mathrm{can}}$ is stable in the same way as in Lemma 4.9. We will define a map

$$\begin{aligned}\Phi_{\mathbf{p}} : \textstyle\prod_{\mathrm{v}} \mathcal{V}(\mathfrak{x}_{\mathrm{v}} \cup \vec{w}_{\mathrm{v}} \cup \vec{w}_{\mathrm{can,v}}) \times (T_0, \infty]^{k_1+k_2+k_3+1} \times \prod_{j=1}^m (((T_{0,j}, \infty] \times S^1)/\sim) \\ \to \mathcal{N}_{\ell+\ell'+\ell''}(\text{source}, (T'^0, \infty]).\end{aligned} \quad (8.9)$$

We first explain the notation in (8.9). $\Sigma_{\rm v}$ is an irreducible components of Σ. $\mathfrak{x}_{\rm v}$ is $\Sigma_{\rm v}$ together with the part of marked points $(z_-, z_+, \vec z)$ on it. $\vec w_{\rm v}$ and $\vec w_{\rm can,v}$ are intersection of $\vec w$ and $\vec w_{\rm can}$ with $\Sigma_{\rm v}$, respectively. $\mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec w_{\rm v} \cup \vec w_{\rm can,v})$ is a neighborhood of $\mathfrak{x}_{\rm v} \cup \vec w_{\rm v} \cup \vec w_{\rm can,v}$ in the moduli space of pointed curves. Namely:

(1) If $\Sigma_{\rm v}$ is a bubble component, then $\mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec w_{\rm v} \cup \vec w_{\rm can,v})$ is an open set of $\overset{\circ}{\mathcal{M}}{}^{\rm cl}_{\ell_{\rm v}}$ for certain $\ell_{\rm v}$. (It is the number of marked or singular points on $\Sigma_{\rm v}$.)

(2) If $\Sigma_{\rm v}$ is a main stream component Σ_a and $a \ne a_1, a_2$, then $\mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec w_{\rm v} \cup \vec w_{\rm can,v})$ is an open set of $\overset{\circ}{\mathcal{M}}_{\ell_{\rm v}}(\text{Source})$. (We include the parametrization $\varphi_{\rm v}$ in $\mathfrak{x}_{\rm v}$ in this case.)

(3) If $\Sigma_{\rm v}$ is a main stream component Σ_a and $a \in \{a_1, a_2\}$, then $\mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec w_{\rm v} \cup \vec w_{\rm can,v})$ is an open set of $\overset{\circ}{\mathcal{N}}_{\ell_{\rm v}}(\text{Source})$. (We include the parametrization $\varphi_{\rm v}$ in $\mathfrak{x}_{\rm v}$ in this case.)

Here we use Notation 4.10 and m is the number of non-transit singular points of Σ.

Recall that the stabilization data at $\mathbf{p}$ contain the local trivialization data in Definition 4.11 (3), which contain the data of a coordinate at each singular point. (See [20, Definition 3.8 (1)].) Using it we can associate a marked Riemann surface to each element of the domain in (8.9). Let Σ' be this curve. Other than Σ' and marked points on it we need to associate a few more data to obtain an element of $\mathcal{N}_{\ell+\ell'+\ell''}(\text{source}, (T'^0, \infty])$. We will explain how to associate those data to an element of the domain in (8.9).

Recall that we take the parameters $T_{1,1}, \dots, T_{1,k_1}$, $T_{2,0}, \dots, T_{2,k_2}$, $T_{3,1}, \dots, T_{3,k_3} \in (T_0, \infty]$ to smooth the transit points.

Case 1: We first consider the case when $T_{2,0} + \cdots + T_{2,k_2} = \infty$. This is equivalent to the condition that at least one of $T_{2,i}$ is infinity. We will obtain an element of $\mathcal{N}_\ell(\text{source}, \infty)$ in this case as follows.

In this case Σ' contains two different components $\Sigma'_{a'_1}$ and $\Sigma'_{a'_2}$ which are obtained by gluing Σ_{a_1} and Σ_{a_2} with other components, respectively. We take them as the first and the second main components, respectively. Using marked points on Σ' corresponding to $z_\pm \in \Sigma$, we can define the notion of mainstream components of Σ'. It is easy to see that $\Sigma'_{a'_1}$ and $\Sigma'_{a'_2}$ are contained in the mainstream.

As mentioned in Definition 4.11 (3), we assume the conditions (a) (b) for the local coordinates at singular points. Since we use them for our gluing, that is, the construction of the map (8.9), we can show the next lemma about the relationship of parametrization of the mainstreams of Σ and of Σ'.

Lemma 8.7. *Let $\varphi_a : \mathbb{R} \times S^1 \to \Sigma_a$ be a parametrization of a mainstream component. We take a biholomorphic map $\varphi'_{a'} : \mathbb{R} \times S^1 \to \Sigma'_{a'} \setminus \{z'_{a',-}, z'_{a',+}\}$ such that*

$$\lim_{\tau \to \pm\infty} \varphi'_{a'}(\tau, t) = z'_{a',\pm}.$$

We use the diffeomorphism $\mathfrak{v}$ *from a complement of the neck region of* Σ *to* Σ'*, which we obtain using the local trivialization data contained in the stabilization data at* $\mathbf{p}$*. Then if the image of* $\mathfrak{v} \circ \varphi_a$ *and* $\varphi'_{a'}$ *intersect, we have*

$$\varphi'_{a'}(\tau, t) = (\mathfrak{v} \circ \varphi_a)(\tau + \tau_0, t + t_0) \tag{8.10}$$

for certain $\tau_0 \in \mathbb{R}$ *and* $t_0 \in S^1$*.*

Using this lemma we can take parametrizations of the main components $\varphi_{a'_i}$ $(i = 1, 2)$ so that τ_0, t_0 become zero for them. By our choice of the local coordinates at transit points given by Definition 4.11 (3) (a) we can choose parametrizations of all the mainstream components so that the number t_0 in (8.10) becomes 0. Note other than the first and second main components, the parametrization of the mainstream component is well-defined only up to a shift on $\mathbb{R}$ direction. Therefore, by Lemma 8.7 we obtain a parametrization of the mainstream. We thus obtain an element of $\mathcal{N}_\ell(\text{source}, \infty)$ in this case.
Case 2: We next consider the case when $T_{2,0} + \cdots + T_{2,k_2} \neq \infty$. We will obtain an element of $\mathcal{N}_\ell(\text{source}) \times \{T\}$ for a certain number $T \in (T'^0, \infty)$. (The number T is determined by the data in the domain of the map (8.9)).

In this case Σ_{a_1} and Σ_{a_2} are glued to become a part of single mainstream component of Σ', which we write as $\Sigma'_{a'_0}$. This will be our main component.

We apply Lemma 8.7 to $a = a_1$ or a_2, and $a' = a'_0$. By Definition 4.11 (3) (a) we may take the same t_0 for both. (Note we use the same $\varphi'_{a'_0}$ for both $a = a_1$ or a_2.) We then have

$$\varphi'_{a'_0}(\tau, t) = (\mathfrak{v} \circ \varphi_{a_i})(\tau - \tau_i, t).$$

We put

$$2T + 1 = \tau_2 - \tau_1. \tag{8.11}$$

Then by shifting $\varphi'_{a'_0}$ in $\mathbb{R}$ direction, we may assume

$$\begin{aligned}\varphi'_{a'_0}(\tau, t) &= (\mathfrak{v} \circ \varphi_{a_1})(\tau + T + 1/2, t),\\ \varphi'_{a'_0}(\tau, t) &= (\mathfrak{v} \circ \varphi_{a_2})(\tau - T - 1/2, t).\end{aligned}$$

By comparing this formula with (8.3), the Hamiltonian term of our Eq. (8.6) is consistent with $\mathfrak{v}$ by this choice. We thus determine the parametrization of the mainstream. Together with T and a'_0 we already defined, this parametrization gives an element of $\mathcal{N}_\ell(\text{source}) \times \{T\}$.

We have thus defined (8.9). We find that this map defines a structure of cornered orbifolds to $\mathcal{N}_\ell(\text{source}, (T'^0, \infty])$ for which (8.9) becomes a diffeomorphism onto its image. We can use this fact to obtain an appropriate coordinate. Namely we identify $(T_0, \infty]^{k_1+k_2+k_3+1}$ with $[0, 1/\log T_0)^{k_1+k_2+k_3+1}$ by $T_i \mapsto 1/\log T_i$. (See Sect. 4.)

Remark 8.8. Note that T is not a part of the coordinate function in a neighborhood of $T = \infty$. It is easy to see that $T = T_{2,0} + \cdots + T_{2,k_2} +$ bounded function. $1/\log T$ is not a smooth function of $1/\log T_i$. (See [11, page 778] [23, Chapter 25] for a related issue.) When we are away from $T = \infty$, T can be taken as a part of coordinates.

The construction of the Kuranishi structure of $\mathcal{N}_\ell(X,\mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]};\alpha_-,\alpha_+)$ in a neighborhood of $\mathcal{N}_\ell(X,\mathcal{J}^{31,\infty},H^{31,\infty};\alpha_-,\alpha_+)$, is similar to the first half of the proof of Theorem 7.4 using the construction we gave above and proceed as follows.

We take a finite set

$$\begin{aligned}\mathcal{A}_\ell(H^{32,\infty},\mathcal{J}^{32,\infty};\alpha_-,\alpha_+) &= \{\mathbf{p}_c \mid c\in\mathfrak{C}_\ell(H^{32,\infty},\mathcal{J}^{32,\infty};\alpha_-,\alpha_+)\}\\ &\subset \mathcal{N}_\ell(X,\mathcal{J}^{32,\infty},H^{32,\infty};\alpha_-,\alpha_+).\end{aligned}$$

For each $c\in\mathfrak{C}_\ell(H^{32,\infty},\mathcal{J}^{32,\infty};\alpha_-,\alpha_+)$ we take obstruction bundle data $\mathfrak{E}_{\mathbf{p}_c}$ centered at $\mathbf{p}_c$[33] and a closed neighborhood $W(\mathbf{p}_c)$ of $\mathbf{p}_c$ in our moduli space

$$\mathcal{N}_\ell(X,\mathcal{J}^{31,[0,\infty]},H^{31,[0,\infty]};\alpha_-,\alpha_+)$$

with the following property. For each element $\mathbf{q}\in W(\mathbf{p}_c)$ there exists $\vec{w}^{\mathbf{q}}_{\mathbf{p}_c}$ such that $\mathbf{q}\cup\vec{w}^{\mathbf{q}}_{\mathbf{p}_c}$ is ϵ_c-close to $\mathbf{p}_c\cup w_{\mathbf{p}_c}\cup\vec{w}_{\rm can}$. (We can use (8.9) to define the notion of ϵ-closeness in the same way as in Definition 4.16.) We also require

$$\bigcup_{c\in\mathfrak{C}_\ell(H^{32,\infty},\mathcal{J}^{32,\infty};\alpha_-,\alpha_+)} \operatorname{Int} W(\mathbf{p}_c)\supset\mathcal{N}_\ell(X,\mathcal{J}^{32,\infty},H^{32,\infty};\alpha_-,\alpha_+). \tag{8.12}$$

We can choose $\epsilon_c>0$ above such that for any $\mathbf{q}\in W(\mathbf{p})$ there exists $\vec{w}^{\mathbf{q}}_{\mathbf{p}}$ uniquely with the following properties:

(1) $\mathbf{q}\cup\vec{w}^{\mathbf{q}}_{\mathbf{p}}$ is $\epsilon_{\mathbf{p}}$-close to $\mathbf{p}_c\cup w_{\mathbf{p}_c}\cup\vec{w}_{\rm can}$.
(2) $\mathbf{q}\cup\vec{w}^{\mathbf{q}}_{\mathbf{p}}$ satisfies the transversal constraint. (The transversal constraint is defined in the same way as in Definition 6.15.)
(3) For each irreducible component $\Sigma_{\rm v}$ of the source curve of $\mathbf{p}$ there exist a finite dimensional complex subspace $\operatorname{Im} I_{\mathbf{p},{\rm v};\mathbf{q}}$ of $L^2_{m,\delta}(\Sigma_{\rm v};u^*TX\otimes\Lambda^{0,1}\Sigma_{\rm v})$ such that the linearization operator of the differential equation at $\mathbf{q}$ in Definition 8.3 (5) (7) is surjective $\mod \oplus_{\rm v}\operatorname{Im}I_{\mathbf{p},{\rm v};\mathbf{q}}$.

The proof of this fact is similar to that of Lemma 4.22.

For each $\mathbf{q}\in\mathcal{N}_\ell(X,\mathcal{J}^{31,[0,\infty]},H^{31,[0,\infty]};\alpha_-,\alpha_+)$ which lies in a neighborhood of $\mathcal{N}_\ell(X,\mathcal{J}^{32,\infty},H^{32,\infty};\alpha_-,\alpha_+)$, we put

$$\mathcal{E}(\mathbf{q})=\{\mathbf{p}_c\mid\mathbf{q}\in W(\mathbf{p}_c)\}.$$

Let $\mathcal{B}$ be a nonempty subset of $\{c\mid\mathbf{p}_c\in\mathcal{E}(\mathbf{q})\}$. We consider $(\mathfrak{Y}\cup\bigcup_{c\in\mathcal{B}}\vec{w}'_c,u',\varphi',a'_1,a'_2)$ (resp. $(\mathfrak{Y}\cup\bigcup_{c\in\mathcal{B}}\vec{w}'_c,u',\varphi',a'_0,T')$) such that for each c, $(\mathfrak{Y}\cup\vec{w}'_c,u',\varphi',a'_1,a'_2)$ and $(\mathfrak{Y}\cup\vec{w}'_c,u',\varphi',a'_0,T')$ (resp. $(\mathfrak{Y}\cup\vec{w}'_c,u',\varphi',a'_1,a'_2)$) are ϵ-close to $\mathbf{q}\cup\vec{w}^{\mathbf{q}}_c$. If $\epsilon>0$ is small, then $(\mathfrak{Y}\cup\vec{w}'_c,u',\varphi',a'_0,T')$ (resp. $(\mathfrak{Y}\cup\vec{w}'_c,u',\varphi',a'_1,a'_2)$) is ϵ-close to $\mathbf{p}_c\cup\vec{w}_c$. Therefore, we can define a complex linear map

$$I_{\mathbf{p}_c,{\rm v};\Sigma',u',\varphi',a'_0,T'}:E_{\mathbf{p}_c,{\rm v}}(\mathfrak{y}_{\mathbf{p}_c})\to C^\infty(\Sigma';(u')^*TX\otimes\Lambda^{0,1})$$

(resp.

$$I_{\mathbf{p}_c,{\rm v};\Sigma',u',\varphi',a'_1,a'_2}:E_{\mathbf{p}_c,{\rm v}}(\mathfrak{y}_{\mathbf{p}_c})\to C^\infty(\Sigma';(u')^*TX\otimes\Lambda^{0,1}))$$

[33]The notion of obstruction bundle data for an element of $\mathcal{N}_\ell(X,\mathcal{J}^{32,\infty},H^{32,\infty};\alpha_-,\alpha_+)$ can be defined in the similar way as before.

in the same way as in (6.8) for each irreducible component v of $\mathbf{p}_c$. (We use the map (8.9) here while we used the map (6.6) in (6.8).) We put

$$E((\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c, u', \varphi', a'_0, T'); \mathbf{q}; \mathcal{B}) = \bigoplus_{c\in\mathcal{B}} \bigoplus_{\mathrm{v}} \operatorname{Im} I_{\mathbf{p}_c,\mathrm{v};\Sigma',u',\varphi',a'_0,T'}, \tag{8.13}$$

or

$$E((\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c, u', \varphi', a'_1, a'_2); \mathbf{q}; \mathcal{B}) = \bigoplus_{c\in\mathcal{B}} \bigoplus_{\mathrm{v}} \operatorname{Im} I_{\mathbf{p}_c,\mathrm{v};\Sigma',u',\varphi',a'_1,a'_2}. \tag{8.14}$$

In the next definition we fix stabilization data at $\mathbf{q}$. Especially the marked points $\vec{w}_{\mathbf{q}}$ are fixed.

Definition 8.9. We define $V(\mathbf{q}, \epsilon_1, \mathcal{B})$ to be the union of isomorphism classes of the following (A) and (B).

(A) An object $(\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, u', \varphi', a'_1, a'_2)$ which satisfies the following.

(1) If Σ'_a is an irreducible component and $\varphi_a : \mathbb{R} \times S^1 \to \Sigma'_a$ is as above, then the composition $u_a \circ \varphi_a$ satisfies the equation

$$\begin{aligned} &\frac{\partial(u_a\circ\varphi_a)}{\partial\tau} + J_{a,\tau}\left(\frac{\partial(u_a\circ\varphi_a)}{\partial t} - \mathfrak{X}_{H^a_{\tau,t}} \circ (u_a\circ\varphi_a)\right) \equiv 0 \\ &\qquad \bmod E((\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c, u', \varphi', a'_1, a'_2); \mathbf{q}; \mathcal{B}). \end{aligned} \tag{8.15}$$

Here $H^a_{\tau,t}$ and $J_{a,\tau}$ are as in Definition 8.3 (5).

(2) If Σ'_{v} is a bubble component, the following equation is satisfied on Σ'_{v}.

$$\overline{\partial}_J u' \equiv 0 \mod E((\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c, u', \varphi', a'_1, a'_2); \mathbf{q}; \mathcal{B}). \tag{8.16}$$

Here the almost complex structure J is as follows. Let $\widehat{\Sigma}'_a$ be the extended mainstream component containing Σ'_{v}. If $a < a'_1$, then $J = J_1$. If $a'_1 < a < a'_2$, then $J = J_2$. If $a'_2 < a$, then $J = J_3$. If $a = a'_1$ and $\varphi_{a'_1}(\tau, t)$ is the root of the tree of sphere components containing Σ'_{v} then $J = J^{21}_\tau$. If $a = a'_2$ and $\varphi_{a'_2}(\tau, t)$ is the root of the tree of sphere components containing Σ'_{v}, then $J = J^{32}_\tau$.

(3) For each $c \in \mathcal{E}(\mathbf{q})$ the additional marked points $\vec{w}'_c$ satisfy the transversal constraint with respect to $\mathbf{p}_c$. (The transversal constraint is defined in the same way as in Definition 6.15.)

(4) The additional marked points $\vec{w}'_{\mathbf{q}}$ satisfy the transversal constraint with respect to $\mathbf{q}$.

(5) $(\mathfrak{Y} \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, u', \varphi', a'_1.a'_2)$ is ϵ_1-close to $\mathbf{q} \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}^{\mathbf{q}}_c \cup \vec{w}_{\mathbf{q}}$.

(B) An object $(\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, u', \varphi', a'_0, T')$ which satisfies the following.

(1) If Σ'_a is an irreducible component and $\varphi_a : \mathbb{R} \times S^1 \to \Sigma_a$ is as above, then the composition $u_a \circ \varphi_a$ satisfies the equation

$$\begin{aligned} &\frac{\partial(u_a \circ \varphi_a)}{\partial \tau} + J_{a,\tau}\left(\frac{\partial(u_a \circ \varphi_a)}{\partial t} - \mathfrak{X}_{H^a_{\tau,t}} \circ (u_a \circ \varphi_a)\right) \equiv 0 \\ &\qquad \mod E((\mathfrak{Y} \cup \bigcup_{c\in\mathcal{B}} \vec{w}'_c, u', \varphi', a'_0, T'); \mathbf{q}; \mathcal{B}). \end{aligned} \tag{8.17}$$

Here $J_{a,\tau}$ and $H^a_{\tau,t}$ are as follows. (See (8.3) and (8.4).)

$$J_{a,\tau} = \begin{cases} J_1 & \text{if } a < a'_0, \\ J^{31}_{\tau,T'} & \text{if } a = a_0, \\ J_3 & \text{if } a > a'_0. \end{cases}$$

$$H^a_{\tau,t} = \begin{cases} H^1_t & \text{if } a < a'_0, \\ H^{31,T'}_{\tau,t} & \text{if } a = a'_0, \\ H^3_t & \text{if } a > a'_0. \end{cases}$$

(2) If $\Sigma'_{\rm v}$ is a bubble component, the following equation is satisfied on $\Sigma'_{\rm v}$:

$$\overline{\partial}_J u' \equiv 0 \mod E((\mathfrak{Y} \cup \vec{w}'_c, u', \varphi', a'_0, T'); \mathbf{q}; \mathcal{B}). \tag{8.18}$$

Here the almost complex structure J is as follows. Let $\widehat{\Sigma}'_a$ be the extended mainstream component containing $\Sigma'_{\rm v}$. If $a < a'_0$, then $J = J_1$. If $a'_0 < a$, then $J = J_3$. If $a = a'_0$ and $\varphi_{a'_0}(\tau, t)$ is the root of the tree of sphere components containing $\Sigma'_{\rm v}$, then $J = J^{31}_{\tau,T'}$.

(3) The same as (3) of (A).

(4) The same as (4) of (A).

(5) $(\mathfrak{Y} \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}'_c \cup \vec{w}'_{\mathbf{q}}, u', \varphi', a'_0, T')$ is ϵ_1-close to $\mathbf{q} \cup \bigcup_{c\in\mathcal{E}(\mathbf{q})} \vec{w}^{\mathbf{q}}_c \cup \vec{w}_{\mathbf{q}}$. In particular, $T' > 1/\epsilon_1$.

The isomorphism among the objects of (A) is defined in the same way as the equivalence relation $\sim_2$ on $\widehat{\mathcal{N}}_\ell(X, \mathcal{J}^{31,\infty}, H^{31,\infty}; \alpha_-, \alpha_+)$, which is the same as Definition 3.7, except we require $\tau_{a_1} = \tau_{a_2} = 0$. The isomorphism among objects of (B) is defined in the same way as in Definition 6.17 (a)(b)(c)(d). (We require T' coincides for two objects to be equivalent.) An object of (A) is never equivalent to an object of (B).

In the same way as in Lemma 4.29, we can prove that $V(\mathbf{q}, \epsilon_1, \mathcal{B})$ is a smooth manifold with boundary and corner if $\epsilon_1 > 0$ and $\epsilon_c > 0$ are small enough. (We use the fact that (8.9) is an open embedding here.)

Then in the same way as in Sect. 4, we can find other data so that $V(\mathbf{q}, \epsilon_1, \mathcal{B})$ together with them is a Kuranishi chart of $\mathbf{q}$. We can also show the existence of coordinate changes. We shrink the Kuranishi neighborhood and discuss in the same way as in Lemmas 4.33, 4.34 and use the exponential decay estimates in the same way as in [16, Chapter 8] to obtain a Kuranishi structure on $\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)$ on a neighborhood of $\mathcal{N}_\ell(X, \mathcal{J}^{31,\infty}, H^{31,\infty}; \alpha_-, \alpha_+)$. We can extend it to the whole $\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}$,

$H^{31,[0,\infty]}; \alpha_-, \alpha_+)$ in the same way as in the proof of Theorem 7.4. We have thus proved Theorem 8.6 (1).

8.3. Proof of Theorem 8.6 (2): Kuranishi structure with outer collar

In this subsection we prove Theorem 8.6 (2).

Remark 8.10. Since the rest of the proof is again a repetition of the construction of previous sections, the readers may safely skip it and go directly to Sect. 9. We provide the details of the proof here for the sake of completeness. The formulas appearing during the proof are lengthy but the argument is a straightforward analogue.

We will modify the Kuranishi structure of $\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)^{\boxplus 1}$ in the collar so that it will be compatible with the fiber product description of its boundary and corners. The way to modify our Kuranishi structure is entirely the same as the proof of Theorem 6.4 (3)(4) except at the boundary corresponding to $T = \infty$. (Namely the subset described in Definition 8.3.) We denote this boundary component $\mathfrak{C}_\infty$ and discuss our construction only on

$$\widehat{S}_m^{\mathfrak{C}_\infty}(\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)^{\boxplus 1}).$$

Let $\mathfrak{A}_r$ be the index set of the critical submanifolds of H^r $(r = 1, 2, 3)$. Let

$$\begin{aligned}
&\alpha_- = \alpha_{1,0}, \alpha_{1,1}, \ldots, \alpha_{1,m_1-1}, \alpha_{1,m_1} \in \mathfrak{A}_1,\\
&\alpha_{2,1}, \ldots, \alpha_{2,m_2-1}, \alpha_{2,m_2} \in \mathfrak{A}_2,\\
&\alpha_{3,1}, \ldots, \alpha_{3,m_3}, \alpha_{3,m_3+1} = \alpha_+ \in \mathfrak{A}_3.
\end{aligned}$$

We put $\vec{\alpha}_1 = (\alpha_{1,0}, \alpha_{1,1}, \ldots, \alpha_{1,m_1-1}, \alpha_{1,m_1})$, $\vec{\alpha}_2 = (\alpha_{2,1}, \ldots, \alpha_{2,m_2-1}, \alpha_{2,m_2})$ and $\vec{\alpha}_3 = (\alpha_{3,1}, \ldots, \alpha_{3,m_3}, \alpha_{3,m_3+1})$.

We consider the fiber product

$$\begin{aligned}
&\mathcal{M}_{\ell_{1,1}}(X, J_1, H^1; \alpha_{1,0}, \alpha_{1,1}) \,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\ell_{1,m_1}}(X, J_1, H^1; \alpha_{1,m_1-1}, \alpha_{1,m_1})\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{N}_{\ell_{(12)}}(X, \mathcal{J}^{21}, H^{21}; \alpha_{1,m_1}, \alpha_{2,1})\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\ell_{2,2}}(X, J_2, H^2; \alpha_{2,1}, \alpha_{2,2}) \,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots\\
&\quad \cdots {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\ell_{2,m_2}}(X, J_2, H^2; \alpha_{2,m_2-1}, \alpha_{2,m_2})\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{N}_{\ell_{(23)}}(X, \mathcal{J}^{32}, H^{32}; \alpha_{2,m_2}, \alpha_{3,1})\\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\ell_{3,2}}(X, J_3, H^3; \alpha_{3,1}, \alpha_{3,2}) \,{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots\\
&\quad \cdots {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\ell_{3,m_3+1}}(X, J_3, H^3; \alpha_{3,m_3}, \alpha_{3,m_3+1}),
\end{aligned} \tag{8.19}$$

which we denote by $\mathcal{N}_{\vec{\ell}_1,\ell_{(12)},\vec{\ell}_2,\ell_{(23)},\vec{\ell}_3}(X, \mathcal{J}^{21}, \mathcal{J}^{32}, H^{21}, H^{32}; \vec{\alpha}_1, \vec{\alpha}_2, \vec{\alpha}_3)$.

We note that a neighborhood of $\widehat{S}_m^{\mathfrak{C}_\infty}(\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)^{\boxplus 1})$ is a union of the direct products

$$\mathcal{N}_{\vec{\ell}_1,\ell_{(12)},\vec{\ell}_2,\ell_{(23)},\vec{\ell}_3}(X, \mathcal{J}^{21}, \mathcal{J}^{32}, H^{21}, H^{32}; \vec{\alpha}_1, \vec{\alpha}_2, \vec{\alpha}_3) \times [-1, 0]^{m-2}$$

for various $\vec{\ell}_1, \ell_{(12)}, \vec{\ell}_2, \ell_{(23)}, \vec{\ell}_3$, $\vec{\alpha}_1, \vec{\alpha}_2, \vec{\alpha}_3$ such that

$$|\vec{\ell}_1| + |\vec{\ell}_2| + |\vec{\ell}_3| + \ell_{(12)} + \ell_{(23)} = \ell,$$
$$\#\vec{\alpha}_1 + \#\vec{\alpha}_2 + \#\vec{\alpha}_3 = m.$$

Note $m_r + 1 = \#\vec{\alpha}_r$ for $r = 1, 3$ and $m_2 = \#\vec{\alpha}_2$.

We will construct a Kuranishi structure for each of

$$\begin{aligned}&\mathcal{N}_{\vec{\ell}_1,\ell_{(12)},\vec{\ell}_2,\ell_{(23)},\vec{\ell}_3}(X, \mathcal{J}^{21}, \mathcal{J}^{32}, H^{21}, H^{32}; \vec{\alpha}_1, \vec{\alpha}_2, \vec{\alpha}_3)^+\\ &\quad= \mathcal{N}_{\vec{\ell}_1,\ell_{(12)},\vec{\ell}_2,\ell_{(23)},\vec{\ell}_3}(X, \mathcal{J}^{21}, \mathcal{J}^{32}, H^{21}, H^{32}; \vec{\alpha}_1, \vec{\alpha}_2, \vec{\alpha}_3)\\ &\qquad\times [-1, 0]^{m_1} \times [-1, 0]^{m_2} \times [-1, 0]^{m_3}.\end{aligned}$$

Let $A_r \sqcup B_r \sqcup C_r = \underline{m_r}$ for $r = 1, 2, 3$. It induces $\mathcal{I}_{A_r,B_r,C_r} : [-1, 0]^{b_r} \to [-1, 0]^{m_r}$ by (5.2). We will formulate compatibility conditions below (Conditions 8.11–8.13), which describe the restriction of the Kuranishi structure

$$\widehat{\mathcal{U}}_{\vec{\ell}_1,\ell_{(12)},\vec{\ell}_2,\ell_{(23)},\vec{\ell}_3}(X, \mathcal{J}^{21}, \mathcal{J}^{32}, H^{21}, H^{32}; \vec{\alpha}_1, \vec{\alpha}_2, \vec{\alpha}_3) \tag{8.20}$$

of the product space $\mathcal{N}_{\vec{\ell}_1,\ell_{(12)},\vec{\ell}_2,\ell_{(23)},\vec{\ell}_3}(X, \mathcal{J}^{21}, \mathcal{J}^{32}, H^{21}, H^{32}; \vec{\alpha}_1, \vec{\alpha}_2, \vec{\alpha}_3)^+$ to the image of the embedding $\mathrm{id} \times \mathcal{I}_{A_1,B_1,C_1} \times \mathcal{I}_{A_2,B_2,C_2} \times \mathcal{I}_{A_3,B_3,C_3}$.

We put $A_r = \{i(A_r, 1), \ldots, i(A_r, a_r)\}$ with

$$i(A_r, 1) < i(A_r, 2) < \cdots < i(A_r, a_r - 1) < i(A_r, a_r).$$

We define $C'_j(A_r)$, $C_j(A_r)$, $\vec{\alpha}_{r,A_r,j}$, $\vec{\alpha}_{r,A_r,j,C_r}$ in the same way as in Notation 6.19 (1)(2)(3). Here we put $i(A_r, a_r + 1) = m_r$, $i(r, A_r, 0) = 1$, except $i(A_1, 0) = 0$, $i(A_3, a_3+1) = m_3+1$. (Compare Remark 6.18 and Notation 6.19 (1).)

Case 1: We first consider the case $A_2 \neq \emptyset$. We consider the fiber product

$$\begin{aligned}&\mathcal{M}_{\vec{\ell}_{1,A_1,1}}(X, J_1, H^1; \vec{\alpha}_{1,A_1,1})\\
&\quad {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{1,A_1,2}}(X, J_1, H^1; \vec{\alpha}_{1,A_1,2}) {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \cdots\\
&\quad {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{1,A_1,j}}(X, J_1, H^1; \vec{\alpha}_{1,A_1,j}) {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \cdots\\
&\quad {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{1,A_1,a_1}}(X, J_1, H^1; \vec{\alpha}_{1,A_1,a_1})\\
&\quad {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{N}_{\vec{\ell}_{1,A_1,a_1+1},\ell_{(12)},\vec{\ell}_{2,A_2,1}}(X, \mathcal{J}^{21}, H^{21}; \vec{\alpha}_{1,A_1,a_1+1}.\vec{\alpha}_{2,A_2,1})\\
&\quad {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{2,A_2,2}}(X, J_2, H^2; \vec{\alpha}_{2,A_2,2}) {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \cdots\\
&\quad {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{2,A_2,j}}(X, J_2, H^2; \vec{\alpha}_{2,A_2,j}) {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \cdots\\
&\quad {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{2,A_2,a_2}}(X, J_2, H^2; \vec{\alpha}_{2,A_2,a_2})\\
&\quad {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{N}_{\vec{\ell}_{2,A_2,a_2+1},\ell_{(23)},\vec{\ell}_{3,A_3,1}}(X, \mathcal{J}^{32}, H^{32}; \vec{\alpha}_{2,A_2,a_2+1}, \vec{\alpha}_{3,A_3,1})\\
&\quad {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{3,A_3,2}}(X, J_3, H^3; \vec{\alpha}_{3,A_3,2}) {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \cdots\\
&{}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{3,A_3,j}}(X, J_3, H^3; \vec{\alpha}_{3,A_3,j}) {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \cdots\\
&\quad {}_{\mathrm{ev}_+} \times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{3,A_3,a_3+1}}(X, J_3, H^3; \vec{\alpha}_{3,A_3,a_3+1}).\end{aligned} \tag{8.21}$$

Here $\vec{\ell}_{r,A_r,j} = (\ell_{r,i(A_r,j-1)+1}, \ldots, \ell_{r,i(A_r,j)})$.

We define $m_j(r, A_r)$ and $m_j(r, A_r, C_r)$ as in Notation 6.19 (5). Then (6.19) holds. We define $\vec{\ell}_{r,A_r,j,C_r}$ in the same way as in Notation 6.19 (6). We put

$$\begin{aligned}\ell'_{(rr+1)} = \ell_{r,i(A_r,a_r)+k_{m_{a_r}(r,A_r,C_r)}+1} \cdots + \ell_{r,m_r} + \\ + \ell_{(rr+1)} + \ell_{r+1,2} + \cdots + \ell_{r+1,\min(i(A_{r+1},1),i(B_{r+1},1))}.\end{aligned} \tag{8.22}$$

We remark that in Proposition 5.5 we determined the Kuranishi structure on

$$\begin{aligned}&\mathcal{M}_{\vec{\ell}_{r,A_r,j},C_r}(X, J_r, H^r; \vec{\alpha}_{r,A_r,j,C_r})^+ \\ &= \mathcal{M}_{\vec{\ell}_{r,A_r,j},C_r}(X, J_r, H^r; \vec{\alpha}_{r,A_r,j,C_r}) \times [-1,0]^{m_j(r,A_r,C_r)-1}.\end{aligned} \tag{8.23}$$

By construction we have

$$\mathcal{M}_{\vec{\ell}_{r,A_r,j}}(X, J_r, H^r; \vec{\alpha}_{r,A_r,j}) \subseteq \widehat{S}_{c_j(r,A_r)}(\mathcal{M}_{\vec{\ell}_{r,A_r,j},C_r}(X, J_r, H^r; \vec{\alpha}_{r,A_r,j,C_r}))$$

and the left hand side is a factor in (8.21).

By restriction it determines a Kuranishi structure of (8.21) times $[-1,0]^*$ except two of the factors

$$\mathcal{N}_{\vec{\ell}_{r,A_r,a_r+1},\ell_{(rr+1)},\vec{\ell}_{r+1,A_{r+1},1}}(X, \mathcal{J}^{r+1r}, H^{r+1r};\ \vec{\alpha}_{r,A_r,a_r+1}, \vec{\alpha}_{r+1,A_{r+1},1}). \tag{8.24}$$

By construction, we can easily show that (8.24) is a component of

$$\begin{aligned}&\widehat{S}_{c_{a_r+1}(r,A_r)+c_1(r+1,A_{r+1})}(\mathcal{N}_{\vec{\ell}_{r,A_r,a_r+1,C_r},\ell'_{(rr+1)},\vec{\ell}_{r+1,A_{r+1},1,C_{r+1}}} \\ &(X, \mathcal{J}^{r+1r}, H^{r+1r}; \vec{\alpha}_{r,A_r,a_r+1,C_r}, \vec{\alpha}_{r+1,A_{r+1},1,C_{r+1}}).\end{aligned} \tag{8.25}$$

Note that we defined the Kuranishi structure on

$$\begin{aligned}&\mathcal{N}_{\vec{\ell}_{r,A_r,a_r+1,C_r},\ell'_{(rr+1)},\vec{\ell}_{r+1,A_{r+1},1,C_{r+1}}}(X, \mathcal{J}^{r+1r}, \\ &H^{r+1r}; \vec{\alpha}_{r,A_r,a_r+1,C_r}, \vec{\alpha}_{r+1,A_{r+1},1,C_{r+1}})^+\end{aligned}$$

during the construction of the morphism $\mathfrak{N}_{r+1r}$ in Proposition 6.21. We write it as

$$\begin{aligned}&\widehat{\mathcal{U}}_{\vec{\ell}_{r,A_r,a_r+1,C_r},\ell'_{(rr+1)},\vec{\ell}_{r+1,A_{r+1},1,C_{r+1}}} \\ &(X, \mathcal{J}^{r+1r}, H^{r+1r}; \vec{\alpha}_{r,A_r,a_r+1,C_r}, \vec{\alpha}_{r+1,A_{r+1},1,C_{r+1}}).\end{aligned} \tag{8.26}$$

Condition 8.11. In the case $A_2 \neq \emptyset$ we require the restriction of the Kuranishi structure (8.20) to the image of $\mathrm{id} \times \mathcal{I}_{A_1,B_1,C_1} \times \mathcal{I}_{A_2,B_2,C_2} \times \mathcal{I}_{A_3,B_3,C_3}$ is the fiber product of the following Kuranishi structures. (Here we use the fiber bundle description (8.21).)

(1) The Kuranishi structure on $\mathcal{M}_{\vec{\ell}_{r,A_r,j}}(X, J_r, H^r; \vec{\alpha}_{r,A_r,j}) \times [-1,0]^{m_j(r,A_r,C_r)-1}$ which is a restriction of those on (8.23) produced in Proposition 5.5.
(2) The Kuranishi structure on (8.24)$\times[-1,0]^{m_{a_r+1}(r,A_r,C_r)} + m_1(r+1, A_{r+1}, C_{r+1}) - 2$ which is a restriction of those on (8.26) produced in Proposition 6.21.

Case 2: We next consider the case $A_2 = \emptyset$ but $B_2 \neq \emptyset$.

We consider the fiber product

$$\begin{aligned}
&\mathcal{M}_{\vec{\ell}_{1,A_1,1}}(X, J_1, H^1; \vec{\alpha}_{1,A_1,1}) \\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{1,A_1,2}}(X, J_1, H^1; \vec{\alpha}_{1,A_1,2})\, {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots \\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{1,A_1,j}}(X, J_1, H^1; \vec{\alpha}_{1,A_1,j})\, {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots \\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{1,A_1,a_1}}(X, J_1, H^1; \vec{\alpha}_{1,A_1,a_1}) \\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{N}_{\vec{\ell}_{1,A_1,a_1+1},\ell_{(12)},\vec{\ell}_2,\ell_{(23)},\vec{\ell}_{3,A_3,1}}(X, \mathcal{J}^{21}, \mathcal{J}^{32}, H^{21}, H^{32} \\
&\quad ; \vec{\alpha}_{1,A_1,a_1+1}, \vec{\alpha}_2, \vec{\alpha}_{3,A_3,1}) \\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{3,A_3,2}}(X, J_3, H^3; \vec{\alpha}_{3,A_3,2})\, {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots \\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{3,A_3,j}}(X, J_3, H^3; \vec{\alpha}_{3,A_3,j})\, {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots \\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\vec{\ell}_{3,A_3,a_3+1}}(X, J_3, H^3; \vec{\alpha}_{3,A_3,a_3+1}).
\end{aligned} \tag{8.27}$$

We define $\vec{\alpha}_{2,C_2}$ by removing $\alpha_{2,i}$, $i \in C_2$ from $\vec{\alpha}_2$.

We define $\vec{\ell}_{2,C_2}$ as follows. We put $\vec{\alpha}_{2,C_2} = \{\alpha_{2,k_s} \mid s = 0, \ldots, m_{2,C_2}\}$, $k_0 < k_1 < \cdots < k_{m_{2,C_2}}$. (Here $m_{2,C_2} = \#\vec{\alpha}_{2,C_2} - 1$.) Note if $i \in (k_s, k_{s+1})_{\mathbb{Z}}$, then $i \in C_2$. We put

$$\ell_{2,C_2,s} = \ell_{2,k_{s-1}+1} + \cdots + \ell_{2,k_s}$$

and $\vec{\ell}_{2,C_2} = (\ell_{2,C_2,1}, \ldots, \ell_{2,C_2,m_{2,C_2}})$. Now we consider the factor

$$\begin{aligned}
&\mathcal{N}_{\vec{\ell}_{1,A_1,a_1+1},\ell_{(12)},\vec{\ell}_2,\ell_{(23)},\vec{\ell}_{3,A_3,1}}(X, \mathcal{J}^{21}, \mathcal{J}^{32}, H^{21}, H^{32} \\
&\quad ; (\vec{\alpha}_{1,A_1,a_1+1}, \vec{\alpha}_2, \vec{\alpha}_{3,A_3,1}))
\end{aligned} \tag{8.28}$$

in (8.27) and lies in the corner of

$$\begin{aligned}
\mathcal{N}_{\vec{\ell}_{1,A_1,a_1+1C_1},\ell'_{(12)},\vec{\ell}_{2,C_2},\ell'_{(23)},\vec{\ell}_{3,A_3,1,C_3}}&(X, \mathcal{J}^{21}, \mathcal{J}^{32}, H^{21}, H^{32}; \\
&\vec{\alpha}_{1,A_1,a_1+1,C_1}, \vec{\alpha}_{2,C_2}, \vec{\alpha}_{3,A_3,1,C_3}),
\end{aligned} \tag{8.29}$$

where

$$\ell'_{(12)} = \ell_{1,i(A_1,a_1)+k_{m_{a_1}(1,A_1,C_1)}+1} + \cdots + \ell_{1,m_1} + \ell_{(12)} + \ell_{2,1} + \cdots + \ell_{2,i(A_2,1)},$$
$$\ell'_{(23)} = \ell_{2,i(A_2,a_2)+k_{m_{a_2}(2,A_2,C_2)}+2} + \cdots + \ell_{2,m_2} + \ell_{(23)} + \ell_{3,1} + \cdots + \ell_{3,i(A_3,1)}.$$

We have a Kuranishi structure

$$\begin{aligned}
&\widehat{\mathcal{U}}_{\vec{\ell}_{1,A_1,a_1+1,C_1},\ell'_{(12)},\vec{\ell}_{2,C_2},\ell'_{(23)},\vec{\ell}_{3,A_3,1,C_3}}(X, \mathcal{J}^{21}, \mathcal{J}^{32}, H^{21}, H^{32}; \\
&\quad \vec{\alpha}_{1,A_1,a_1+1,C_1}, \vec{\alpha}_{2,C_2}, \vec{\alpha}_{3,A_3,1,C_3})
\end{aligned} \tag{8.30}$$

on (8.28)$\times[-1,0]^*$. (More precisely, we are during the process of producing it. Here $* = \#\vec{\alpha}_{1,A_1,a_1+1,C_1} + \#\vec{\alpha}_{2,C_2} + \#\vec{\alpha}_{3,A_3,1,C_3} - 2$.)

Condition 8.12. In the case $A_2 = \emptyset$, $B_2 \neq \emptyset$, we require the restriction of the Kuranishi structure (8.20) to the image of $\mathrm{id}\times\mathcal{I}_{A_1,B_1,C_1}\times\mathcal{I}_{A_2,B_2,C_2}\times\mathcal{I}_{A_3,B_3,C_3}$ is the fiber product of the following Kuranishi structures. (We use the fiber product description (8.27).)

(1) The Kuranishi structure on $\mathcal{M}_{\vec{\ell}_{r,A_r,j}}(X, J_r, H^r; \vec{\alpha}_{r,A_r,j}) \times [-1,0]^{m_j(r,A_r,C_r)-1}$ which is a restriction of those on (8.23) produced in Proposition 5.5.
(2) The Kuranishi structure on (8.28)$\times[-1,0]^*$ ($* = \#\vec{\alpha}_{1,A_1,a_1+1,C_1} + \#\vec{\alpha}_{2,C_2} + \#\vec{\alpha}_{3,A_3,1,C_3} - 2$) which is a restriction of (8.30).

Case 3: We finally consider the case when $A_2 = B_2 = \emptyset$.

In this case, using the notation above, we have $\vec{\alpha}_{2,C_2} = \emptyset$. We require the compatibility with the Kuranishi structures on the part $T < \infty$ in this case, as follows.

We denote the fiber product:

$$\begin{aligned}
&\mathcal{M}_{\ell_{1,1}}(X, J_1, H^1; \alpha_{1,0}, \alpha_{1,1})\ {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots \\
&\quad \cdots {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\ell_{1,m_1}}(X, J_1, H^1; \alpha_{1,m_1-1}, \alpha_{1,m_1}) \\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{N}_{\ell'}(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_{1,m_1}, \alpha_{3,1}) \\
&\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\ell_{3,2}}(X, J_3, H^3; \alpha_{3,1}, \alpha_{3,2})\ {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots \\
&\quad \cdots {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\ell_{3,m_3+1}}(X, J_3, H^3; \alpha_{3,m_3}, \alpha_{3,m_3+1})
\end{aligned} \tag{8.31}$$

by $\mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_3}(X, H^{31,[0,\infty]}, \mathcal{J}^{31,[0,\infty]}; \vec{\alpha}_1, \vec{\alpha}_3)$. Note this is also a component of the corner of $\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)$ if $|\vec{\ell}_1| + \ell' + |\vec{\ell}_2| = \ell$. So we are during the process of constructing Kuranishi structures on $\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)^+$, which is the direct product of $\mathcal{N}_\ell(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)$ with $[-1,0]^*$. ($* = \#\vec{\alpha}_1 + \#\vec{\alpha}_3 - 2$.) Let us denote by $\widehat{\mathcal{U}}_{\vec{\ell}_1,\ell',\vec{\ell}_3}(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \vec{\alpha}_1, \vec{\alpha}_3)$ the Kuranishi structure on it.

We put

$$\ell'_2 = \ell - |\vec{\ell}_{1,C_1}| - |\vec{\ell}_{3,C_3}|.$$

We then observe that (8.28) lies in the corner of

$$\mathcal{N}_{\vec{\ell}_{1,A_1,a_1+1,C_1},\ell'_2,\vec{\ell}_{3,A_3,1,C_3}}(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \vec{\alpha}_{1,A_1,a_1+1,C_1}, \vec{\alpha}_{3,A_3,1,C_3}). \tag{8.32}$$

Condition 8.13. In the case $A_2 = B_2 = \emptyset$, we require the restriction of the Kuranishi structure (8.20) to the image of $\mathrm{id} \times \mathcal{I}_{A_1,B_1,C_1} \times \mathcal{I}_{A_2,B_2,C_2} \times \mathcal{I}_{A_3,B_3,C_3}$ is the fiber product of the following Kuranishi structures. (We use the fiber product description (8.27).)

(1) The Kuranishi structure on $\mathcal{M}_{\vec{\ell}_{r,A_r,j}}(X, J_r, H^r; \vec{\alpha}_{r,A_r,j}) \times [-1,0]^{m_j(r,A_r,C_r)-1}$ which is a restriction of those on (8.23) produced in Proposition 5.5.
(2) The Kuranishi structure on (8.28)$\times[-1,0]^*$ which is a restriction of the Kuranishi structure

$$\widehat{\mathcal{U}}_{\vec{\ell}_{1,A_1,a_1+1,C_1},\ell'_2,\vec{\ell}_{3,A_3,1,C_3}}(X, H^{31,[0,\infty]}, \mathcal{J}^{31,[0,\infty]}; \vec{\alpha}_{1,A_1,a_1+1,C_1}, \vec{\alpha}_{3,A_3,1,C_3}),$$

on (8.32)$\times[-1,0]^*$. Here $* = \#\vec{\alpha}_{1,A_1,a_1+1,C_1} + \#\vec{\alpha}_{3,A_3,1,C_3} - 2$.

We have thus described the conditions we require at $T = \infty$.

There are similar compatibility conditions at $T < \infty$ and $T = 0$. We require such conditions to the Kuranishi structure $\widehat{\mathcal{U}}_{\vec{\ell}_1,\ell',\vec{\ell}_3}(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \vec{\alpha}_1, \vec{\alpha}_3)$. We omit the detailed description of this compatibility condition since it is the same as that for the case of Condition 6.20 and Proposition 6.21.

Proposition 8.14. *There exists a K-system*

$$\{(\mathcal{N}_{\vec{\ell}_1,\ell_{(12)},\vec{\ell}_2,\ell_{(23)},\vec{\ell}_3}(X, \mathcal{J}^{21}, \mathcal{J}^{32}, H^{21}, H^{32}; \vec{\alpha}_1, \vec{\alpha}_2, \vec{\alpha}_3)^+,$$
$$\widehat{\mathcal{U}}_{\vec{\ell}_1,\ell_{(12)},\vec{\ell}_2,\ell_{(23)},\vec{\ell}_3}(X, \mathcal{J}^{21}, \mathcal{J}^{32}, H^{21}, H^{32}; \vec{\alpha}_1, \vec{\alpha}_2, \vec{\alpha}_3))\}$$

whose Kuranishi structure is as in (8.20) with the following properties.

(1) *They satisfy Conditions 8.11–8.13.*

(2) *There exists a K-system*

$$\{(\mathcal{N}_{\vec{\ell}_1,\ell',\vec{\ell}_3}(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \vec{\alpha}_1, \vec{\alpha}_3),$$
$$\widehat{\mathcal{U}}_{\vec{\ell}_1,\ell',\vec{\ell}_3}(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \vec{\alpha}_1, \vec{\alpha}_3))\}$$

which satisfies a compatibility condition similar to Condition 6.20.

(3) *A similar compatibility condition is satisfied at $T = 0$.*

(4) *Let $\mathfrak{C}$ be the union of the boundary components of the underlying topological space $\mathcal{N}_{\vec{\ell}_1,\ell_{(12)},\vec{\ell}_2,\ell_{(23)},\vec{\ell}_3}(X, \mathcal{J}^{21}, \mathcal{J}^{32}, H^{21}, H^{32}; \vec{\alpha}_1, \vec{\alpha}_2, \vec{\alpha}_3)^+$ corresponding to $\partial([0,1]^*)$. Then* (8.20) *is $\mathfrak{C}$-collared.*

(5) *If $\vec{\alpha}_1 = \{\alpha_-\}$ and $\vec{\alpha}_3 = \{\alpha_+\}$, then $\widehat{\mathcal{U}}_{\emptyset,\ell,\emptyset}(X, \mathcal{J}^{31,[0,\infty]}, H^{31,[0,\infty]}; \alpha_-, \alpha_+)$ coincides with the Kuranishi structure we produced in Theorem* 8.6 *(1).*

Proof. The proof is entirely similar to the proof of Proposition 6.21 etc. □

We can use Proposition 8.14 to complete the proof of Theorem 8.6 (2) in the same way as before. (We use smoothing of corners and [23, Lemma 18.40] to show that at $T = \infty$ we get the Kuranishi structure used to define the composition.) □

The proof of Theorem 8.2 is complete. □

9. Well-definedness of Hamiltonian Floer cohomology

We now use the results of the previous sections to conclude the well-definedness of the Floer cohomology of a periodic Hamiltonian system. Namely we prove the next theorems.

Theorem 9.1. *Let $H : X \times S^1 \to \mathbb{R}$ be a smooth function such that $\mathrm{Per}(H)$ is Morse–Bott non-degenerate in the sense of Condition 2.1. Then we can associate the Floer cohomology $HF(X, H; \Lambda_{0,\mathrm{nov}})$ which is independent of various choices involved in the definition.*

Recall $\Lambda_{0,\mathrm{nov}}$ is the Novikov ring defined in (1.1). We define the Novikov field Λ_{nov} as its field of fractions by allowing λ_i to be negative.

Theorem 9.2. *Let $H^r : X \times S^1 \to \mathbb{R}$ $(r = 1, 2)$ be as in Theorem 9.1. Then the Floer cohomologies $HF(X, H^r; \Lambda_{\rm nov}) = HF(X, H^r; \Lambda_{0,\rm nov}) \otimes_{\Lambda_{0,\rm nov}} \Lambda_{\rm nov}$ $(r = 1, 2)$ over the Novikov field $\Lambda_{\rm nov}$ satisfy*

$$HF(X, H^1; \Lambda_{\rm nov}) \cong HF(X, H^2; \Lambda_{\rm nov}).$$

Remark 9.3. Using Remark 6.22, we can prove the Lipschitz continuity of torsion exponent of the Floer cohomology over $\Lambda_{0,\rm nov}$ with respect to the distance

$$d(H^1, H^2) = \int_{t \in S^1} \sup_{x \in X} |H^1(t, x) - H^2(t, x)| \mathrm{d}t$$

on the set of the Hamiltonians. We omit the discussion about it since we can derive it from a similar result on the Floer cohomology of Lagrangian intersection. (See [14].) We can also use Remark 6.22 to derive more precise results about the filtration of the Floer cohomology of a periodic Hamiltonian system. We discussed it in detail in [22].

The proofs of Theorems 9.1, 9.2 occupy the main part of this section.

Situation 9.4. (1) Let $H : X \times S^1 \to \mathbb{R}$ be a smooth function such that $\mathrm{Per}(H)$ is Morse–Bott non-degenerate. We write $H = H^1$. We define $H^{11} : X \times \mathbb{R} \times S^1 \to \mathbb{R}$ by $H^{11}(x, \tau, t) = H^1(x, t)$.

(2) Let J and J' be two choices of tame almost complex structures. We define $\mathcal{J} = \{J_\tau\}$ such that $J_\tau = J$ for $\tau < -1$ and $J_\tau = J'$ for $\tau > 1$. □

Construction 9.5. Suppose we are in Situation 9.4.

(1) We use H and J and apply Theorem 2.9 (1) to obtain a linear K-system $\mathcal{F}_X(H, J)$ whose space of connecting orbits is $\mathcal{M}((X, J), H; \alpha_-, \alpha_+)$. (We made choices to define it.)

(2) We then apply [23, Theorem 16.9] to obtain a chain complex (we made choices to define it) and then use [23, Definition 16.12] to obtain a cochain complex over the universal Novikov ring $\Lambda_{0,\rm nov}$, which we denote by $CF((X, J), H; \Lambda_{0,\rm nov})$. By [23, Theorem 16.9 (2)] the cohomology of this cochain complex is independent of the choices made in Item (2). We denote this cohomology group as

$$HF((X, J), H; \Lambda_{0,\rm nov}).$$

(3) We start from H and J' and make choices in Item (2) to obtain

$$HF((X, J'), H; \Lambda_{0,\rm nov}).$$

However, [23, Theorem 16.9 (2)] does not imply that $HF((X, J), H; \Lambda_{0,\rm nov})$ is independent of the choices made in Item (1). The statement of Theorem 9.1 is independence of the Floer cohomology (over $\Lambda_{0,\rm nov}$) of the choices made in Item (2) as well as the almost complex structure J. The statement of Theorem 9.2 is independence of the Floer cohomology (over $\Lambda_{\rm nov}$) of the choices made in Item (1) as well as Item (2).

Construction 9.6. Suppose we are in Situation 9.4. We also assume that we have made all the choices involved in Construction 9.5 (1)(2)(3).

We apply Theorem 6.4 to obtain a morphism from $\mathcal{F}_X(H,J)$ to $\mathcal{F}_X(H,J')$. Here $\mathcal{F}_X(H,J)$ is defined in Situation 9.4 (1) and $\mathcal{F}_X(H,J')$ is defined in Situation 9.4 (3). We denote this morphism by $\mathfrak{N}_{11}(\mathcal{J},H^{11})$.

We make choices to define $\mathfrak{N}_{11}(\mathcal{J},H^{11})$. The interpolation space of $\mathfrak{N}_{11}(\mathcal{J},H^{11})$ is $\mathcal{N}(X,\mathcal{J},H^{11};\alpha_-,\alpha_+)^{\boxplus 1}$.

Lemma 9.7. *We can make the choice in Construction 9.6 so that $\mathfrak{N}_{11}(\mathcal{J},H^{11})$ is a morphism of energy loss 0. Namely we have*

(1) $\mathcal{N}(X,\mathcal{J},H^{11};\alpha_-,\alpha_+)^{\boxplus 1}=\emptyset$ *if* $E(\alpha_-)>E(\alpha_+)$ *or* $E(\alpha_-)=E(\alpha_+)$, $\alpha_-\neq\alpha_+$.
(2) $\mathcal{N}(X,\mathcal{J},H^{11};\alpha,\alpha)^{\boxplus 1}=R_\alpha$. *The evaluation maps on it are the identity maps.*

Proof. Using the fact that $H^{11}_{\tau,t}=H^1_t$ and is τ independent, (1) is an immediate consequence of Remark 6.22. To prove (2) we first observe that $\mathcal{N}(X,\mathcal{J},H^{11};\alpha,\alpha)=R_\alpha$ set-theoretically. In fact, $\mathcal{N}(X,\mathcal{J},H^{11};\alpha,\alpha)$ consists of $((\Sigma,\vec{z}_\pm),\varphi,u)$ where $\Sigma=S^2$ (without bubbles) and $u(\varphi(\tau,t))=\gamma(t)$ with $\gamma\in R_\alpha$. We also remark that this moduli space is Fredholm regular. Therefore, we can make our choice so that the obstruction bundle is 0 in this particular case. (Since $\partial\mathcal{N}(X,\mathcal{J},H^{11};\alpha,\alpha)=\emptyset$, we do not need to study the compatibility with the choices made in Construction 9.5 (1)(3).) Item (2) holds for this choice. □

Proof of Theorem 9.1. We now apply [23, Theorem 16.31] to the morphism obtained by taking the choice made in Lemma 9.7. We then obtain a chain map

$$\psi_{\mathcal{J},H^{11}}: CF((X,J),H;\Lambda_{0,\mathrm{nov}})\to CF((X,J'),H;\Lambda_{0,\mathrm{nov}}). \tag{9.1}$$

We note that

$$CF((X,J),H;\Lambda_{0,\mathrm{nov}})=CF((X,J'),H;\Lambda_{0,\mathrm{nov}})$$

as $\Lambda_{0,\mathrm{nov}}$-modules. (Floer's boundary operator may be different, however.)

Using the fact that energy loss of $\mathfrak{N}_{11}(\mathcal{J},H^{11})$ is zero, especially Lemma 9.7 (2), we find that

$$\psi_{\mathcal{J},H^{11}}\equiv \mathrm{id}\mod T^\epsilon\Lambda_{0,\mathrm{nov}}$$

for some $\epsilon>0$. Therefore, $\psi_{\mathcal{J},H^{11}}$ has an inverse, which automatically becomes a chain map. Therefore, we have

$$HF((X,J),H;\Lambda_{0,\mathrm{nov}})\cong HF((X,J'),H;\Lambda_{0,\mathrm{nov}}),$$

as required. □

Proof of Theorem 9.2

Situation 9.8. (1) Let $H^r: X\times S^1\to\mathbb{R}$ $(r=1,2)$ be smooth functions such that $\mathrm{Per}(H^r)$ are Morse–Bott non-degenerate. Let J^r $(r=1,2)$ be tame almost complex structures.

(2) For $r = 1, 2$ we make choices as in Construction 9.5 to define chain complexes $CF((X, J^r), H^r; \Lambda_{0,\mathrm{nov}})$ with $\Lambda_{0,\mathrm{nov}}$ coefficients, and their cohomology groups $HF((X, J^r), H^r; \Lambda_{0,\mathrm{nov}})$.

(3) We take $H^{21} : X \times \mathbb{R} \times S^1 \to \mathbb{R}$ and $\mathcal{J}^{21}$ as in Situation 6.1. We exchange the role of H^1, J^1 and H^2, J^2 and take H^{12}, $\mathcal{J}^{12}$. □

We now apply Theorem 6.4 to $H^{21}, \mathcal{J}^{21}$ and obtain a morphism

$$\mathfrak{N}_{21} : \mathcal{F}_X(H^1, J^1) \to \mathcal{F}_X(H^2, J^2).$$

We also apply Theorem 6.4 to $H^{12}, \mathcal{J}^{12}$ and obtain a morphism

$$\mathfrak{N}_{12} : \mathcal{F}_X(H^2, J^2) \to \mathcal{F}_X(H^1, J^1).$$

We apply [23, Theorem 16.31 (1)] to obtain chain maps

$$\psi_{12} : CF((X, J^1), H^1; \Lambda_{\mathrm{nov}}) \to CF((X, J^2), H^2; \Lambda_{\mathrm{nov}})$$

and

$$\psi_{21} : CF((X, J^2), H^2; \Lambda_{\mathrm{nov}}) \to CF((X, J^1), H^1; \Lambda_{\mathrm{nov}}).$$

We consider the particular case of $J = J' = J^1$ in Situation 9.8. Then we obtain

$$\mathfrak{N}_{11} : \mathcal{F}_X(H^1, J^1) \to \mathcal{F}_X(H^1, J^1).$$

In a similar way we obtain

$$\mathfrak{N}_{22} : \mathcal{F}_X(H^2, J^2) \to \mathcal{F}_X(H^2, J^2).$$

We denote the chain map (9.1) in this case by

$$\psi_{0,rr} : CF((X, J^r), H^r; \Lambda_{0,\mathrm{nov}}) \to CF((X, J^r), H^r; \Lambda_{0,\mathrm{nov}}). \tag{9.2}$$

This is a chain isomorphism in the proof of Theorem 9.1 by Lemma 9.7. We change the coefficient ring to Λ_{nov} by taking tensor product and obtain

$$\psi_{rr} : CF((X, J^r), H^r; \Lambda_{\mathrm{nov}}) \to CF((X, J^r), H^r; \Lambda_{\mathrm{nov}}). \tag{9.3}$$

This is also a chain isomorphism.

Lemma 9.9. *The composition $\mathfrak{N}_{12} \circ \mathfrak{N}_{21}$ (resp. $\mathfrak{N}_{21} \circ \mathfrak{N}_{12}$) is homotopic to $\mathfrak{N}_{11}$ (resp. $\mathfrak{N}_{22}$).*

Proof. This is a special case of Theorem 8.2. □

Lemma 9.9 and [23, Theorem 16.31 (3)] imply that $\psi_{21} \circ \psi_{12}$ (resp. $\psi_{12} \circ \psi_{21}$) is chain homotopic to ψ_{22} (resp. ψ_{11}). Since ψ_{22}, ψ_{11} are chain homotopy equivalences, it implies that ψ_{21} is a chain homotopy equivalence. Therefore,

$$\psi_{21*} : HF((X, J^1), H^1; \Lambda_{\mathrm{nov}}) \cong HF((X, J^2), H^2; \Lambda_{\mathrm{nov}}) \tag{9.4}$$

as required. □

We will discuss the well-definedness of the isomorphisms in Theorems 9.1 and 9.2 of various choices more precisely below.

In Situation 9.8 we obtain a chain map ψ_{21} (of Λ_{nov} coefficients) which induces an isomorphism of the Floer cohomology (9.4).

In Situation 9.4 we obtain a chain map between two chain complexes of $\Lambda_{0,\mathrm{nov}}$ coefficients, which we denote by

$$\psi_{0;J'J} : CF((X,J),H;\Lambda_{0,\mathrm{nov}}) \to CF((X,J'),H;\Lambda_{0,\mathrm{nov}}). \tag{9.5}$$

This is nothing but the chain map (9.1).[34] It induces an isomorphism $\psi_{0;J'J*}$ on cohomology groups.

Finally when we use the same almost complex structure J and the other choices for the domain and the target, we obtain (9.2) and (9.3).

Lemma 9.10. *The maps ψ_{21*}, $\psi_{0;J'J*}$ and ψ_{rr*} are independent of various choices involved in the construction.*

Proof. We consider the two choices to define the morphism $\mathfrak{N}_{21}$. We denote by $\mathfrak{N}^a_{21}$ and $\mathfrak{N}^b_{21}$ the morphism obtained by those two choices. By Lemma 7.2 and Theorem 7.4, two morphisms $\mathfrak{N}^a_{21}$ and $\mathfrak{N}^b_{21}$ are homotopic each other. The isomorphism (9.4) is induced from $\mathfrak{N}_{21}$ using [23, Theorem 16.31 (1)]. We can use [23, Theorem 16.31 (2)] that the homomorphism ψ^a_{12} induced from $\mathfrak{N}^a_{21}$ is chain homotopic to the homomorphism ψ^b_{12} induced from $\mathfrak{N}^b_{21}$. This implies the independence of ψ_{21*} of the choices.

The independence of $\psi_{0;J'J*}$ and ψ_{rr*} are proved in the same way. We only need to note that in that situation we can take the homotopy to be of energy loss zero. □

This lemma together with the next lemma imply that the group $HF(X,H;\Lambda_{\mathrm{nov}})$ is independent of H, J and other choices up to *canonical* isomorphism.

Lemma 9.11. (1) *We consider H^r, J^r for $r = 1,2,3$ and isomorphisms ψ_{21*}, ψ_{32*}, ψ_{31*} as above. Then*

$$\psi_{31*} = \psi_{32*} \circ \psi_{21*}.$$

(2) *We fix H and take three choices J_1, J_2 and J_3 of almost complex structures as well as other choices in Construction 9.5. Then we have*

$$\psi_{0,J_3J_1*} = \psi_{0,J_3J_2*} \circ \psi_{0,J_2J_1*}.$$

(3) *$\psi_{0,rr}$ induces the identity map*

$$\psi_{0,rr*} : HF((X,J^r),H^r;\Lambda_{0,\mathrm{nov}}) \to HF((X,J^r),H^r;\Lambda_{0,\mathrm{nov}}).$$

Proof. (1) is a consequence of Theorem 8.2 and [23, Theorem 16.31 (3)]. The proof of (2) is similar. (We again note that in this situation we obtain a homotopy with energy loss 0.)

(3) We first observe that $\mathfrak{N}_{rr} \circ \mathfrak{N}_{rr}$ is homotopic to $\mathfrak{N}_{rr}$. This is a consequence of Theorem 8.2. Furthermore by its proof we can show that the energy loss of the homotopy between them is 0. Therefore, by [23, Theorem 16.31 (2)] we find that $\psi_{0,rr*} \circ \psi_{0,rr*} = \psi_{0,rr*}$. Since $\psi_{0,rr*}$ is an isomorphism, this implies that $\psi_{0,rr*}$ is the identity map. □

[34] The domain and the target of the map (9.5) are different not only because we use different almost complex structures but also we made various choices to define the linear K-system and various choices to define a map between spaces of differential forms via smooth correspondence by K-spaces.

Remark 9.12. We did not use the identity morphism in this section. (In several results such as [23, Theorem 16.9 (2)], we used the identity morphism in their proofs.) We can actually prove the following. See also [23, Section 18.11].

Claim 9.13. *In the case $\mathcal{J}$ and H^{11} are τ independent families, the morphism $\mathfrak{N}_{11}(\mathcal{J}, H^{11})$ is homotopic to the identity morphism.*

This immediately implies Lemma 9.11 (3) for example.

The strata of $\mathfrak{N}_{11}(\mathcal{J}, H^{11})$ is actually the same as the ones appearing in the definition of the identity morphism. Let us explain this fact below. Recall that we used $\tau \in \mathbb{R}$ independent H^{11} and J to define our morphism $\mathfrak{N}_{11}(\mathcal{J}, H^{11})$. Therefore, an element $((\Sigma, \vec{z}_{\pm}), u)$ of the interpolation space $\mathcal{N}(X, \mathcal{J}, H^{11}; \alpha_-, \alpha_+)$ is the same as an element of $\mathcal{M}(X, J, H; \alpha_-, \alpha_+)$, except we add the data to specify the main component and fix a parametrization of the main component. (Namely the isomorphism between two elements is required to commutes strictly with the parametrization of the main component.) This causes two points where $\mathcal{N}(X, \mathcal{J}, H^{11}; \alpha_-, \alpha_+)$ is different from $\mathcal{M}(X, J, H; \alpha_-, \alpha_+)$.

(1) In the case the main component represents an element of $\mathcal{N}^{\mathrm{reg}}(X, H; \alpha, \alpha')$ $\alpha \neq \alpha'$, it has an extra parameter $\in \mathbb{R}$ other than those in $\mathcal{M}^{\mathrm{reg}}(X, H; \alpha, \alpha')$ which specify the parametrization φ_{a_0} of the main component Σ_{a_0}.
(2) There is a case when the main component corresponds to an element of $\mathcal{N}^{\mathrm{reg}}(X, H; \alpha, \alpha)$.

In the case (1) the moduli parameter of the main component is $\mathcal{N}^{\mathrm{reg}}(X, H; \alpha, \alpha')$ which is isomorphic to $\mathcal{M}^{\mathrm{reg}}(X, H; \alpha, \alpha') \times \mathbb{R}$. Thus the strata [23, Definition 18.55 (1)(a)] appears.

In the case (2) we have R_α as the parameter space of the main component. In this case $\mathcal{M}^{\mathrm{reg}}(X, H; \alpha, \alpha)$ is an empty set. The map which is constant in the $\mathbb{R}$ direction corresponds to an element of R_α. Thus the strata [23, Definition 18.55 (1)(b)] appears.

Thus we find a K-space $\mathcal{N}(X, H; \alpha, \alpha')$ which is similar to the interpolation space of the identity morphism. We remark, however, that to prove Claim 9.13, we need to show not only the underlying topological space but also their Kuranishi structures coincide. In other words, the Kuranishi structure on $\mathcal{N}(X, \mathcal{J}, H^{11}; \alpha_-, \alpha_+)$ should be induced by the forgetful map

$$\mathcal{N}(X, \mathcal{J}, H^{11}; \alpha_-, \alpha_+) \to \mathcal{M}(X, J, H; \alpha_-, \alpha_+).$$

It is possible to find such a Kuranishi structure on $\mathcal{N}(X, \mathcal{J}, H^{11}; \alpha_-, \alpha_+)$. However, since we postpone the thorough detail of the discussion of the forgetful map to [24], we do not prove Claim 9.13 here. For this reason, we organize the proof in this section in a slightly different way.

10. Calculation of Hamiltonian Floer cohomology

Definition 10.1. Let (X, ω) be a compact symplectic manifold. We define the *trivial linear K-system of X* as follows. Here we use the same item numbers and the notation in [23, Condition 16.1].

(I) We define an additive group $\mathfrak{G} = \pi_2(X)/\sim$, where $\alpha \sim \alpha'$ if and only if $\omega[\alpha] = \omega[\alpha']$ and $c_1(TX)[\alpha] = c_1(TX)[\alpha']$. Group homomorphisms $E : \mathfrak{G} \to \mathbb{R}$ and $\mu : \mathfrak{G} \to \mathbb{Z}$ are induced by $[\alpha] \mapsto \omega[\alpha]$ and $\mu([\alpha]) = 2c_1(TX)[\alpha]$, respectively.

(II) As a set $\mathfrak{A} = \mathfrak{G}$ with left multiplication as the $\mathfrak{G}$ action on $\mathfrak{A}$ itself. The maps $E : \mathfrak{A} \to \mathbb{R}$ and $\mu : \mathfrak{A} \to \mathbb{Z}$ are as above.

(III) For each $\alpha \in \mathfrak{A}$, $R_\alpha = X$.

(IV) $\mathcal{M}(\alpha_-, \alpha_+) = \emptyset$ always.

(VII) o_{R_α} is the canonical orientation of the symplectic manifold X. The orientation isomorphism $\mathrm{OI}_{\alpha_-,\alpha_+}$ in [23, (16.2)] is trivial.

Other items in [23, Condition 16.1] are satisfied in a trivial way. We denote the trivial linear K-system of X by $\mathcal{F}_X^{\rm tri}$.

The main result of this section is the following.

Theorem 10.2. *Suppose we are in Situation 9.4. Let $\mathcal{F}_X(H,J)$ be as in Construction 9.5 (1). Then there exist morphisms of linear K-systems $\mathfrak{N}_{*(H,J)} : \mathcal{F}_X(H,J) \to \mathcal{F}_X^{\rm tri}$ and $\mathfrak{N}_{(H,J)*} : \mathcal{F}_X^{\rm tri} \to \mathcal{F}_X(H,J)$ with the following properties.*

(1) *The composition $\mathfrak{N}_{*(H,J)} \circ \mathfrak{N}_{(H,J)*} : \mathcal{F}_X^{\rm tri} \to \mathcal{F}_X^{\rm tri}$ is homotopic to a morphism of energy loss 0.*[35]

(2) *The composition $\mathfrak{N}_{(H,J)*} \circ \mathfrak{N}_{*(H,J)} : \mathcal{F}_X(H,J) \to \mathcal{F}_X(H,J)$ is homotopic to the morphism $\mathfrak{N}_{11}(\mathcal{J}, H^{11})$ in Construction 9.6, where $\mathcal{J}$ is the trivial family and $H^{11} \equiv H$ is constant in the $\mathbb{R}$ factor.*

Corollary 10.3.

$$HF((X,H);\Lambda_{\rm nov}) \cong H(X;\Lambda_{\rm nov}).$$

The corollary is an immediate consequence of Theorem 10.2, [23, Theorem 16.39 (4)(5)], [23, Lemma 19.45] and an obvious fact that the Floer cohomology of the trivial linear K-system is $H(X;\Lambda_{\rm nov})$. We note that Corollary 10.3 implies (1.2).

Proof of Theorem 10.2. We consider a smooth function $H^{*1} : X \times \mathbb{R} \times S^1 \to \mathbb{R}$ such that:

(1) If $\tau < -1$ then $H^{*1}(x,\tau,t) = H(x,t)$.

(2) If $\tau > 1$ then $H^{*1}(x,\tau,t) = 0$.

Since we already proved J independence of the Floer cohomology, we fix J in this section and do not include it in the notation.

Let $\alpha_- \in \mathfrak{A}$ where $\mathfrak{A}$ is the index set of the space of contractible periodic orbits of $\mathcal{F}_X(H,J)$. Let $\alpha_+ \in \mathfrak{G} = \pi_2(X)/\sim$ as in Definition 10.1.

[35] Namely it satisfies Lemma 9.7 (1)(2).

Definition 10.4. The set $\widehat{\mathcal{N}}'_\ell(X, H^{*1}; \alpha_-, \alpha_+)$ consists of triples

$$((\Sigma, (z_-, z_+, \vec{z}), a_0), u, \varphi)$$

satisfying the following conditions:

(1) $(\Sigma, (z_-, z_+, \vec{z}))$ is a genus zero semi-stable curve with $\ell+2$ marked points.
(2) Σ_{a_0} is one of the mainstream components. We call it the *main component*.
(3) $\varphi = (\varphi_a)$ where $\varphi_a : \mathbb{R} \times S^1 \to \Sigma_a \setminus \{z_{a,-}, z_{a+}\}$ is a parametrization of mainstream component Σ_a with $a \le a_0$ and φ_a is a biholomorphic map such that

$$\lim_{\tau \to \pm} \varphi_a(\tau, t) = z_{a,\pm}.$$

(4) For each extended mainstream component $\widehat{\Sigma}_a$, the map u induces $u_a : \widehat{\Sigma}_a \setminus \{z_{a,-}, z_{a,+}\} \to X$ which is a continuous map
(5) If Σ_a is a mainstream component with $a \le a_0$ and $\varphi_a : \mathbb{R} \times S^1 \to \Sigma_a$ is as in (3), then the composition $u_a \circ \varphi_a$ satisfies the equation

$$\frac{\partial(u_a \circ \varphi_a)}{\partial \tau} + J\left(\frac{\partial(u_a \circ \varphi_a)}{\partial t} - \mathfrak{X}_{H^a_{\tau,t}} \circ (u_a \circ \varphi_a)\right) = 0 \qquad (10.1)$$

where

$$H^a_{\tau,t} = \begin{cases} H^1_t & \text{if } a < a_0, \\ H^{*1}_{\tau,t} & \text{if } a = a_0. \end{cases}$$

(6)

$$\int_{\mathbb{R} \times S^1} \left\| \frac{\partial(u \circ \varphi_a)}{\partial \tau} \right\|^2 d\tau dt < \infty.$$

(7) If Σ_{v} is a bubble component or a mainstream component Σ_a with $a > a_0$, then u is pseudo-holomorphic on Σ_{v}.
(8) Let Σ_{a_1} and Σ_{a_2} be mainstream components and $z_{a_1,+} = z_{a_2,-}$. Then

$$\lim_{\tau \to +\infty} (u_{a_1} \circ \varphi_{a_1})(\tau, t) = \lim_{\tau \to -\infty} (u_{a_2} \circ \varphi_{a_2})(\tau, t)$$

holds for each $t \in S^1$ if $a_2 \le a_0$. ((6) and Lemma 6.3 imply that the left and right hand sides both converge.)

If $a_2 > a_0$ then we require that u is continuous at $z_{a_1,+} = z_{a_2,-}$.
(9) If Σ_a is a mainstream component and $z_{a,-} = z_-$, then there exists $(\gamma_-, w_-) \in R_{\alpha_-}$ such that

$$\lim_{\tau \to -\infty} (u_a \circ \varphi_a)(\tau, t) = \gamma_-(t).$$

Moreover, $[u_*[\Sigma]] \# w_-$ represents the class $[\alpha_+]$, where $\#$ is the obvious concatenation.
(10) We assume that $((\Sigma, (z_-, z_+, \vec{z}), a_0), u, \varphi)$ is stable in the sense of Definition 10.5 below.

Assume that $((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ satisfies (1)–(9) above. The *extended automorphism group* $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ of $((\Sigma,(z_-,z_+,\vec{z}),a_0)),u,\varphi)$ consists of map $v:\Sigma\to\Sigma$ such that it satisfies (1)(2)(5) of Definition 6.7, and (3) of Definition 6.7 for φ_a with $a\le a_0$, and $\tau_{a_0}=0$.

Definition 10.5. An object $((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ satisfies (1)–(9) above is said to be *stable* if $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ is a finite group.

We can define the equivalence relation $\sim_2$ on $\widehat{\mathcal{N}}'_\ell(X,H^{*1};\alpha_-,\alpha_+)$ in the same way as in Definition 3.7 except we require $\tau_{a_0}=0$ and require (3) only for $a\le a_0$. We put

$$\mathcal{N}'_\ell(X,H^{*1};\alpha_-,\alpha_+)=\widehat{\mathcal{N}}'_\ell(X,H^{*1};\alpha_-,\alpha_+)/\sim_2. \tag{10.2}$$

This space $\mathcal{N}'_\ell(X,H^{*1};\alpha_-,\alpha_+)$ (more precisely, $\mathcal{N}'_\ell(X,H^{*1};\alpha_-,\alpha_+)^{\boxplus 1}$) will be the underlying topological space of the interpolation space of the morphism $\mathfrak{N}_{*(H,J)}$.

Remark 10.6. We use the $\mathbb{R}\times S^1$ parametrized family of Hamiltonians H^{*1} in exactly the same way as in Definition 6.6 etc. and obtain the space $\mathcal{N}_\ell(X,\mathcal{J},H^{*1};\alpha_-,\alpha_+)$ as in Definition 6.10. (Here $\mathcal{J}$ is the family $J_{\tau,t}=J_t$ of almost complex structures.)

The main difference between $\mathcal{N}_\ell(X,\mathcal{J},H^{*1};\alpha_-,\alpha_+)$ and $\mathcal{N}'_\ell(X,H^{*1};\alpha_-,\alpha_+)$ is that in the latter we do not put parametrizations on the mainstream components Σ_a with $a>a_0$. Note because of the equivalence relation $\sim_2$ the parametrizations of the mainstream components Σ_a (for $a\ne a_0$) is a part of the data of elements of $\mathcal{N}_\ell(X,\mathcal{J},H^{*1};\alpha_-,\alpha_+)$ up to the translation of the $\mathbb{R}$ direction. So for a given element of $\mathcal{N}_\ell(X,\mathcal{J},H^{*1};\alpha_-,\alpha_+)$ which has exactly m mainstream components Σ_a with $a>a_0$, the corresponding element in $\mathcal{N}'_\ell(X,H^{*1};\alpha_-,\alpha_+)$ is parametrized by $(S^1)^m$. In other words, there exists a map

$$\pi:\mathcal{N}_\ell(X,\mathcal{J},H^{*1};\alpha_-,\alpha_+)\to\mathcal{N}'_\ell(X,H^{*1};\alpha_-,\alpha_+)$$

whose fibers are $(S^1)^m$. (Note π is not a fiber bundle and the dimension of the fiber m depends on the strata.) In this article, however, we do not define an equivariant Kuranishi structure of $\mathcal{N}_\ell(X,\mathcal{J},H^{*1};\alpha_-,\alpha_+)$ by the strata-wise $(S^1)^m$ action, but can use the Kuranishi structure on the 'quotient space' of this action, which is nothing but $\mathcal{N}'_\ell(X,H^{*1};\alpha_-,\alpha_+)$. This is the reason we do not need to study an S^1 equivariant Kuranishi structure in this article.

We will discuss Kuranishi structure on $\mathcal{N}'_\ell(X,H^{*1};\alpha_-,\alpha_+)$. Before doing so we define another moduli space which will be the interpolation space of $\mathfrak{N}_{(H,J)*}$.

We put

$$H^{1*}(x,\tau,t)=H^{*1}(x,-\tau,t).$$

Let $\alpha_+\in\mathfrak{A}$ where $\mathfrak{A}$ is the index set of contractible periodic orbits of $\mathcal{F}_X(H,J)$. Let $\alpha_-\in\mathfrak{G}=\pi_2(X)/\sim$ as before.

Definition 10.7. The set $\widehat{\mathcal{N}}'_\ell(X,H^{1*};\alpha_-,\alpha_+)$ consists of triples $((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ satisfying the following conditions:

(1) The same as Definition 10.4 (1).
(2) The same as Definition 10.4 (2).
(3) The same as Definition 10.4 (3), except we replace $a \le a_0$ by $a \ge a_0$.
(4) The same as Definition 10.4 (4).
(5) The same as Definition 10.4 (5), except we replace $a \le a_0$ by $a \ge a_0$ and we put

$$H^a_{\tau,t} = \begin{cases} H^1_t & \text{if } a > a_0, \\ H^{1*}_{\tau,t} & \text{if } a = a_0. \end{cases}$$

(6) The same as Definition 10.4 (6).
(7) The same as Definition 10.4 (7), except we replace $a > a_0$ by $a < a_0$.
(8) Let Σ_{a_1} and Σ_{a_2} be mainstream components. If $z_{a_1,+} = z_{a_2,-}$, then

$$\lim_{\tau\to+\infty}(u_{a_1}\circ\varphi_{a_1})(\tau,t) = \lim_{\tau\to-\infty}(u_{a_2}\circ\varphi_{a_2})(\tau,t)$$

holds for each $t \in S^1$ if $a_1 \ge a_0$. ((6) and Lemma 6.3 imply that the left and right hand sides both converge.)

If $a_1 < a_0$, then we require that u is continuous at $z_{a_1,+} = z_{a_2,-}$.
(9) If Σ_a is mainstream components and $z_{a,+} = z_+$, then there exist $(\gamma_+, w_+) \in R_{\alpha_+}$ such that

$$\lim_{\tau\to+\infty}(u_a\circ\varphi_a)(\tau,t) = \gamma_+(t).$$

Moreover, $[u_*[\Sigma]]\#[\alpha_-] \cong [w_+]$ where $\#$ is the obvious concatenation.
(10) We assume that $((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi)$ is stable, which can be defined in the same way as in Definition 10.5 above.

We can define an equivalence relation $\sim_2$ in the same way and define

$$\mathcal{N}'_\ell(X,H^{1*};\alpha_-,\alpha_+) = \widehat{\mathcal{N}}'_\ell(X,H^{1*};\alpha_-,\alpha_+)/\sim_2 . \tag{10.3}$$

When X is a point, we denote $\mathcal{N}'_\ell(X,H^{*1};\alpha_-,\alpha_+)$ (resp. $\mathcal{N}'_\ell(X,H^{1*};\alpha_-,\alpha_+)$) in this case by $\mathcal{N}'_\ell(\text{source},\text{right})$ (resp. $\mathcal{N}'_\ell(\text{source},\text{left})$).

Example 10.8. The space $\mathcal{N}'_1(\text{source},\text{right})$ is homeomorphic to the disc D^2. In fact, $\mathcal{N}'_1(\text{source},\text{right})$ consists of three strata. One is the case when there is one mainstream component. This stratum is homeomorphic to $\mathbb{R}\times S^1$. The second is the case when there are two mainstream components Σ_a, Σ_{a_0} with $a < a_0$. (Here Σ_{a_0} is the main component.) The marked point is on Σ_a. If it is $\varphi_a(\tau,t)$, then t is the well-defined parameter of this stratum, which is homeomorphic to S^1. The third is the case when there are two mainstream components Σ_a, Σ_{a_0} with $a > a_0$. The marked point is on Σ_a. Since the parametrization of Σ_a is not a part of the data of an element of $\mathcal{N}'_1(\text{source},\text{right})$, this stratum is one point.

We remark that $\mathcal{N}_1(\text{source})$ is homeomorphic to $S^1\times[0,1]$. We shrink one of the boundary components by the S^1 action to obtain $\mathcal{N}'_1(\text{source},\text{right})$.

The next proposition proves a part of Theorem 10.2.

Proposition 10.9. (1) *We can define topologies on $\mathcal{N}'_\ell(X,H^{*1};\alpha_-,\alpha_+)$ and $\mathcal{N}'_\ell(X,H^{1*};\alpha_-,\alpha_+)$ so that they are compact and Hausdorff.*

(2) *We can define Kuranishi structures on them.*
(3) *We can define Kuranishi structures on the spaces*[36] $\mathcal{N}'_\ell(X, H^{*1}; \alpha_-, \alpha_+)^{\boxplus 1}$ *and* $\mathcal{N}'_\ell(X, H^{1*}; \alpha_-, \alpha_+)^{\boxplus 1}$.
(4) *Together with other objects* $\mathcal{N}'_\ell(X, H^{*1}; \alpha_-, \alpha_+)^{\boxplus 1}$ *defines a morphism* $\mathfrak{N}_{*(H,J)} : \mathcal{F}_X(H, J) \to \mathcal{F}_X^{\rm tri}$*, of which it will be an interpolation space.*
(5) *Together with other objects* $\mathcal{N}'_\ell(X, H^{1*}; \alpha_-, \alpha_+)^{\boxplus 1}$ *defines a morphism* $\mathfrak{N}_{(H,J)*} : \mathcal{F}_X^{\rm tri} \to \mathcal{F}_X(H, J)$*, of which it will be an interpolation space.*

Proof. We prove the case of $\mathcal{N}'_\ell(X, H^{*1}; \alpha_-, \alpha_+)$ since the case of $\mathcal{N}'_\ell(X, H^{1*}; \alpha_-, \alpha_+)$ is entirely similar.

The proof is classical and similar to that of Theorem 6.4. Indeed Proposition 10.9 (1) can be proved in the same way as the proof of Lemma 6.13. (See Definitions 3.13, 3.17 and [25, Lemma 10.4 and Theorem 11.1] etc.) Thus we will describe the point where the proof is different below.

The notion of symmetric stabilization $\vec{w}$ is slightly different. We put marked points not only on bubble components but also on the mainstream components Σ_a with $a > a_0$ such that Σ_a contains no singular or marked points other than transit points.

We put canonical marked points only on the mainstream component Σ_a with $a < a_0$ (such that Σ_a contains no singular or marked points other than transit points.)

The notation of obstruction bundle data is modified as follows. Definition 4.11 (1) (symmetric stabilization) is modified as above. Definition 4.11 (2),(a) is changed as follows. Let $\Sigma_{\rm v} = \Sigma_a$ be a mainstream component.

(i) $\mathcal{V}(\mathfrak{x}_a \cup \vec{w}_{{\rm can},a})$ is an open subset of $\overset{\circ}{\mathcal{M}}_{\ell_a+\ell'_a+\ell''_a}(\text{source})$ if $a < a_0$.
(ii) $\mathcal{V}(\mathfrak{x}_a \cup \vec{w}_a)$ is an open subset of $\overset{\circ}{\mathcal{N}}_{\ell_a+\ell'_a+\ell''_a}(\text{source})$ if $a = a_0$.
(iii) $\mathcal{V}(\mathfrak{x}_a \cup \vec{w}_a)$ is an open subset of $\overset{\circ}{\mathcal{M}}{}^{\rm cl}{}_{\ell_a+\ell'_a+\ell''_a}$ if $a > a_0$.

Definition 4.11 (3) is mostly the same but Definition 4.11 (3) (a) for transit points $z_{a,-}$ with $a > a_0$ (or $z_{a,+}$ with $a > a_0$, $z_{a,+} \neq z_+$) is slightly modified as follows. As we mentioned in Remark 10.6, we did not fix a parametrization φ_a for each mainstream component Σ_a with $a > a_0$. Instead we take any biholomorphic map $\varphi_a : \mathbb{R} \times S^1 \to \Sigma_a \backslash \{z_{a,-}, z_{a,+}\}$ such that $\lim_{\tau \to \pm\infty} \varphi_a(\tau, t) = z_{a,\pm}$. It is determined up to the $\mathbb{R} \times S^1$ action on the source.

For transit points with extra S^1 factor, we perform the process of outer collaring for the factor $(T_{0,j}, \infty] \cong [0, s_{0,j})$, where $s_{0,j} = 1/\log T_{0,j}$. Namely, we change $\prod([0, s_{0,j}) \times S^1)/\sim$ to $\prod([-1, s_{0,j}) \times S^1)/\sim'$, where $\sim'$ is the equivalence relation given by

$$(s_1, t_1) \sim' (s_2, t_2) \text{ if and only if either } (s_1, t_1) = (s_2, t_2) \text{ or } s_1 = s_2 = -1.$$

Write $D_j = ([-1, s_{0,j}) \times S^1)/\sim'$. Namely, we fill a smaller disk around the origin of $([0, s_{0,j}) \times S^1)/\sim$. We call this procedure the *fattening* of the origin of the complex smoothing parameter. See Fig. 10.

[36] This is defined from $\mathcal{N}'_\ell(X, H^{*1}; \alpha_-, \alpha_+)$ in the same way as before by including $t_z \in [-1, 0]$ for any transit point z. See also Remark 5.6.

Note that the boundary of $\prod D_j$ is empty and does not contribute to the coodimension one boundary of the moduli spaces, which we are dealing with in this proof. Recall that the exponential decay estimates [16, Chapter 8] were used during the outer collar construction in Sects. 5, 6. We also have the exponential decay estimates for derivatives involving the S^1 directions in the smoothing parameters [16, Chapter 8]. Therefore, the fattening construction goes in a similar way to the outer collar construction. The same is also for the interpolation of Kuranishi structures among various Kuranishi structures. Hence we have (3).

Definition 4.11 (4)–(6) are the same.

Definition 4.11 (7) is slightly changed as follows. We also take codimension 2 submanifolds $\mathcal{D}_i$ for the additional marked points w_i lying on the mainstream components Σ_a with $a > a_0$.

We have thus defined the notion of obstruction bundle data. We define the notion of stabilization data in Definition 4.26 as a part of the obstruction bundle data modified above.

Let $\mathbf{p} = [((\Sigma, (z_-, z_+, \vec{z}), a_0), u, \varphi)] \in \mathcal{N}'_\ell(X, H^{1*}; \alpha_-, \alpha_+)$ and we take a symmetric stabilization $\vec{w}$ and the canonical marked points $\vec{w}_{\rm can}$ in the above sense. Suppose we have obstruction bundle data or stabilization data on $\mathbf{p}$. We will define a map

$$\begin{aligned}\Phi_{\mathbf{p}} : \prod \mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec{w}_{\rm v} \cup \vec{w}_{\rm can,v}) \times (T_0, \infty]^{k_1} \times \prod_{j=1}^{m+k_2} \left(((T_{0,j}, \infty] \times S^1)/\sim\right) \\ \to \mathcal{N}'_{\ell+\ell'+\ell''}(\text{source}, \text{left}),\end{aligned} \tag{10.4}$$

which is similar to but slightly different from (6.6). Here k_1 is the number of transit points $z_{a,+}$ with $a < a_0$, k_2 is the number of transit points $z_{a,-}$ with $a > a_0$. (Note $k_1 + k_2$ is the number of all transit points. m is the number of non-transit singular points.)

The definition of (10.4) is mostly the same as (6.6). The main difference is that we have an extra S^1 factor for each transit point $z_{a,-}$ with $a > a_0$. We use this parameter as follows. In the case when the components will not be glued with the main component, the role of the S^1 factor is the same as that for the case of non-transit singular points in the definition of the map (4.8). So we consider the case when all the parameters T_j corresponding to transit points $z_{a,-}$ with $a > a_0$ are finite. In this case we first use the parameters $(T_j, \theta_j) \in ((T_{0,j}, \infty] \times S^1)/\sim$ (which is associated to those transit points) with local coordinates at those transit points to glue the spaces to obtain Σ'. We define the parametrization of the main component $\Sigma'_{a'_0}$ as follows. We identify $\Sigma_{a_0} \cong \mathbb{R} \times S^1$. Then we can embed $[-T_0, T_0] \times S^1 \subset \Sigma_{a_0}$ into $\Sigma'_{a'_0}$ for sufficiently large number T_0 in a canonical way. Let $\mathfrak{v}$ be this embedding. We require that $\varphi'_{a'_0} = \varphi_{a_0} \circ \mathfrak{v}$ on $[-T_0, T_0] \times S^1$. This condition determines $\varphi'_{a'_0}$ uniquely.

The definition of (10.4) is the same as (6.6) in the other points.

Using (10.4) we can define the notion of ϵ-closeness in the same way as in (4.16).

The definition of the notion of transversal constraint is the same as Definition 4.19, except we apply Definition 4.19 (1) to the marked points w'_i corresponding to ones on the mainstream component Σ_a with $a > a_0$ also.

We can then define an embedding

$$I_{\mathbf{p},\mathrm{v};\Sigma',u',\varphi'} : E_{\mathbf{p},\mathrm{v}}(\mathfrak{y}) \to C^\infty(\Sigma'; (u')^*TX \otimes \Lambda^{0,1})$$

of obstruction spaces in the same way as in Definition 4.21.

In the rest of the construction of the Kuranishi structure, there is nothing to change the proof of Theorem 4.1 (1) and we obtain a Kuranishi structure on $\mathcal{N}'_\ell(X, H^{*1}; \alpha_-, \alpha_+)$ in the same way.

Note that the normalized corner $\widehat{S}_m(\mathcal{N}'_\ell(X, H^{*1}; \alpha_-, \alpha_+))$ of this K-space is the disjoint union of the following fiber products.

$$\begin{aligned}&\mathcal{M}_{\ell_{1,1}}(X, J_1, H^1; \alpha_{1,0}, \alpha_{1,1})\ {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots \\ &\quad \cdots\ {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\ell_{1,m}}(X, J_1, H^1; \alpha_{1,m-1}, \alpha_{1,m}) \\ &\quad {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{N}'_{\ell'}(X, H^{*1}; \alpha_{1,m}, \alpha_+),\end{aligned} \tag{10.5}$$

where $\alpha_- = \alpha_{1,0}, \alpha_{1,1}, \ldots, \alpha_{1,m-1}, \alpha_{1,m} \in \mathfrak{A}_1$ and $\ell_{1,1} + \cdots + \ell_{1,m} + \ell' = \ell$.

The important remark here is that there is no fiber product factor appearing to the 'right' of $\mathcal{N}'_{\ell'}(X, H^{*1}; \alpha_{1,m}, \alpha_+)$. Note the fiber product factors such as $\mathcal{M}_{\ell_{1,j}}(X, J_1, H^1; \alpha_{1,j-1}, \alpha_{1,j})$ are attached to the main component at the part where $\tau \to -\infty$. One of the factors of $(T_0, \infty]^{k_1}$ in (10.4) parametrizes the way we glue it on the transit points (which are $z_{a,+}$ with $a < a_0$). On the other hand, the way we glue mainstream components Σ_a with $a > a_0$ at the transit points (which are $z_{a,-}$ with $a > a_0$) are parametrized by one of the factors in $\prod_{j=1}^{m+k_2} \left(((T_{0,j}, \infty] \times S^1)/\sim\right)$ in (10.4). Because of the S^1 factors, this parameter does *not* correspond to the boundary or the corner of the K-space. This is the reason there is no fiber product factor appearing to the 'right' in (10.5).

We have thus completed the construction of the Kuranishi structure on

$$\mathcal{N}'_\ell(X, H^{*1}; \alpha_-, \alpha_+).$$

So we have proved Proposition 10.9 (2).

We next modify the Kuranishi structure of $\mathcal{N}'_\ell(X, H^{*1}; \alpha_-, \alpha_+)^{\boxplus 1}$ to obtain an interpolation space of the morphism $\mathfrak{N}_{*(H,J)} : \mathcal{F}_X(H, J) \to \mathcal{F}_X^{\mathrm{tri}}$, (which is the proof of Proposition 10.9 (3) (4).) For this purpose we modify the Kuranishi structure of $\mathcal{N}'_\ell(X, H^{*1}; \alpha_-, \alpha_+)^{\boxplus 1}$ obtained by [23, Lemma-Definition 17.38] on (10.5) $\times[-1, 0]^m$. The way to modify it is the same as the proof of Theorem 6.4 (3)(4). So we do not repeat it here. We remark again that the boundary and corner in (10.5) appear only to the 'left' from the main component. This implies that the spaces $\mathcal{N}'_\ell(X, H^{*1}; \alpha_-, \alpha_+)^{\boxplus 1}$ (after adjusting the Kuranishi structure so that it is compatible with the fiber product description (10.5)) will become the interpolation space of the morphism $: \mathcal{F}_X(H, J) \to \mathcal{F}_X^{\mathrm{tri}}$. In fact, in the trivial linear K-system $\mathcal{F}_X^{\mathrm{tri}}$ all the spaces of connecting orbits $\mathcal{M}(\alpha_-, \alpha_+)$ are empty sets.

The proof of Proposition 10.9 is now complete. $\square$

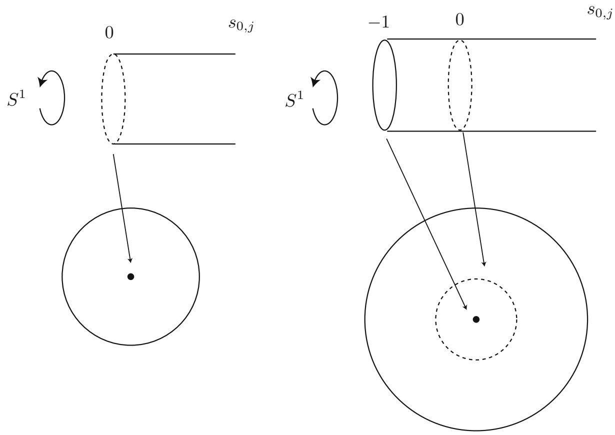


FIGURE 10. Flattering of the origin of the complex smoothing parameter

Remark 10.10. The constructed Kuranishi structures on $\mathcal{N}'_\ell(X, H^{*1}; \alpha_-, \alpha_+)^{\boxplus 1}$ and $\mathcal{N}'_\ell(X, H^{*1}; \alpha_-, \alpha_+)^{\boxplus 1}$ are $\mathfrak{C}$-collared in the sense of Remark 5.6. The Kuranishi structures on $\mathcal{N}'_\ell(X, H^{*1}; \alpha_-, \alpha_+)^{\boxplus 1}$ and $\mathcal{N}'_\ell(X, H^{1*}; \alpha_-, \alpha_+)^{\boxplus 1}$ are also compatible with the fattening in the following sense.

Definition 10.11. We call a Kuranishi structure $\widehat{\mathcal{U}}$ on $Z \times D^2$ *compatible with the fattening* if $\widehat{\mathcal{U}}$ is the product of a Kuranishi structure on Z and the trivial Kuranishi structure on D^2.

We have thus constructed morphisms in Theorem 10.2. We will prove their properties (1)(2) of Theorem 10.2. The proof is similar to that of Theorem 8.2. We will define the interpolation space of the homotopy.

We use smooth functions $H^{*1*,\cdot} : X \times [0,\infty) \times \mathbb{R} \times S^1 \to \mathbb{R}$, $H^{1*1,\cdot} : X \times [0,\infty) \times \mathbb{R} \times S^1 \to \mathbb{R}$ as follows. (Here $H^{*1*,\cdot}(x,T,\tau,t) = H^{*1*,T}(x,\tau,t)$ and $H^{1*1,\cdot}(x,T,\tau,t) = H^{1*1,T}(x,\tau,t)$.)

(1) $H^{*1*,0}(x,\tau,t) \equiv 0$ and $H^{1*1,0}(x,\tau,t) \equiv H(x,t)$.
(2)
$$H^{1*1,T}(x,\tau,t) = \begin{cases} H^{*1}(x,\tau+T,t) & \text{if } \tau \le 0, \\ H^{1*}(x,\tau-T,t) & \text{if } \tau \ge 0, \end{cases}$$
for sufficiently large T.
(3)
$$H^{*1*,T}(x,\tau,t) = \begin{cases} H^{1*}(x,\tau+T,t) & \text{if } \tau \le 0, \\ H^{*1}(x,\tau-T,t) & \text{if } \tau \ge 0, \end{cases}$$
for sufficiently large T.

We define the moduli space $\mathcal{N}'_\ell(X, H^{1*1,T}; \alpha_-, \alpha_+)$ for each T by Definition 6.10. We next define $\mathcal{N}'_\ell(X, H^{*1*,T}; \alpha_-, \alpha_+)$. Let $\alpha_\pm \in \pi_2(X)/\sim$, where $\pi_2(X)/\sim$ is as in Definition 10.1.

Definition 10.12. The set $\widehat{\mathcal{N}}'_\ell(X, H^{*1*,T}; \alpha_-, \alpha_+)$ consists of triples

$$((\Sigma, (z_-, z_+, \vec{z}), a_0), u, \varphi)$$

satisfying the following conditions:

(1) $(\Sigma, (z_-, z_+, \vec{z}))$ is a genus zero semi-stable curve with $\ell+2$ marked points.
(2) Σ_{a_0} is one of the mainstream components. We call it the *main component.*
(3) $\varphi = \varphi_{a_0}$ where $\varphi_{a_0} : \mathbb{R} \times S^1 \to \Sigma_{a_0} \backslash \{z_{a_0,-}, z_{a_0+}\}$ is a parametrization of the main component Σ_{a_0} and φ_{a_0} is a biholomorphic map such that

$$\lim_{\tau \to \pm} \varphi_{a_0}(\tau, t) = z_{a_0,\pm}.$$

(4) For each extended mainstream component $\widehat{\Sigma}_a$, the map u induces $u_a : \widehat{\Sigma}_a \backslash \{z_{a,-}, z_{a,+}\} \to X$ which is a continuous map.
(5) If Σ_{a_0} is the main component and $\varphi_{a_0} : \mathbb{R} \times S^1 \to \Sigma_{a_0}$ is as in (3), then the composition $u_{a_0} \circ \varphi_{a_0}$ satisfies the equation

$$\frac{\partial(u_{a_0} \circ \varphi_{a_0})}{\partial \tau} + J\left(\frac{\partial(u_{a_0} \circ \varphi_{a_0})}{\partial t} - \mathfrak{X}_{H^{*1*,T}_{\tau,t}} \circ (u_{a_0} \circ \varphi_{a_0})\right) = 0. \quad (10.6)$$

(6) Void. (In our situation, the finiteness of the energy is a consequence of the continuity of u in Item (8) below. See [30, (2.14)].)
(7) If $\Sigma_{\rm v}$ is a bubble component or a mainstream component Σ_a with $a \neq a_0$, then u is pseudo-holomorphic on $\Sigma_{\rm v}$.
(8) u defines a continuous map on Σ.
(9) $[u_*[\Sigma]] \# [\alpha_-] = [\alpha_+]$.
(10) We assume that $((\Sigma, (z_-, z_+, \vec{z}), a_0), u, \varphi)$ is stable in the sense of Definition 10.13 below.

Assume that $((\Sigma, (z_-, z_+, \vec{z}), a_0), u, \varphi)$ satisfies (1)–(9) above. The *extended automorphism group* $\mathrm{Aut}^+((\Sigma, (z_-, z_+, \vec{z}), a_0), u, \varphi)$ of $((\Sigma, (z_-, z_+, \vec{z}), a_0)), u, \varphi)$ consists of map $v : \Sigma \to \Sigma$ such that it satisfies (1)(2)(5) of Definition 6.7, and (3) of Definition 6.7 for φ_{a_0}, and $\tau_{a_0} = 0$.

Definition 10.13. An object $((\Sigma, (z_-, z_+, \vec{z}), a_0), u, \varphi)$ satisfies (1)–(9) above is said to be *stable* if $\mathrm{Aut}^+((\Sigma, (z_-, z_+, \vec{z}), a_0), u, \varphi)$ is a finite group.

We can define the equivalence relation $\sim_2$ on $\widehat{\mathcal{N}}_\ell(X, H^{*1*}; \alpha_-, \alpha_+)$ in the same way as in Definition 3.7 except we require $\tau_{a_0} = 0$ and require (3) only for $a = a_0$. We put

$$\mathcal{N}'_\ell(X, H^{*1*,T}; \alpha_-, \alpha_+) = \widehat{\mathcal{N}}'_\ell(X, H^{*1*,T}; \alpha_-, \alpha_+)/\sim_2 . \quad (10.7)$$

When X is a point, we denote the space $\mathcal{N}'_\ell(X, H^{*1*,T}; \alpha_-, \alpha_+)$ by $\mathcal{N}'_\ell(\text{source}; *1*, \text{finite})$.

For a later purpose, we introduce the following moduli space, which is one-dimensional higher than $\mathcal{N}'_\ell(X, H^{1*1,T}; \alpha_-, \alpha_+)$. By the definition of $H^{1*1,T}$, when $T > 1$, we have $H^{1*1,T}(x, \tau, t) = 0$ for $|\tau| < T - 1$. So we can consider the following condition in place of (10.6) in Definition 10.12 (5)

$$\frac{\partial(u_{a_0} \circ \varphi_{a_0})}{\partial \tau} + J\left(\frac{\partial(u_{a_0} \circ \varphi_{a_0})}{\partial t} - \mathfrak{X}_{H^{1*1,T}_{\tau,t}} \circ (u_{a_0} \circ \varphi_{a_0})\right) = 0 \text{ for } \tau \leq 0,$$

and

$$\frac{\partial(u_{a_0}\circ\varphi_{a_0})}{\partial\tau}+J\left(\frac{\partial(u_{a_0}\circ\varphi_{a_0})}{\partial t}-\mathfrak{X}_{H^{1*1,T}_{\tau,t+t_0}}\circ(u_{a_0}\circ\varphi_{a_0})\right)=0 \text{ for } \tau\geq 0,$$

where $t_0\in\mathbb{R}/\mathbb{Z}$. Then we consider the space of

$$((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi,t_0)$$

satisfying Conditions in Definition 10.12 with (5) replaced by the condition above and obtain the moduli space $\mathcal{N}^{\bullet}_{\ell}(X,H^{1*1,T};\alpha_-,\alpha_+)$. (In a later argument, we only need such a one-dimensional higher moduli space for $\mathcal{N}'_{\ell}(X,H^{1*1,T};\alpha_-,\alpha_+)$ but not for $\mathcal{N}'_{\ell}(X,H^{*1*,T};\alpha_-,\alpha_+)$.)

We consider

$$\bigcup_{T\in[0,\infty)}\mathcal{N}'_{\ell}(X,H^{*1*,T};\alpha_-,\alpha_+)\times\{T\}, \tag{10.8}$$

$$\bigcup_{T\in[0,\infty)}\mathcal{N}'_{\ell}(X,H^{1*1,T};\alpha_-,\alpha_+)\times\{T\}, \tag{10.9}$$

$$\bigcup_{T\in[10,\infty)}\mathcal{N}^{\bullet}_{\ell}(X,H^{1*1,T};\alpha_-,\alpha_+)\times\{T\}, \tag{10.10}$$

and will compactify them by adding certain spaces at $T=\infty$ as follows.

Definition 10.14. The set $\widehat{\mathcal{N}}'_{\ell}(X,H^{*1*,\infty};\alpha_-,\alpha_+)$ consists of triples

$$((\Sigma,(z_-,z_+,\vec{z}),a_1,a_2),u,\varphi)$$

satisfying the following conditions:

(1) $(\Sigma,(z_-,z_+,\vec{z}))$ is a genus zero semi-stable curve with $\ell+2$ marked points.
(2) $\Sigma_{a_1},\Sigma_{a_2}$ are mainstream components such that $a_1<a_2$. We call them *the first main component* and *the second main component*.
(3) $\varphi=(\varphi_a)$ where $\varphi_a:\mathbb{R}\times S^1\to\Sigma_a\setminus\{z_{a,-},z_{a+}\}$ is a parametrization of the mainstream component Σ_a with $a_1\leq a\leq a_2$ and φ_a is a biholomorphic map such that

$$\lim_{\tau\to\pm}\varphi_a(\tau,t)=z_{a,\pm}.$$

(4) For each extended mainstream component $\widehat{\Sigma}_a$, the map u induces $u_a:\widehat{\Sigma}_a\setminus\{z_{a,-},z_{a,+}\}\to X$ which is a continuous map.
(5) If Σ_a is a mainstream component with $a_1\leq a\leq a_2$ and $\varphi_a:\mathbb{R}\times S^1\to\Sigma_a$ is as in (3), then the composition $u_a\circ\varphi_a$ satisfies the equation

$$\frac{\partial(u_a\circ\varphi_a)}{\partial\tau}+J\left(\frac{\partial(u_a\circ\varphi_a)}{\partial t}-\mathfrak{X}_{H^{a,\infty}_{\tau,t}}\circ(u_a\circ\varphi_a)\right)=0. \tag{10.11}$$

Here $H^{a,\infty}=H^{*1}$ if $a=a_1$, $H^{a,\infty}=H^1$ if $a_1<a<a_2$, and $H^{a,\infty}=H^{1*}$ if $a=a_2$.
(6)

$$\int_{\mathbb{R}\times S^1}\left\|\frac{\partial(u\circ\varphi_a)}{\partial\tau}\right\|^2 d\tau dt<\infty.$$

(7) If $\Sigma_{\rm v}$ is a bubble component or a mainstream component Σ_a with $a < a_1$ or $a > a_2$, then u is pseudo-holomorphic on $\Sigma_{\rm v}$.
(8) Let Σ_a and $\Sigma_{a'}$ be mainstream components. If $z_{a,+} = z_{a',-}$ and $a_1 \le a < a' \le a_2$, then

$$\lim_{\tau\to+\infty}(u_a\circ\varphi_a)(\tau,t) = \lim_{\tau\to-\infty}(u_{a'}\circ\varphi_{a'})(\tau,t)$$

holds for each $t \in S^1$. ((6) and Lemma 6.3 imply that the left and right hand sides both converge.)

If $a < a_1$, then we require that u is continuous at $z_{a,+}$. If $a > a_2$, then we require that u is continuous at $z_{a,-}$.
(9) $[u_*[\Sigma]]\#[\alpha_-] = [\alpha_+]$.
(10) We assume that $((\Sigma,(z_-,z_+,\vec z),a_1,a_2),u,\varphi)$ is stable in the sense of Definition 10.15 below.

Assume that $((\Sigma,(z_-,z_+,\vec z),a_1,a_2),u,\varphi)$ satisfies (1)–(9) above. The *extended automorphism group* $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec z),a_1,a_2),u,\varphi)$ of $((\Sigma,(z_-,z_+,\vec z),a_1,a_2)),u,\varphi)$ consists of map $v:\Sigma\to\Sigma$ such that it satisfies (1)(2)(5) of Definition 6.7, and (3) of Definition 6.7 for φ_a with $a_1 \le a \le a_2$, and $\tau_a = 0$ if $a = a_1$ or $a = a_2$.

Definition 10.15. An object $((\Sigma,(z_-,z_+,\vec z),a_1,a_2),u,\varphi)$ satisfies (1)–(9) above is said to be *stable* if $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec z),a_1,a_2),u,\varphi)$ is a finite group.

Definition 10.16. The set $\widehat{\mathcal N}'_\ell(X,H^{1*1,\infty};\alpha_-,\alpha_+)$ consists of triples

$$((\Sigma,(z_-,z_+,\vec z),a_1,a_2),u,\varphi)$$

satisfying the following conditions:
(1) $(\Sigma,(z_-,z_+,\vec z))$ is a genus zero semistable curve with $\ell+2$ marked points.
(2) $\Sigma_{a_1},\Sigma_{a_2}$ are mainstream components such that $a_1 < a_2$. We call them *the first main component* and *the second main component*.
(3) $\varphi = (\varphi_a)$ where $\varphi_a : \mathbb R\times S^1 \to \Sigma_a\backslash\{z_{a,-},z_{a+}\}$ is a parametrization of mainstream component Σ_a with $a \le a_1$ or $a \ge a_2$, and φ_a is a biholomorphic map such that

$$\lim_{\tau\to\pm}\varphi_a(\tau,t) = z_{a,\pm}.$$

(4) For each extended mainstream component $\widehat\Sigma_a$, the map u induces $u_a : \widehat\Sigma_a\backslash\{z_{a,-},z_{a,+}\}\to X$ which is a continuous map
(5) If Σ_a is a mainstream component with $a \le a_1$ or $a \ge a_2$ and $\varphi_a : \mathbb R\times S^1 \to \Sigma_a$ is as in (3), then the composition $u_a\circ\varphi_a$ satisfies the equation

$$\frac{\partial(u_a\circ\varphi_a)}{\partial\tau} + J\left(\frac{\partial(u_a\circ\varphi_a)}{\partial t} - \mathfrak X_{H^{a,\infty}_{\tau,t}}\circ(u_a\circ\varphi_a)\right) = 0. \qquad (10.12)$$

Here $H^{a,\infty} = H^{1*}$ if $a = a_1$, $H^{a,\infty} = H^1$ if $a < a_1$ or $a > a_2$, and $H^{a,\infty} = H^{*1}$ if $a = a_2$.
(6)

$$\int_{\mathbb R\times S^1}\left\|\frac{\partial(u\circ\varphi_a)}{\partial\tau}\right\|^2 d\tau dt < \infty.$$

(7) If $\Sigma_{\rm v}$ is a bubble component or a mainstream component Σ_a with $a_1 < a < a_2$, then u is pseudo-holomorphic on $\Sigma_{\rm v}$.
(8) Let Σ_a and $\Sigma_{a'}$ be mainstream components. If $z_{a,+} = z_{a',-}$ and $a \le a' \le a_1$ or $a_2 \le a \le a'$, then

$$\lim_{\tau\to+\infty}(u_a \circ \varphi_a)(\tau,t) = \lim_{\tau\to-\infty}(u_{a'} \circ \varphi_{a'})(\tau,t)$$

holds for each $t \in S^1$. ((6) and Lemma 6.3 imply that the left and right hand sides both converge.)

If $a_1 < a \le a_2$, then we require that u is continuous at $z_{a,-}$.
(9) If Σ_a is mainstream components and $z_{a,-} = z_-$ (resp. $z_{a,+} = z_+$), then there exist $(\gamma_-, w_-) \in R_{\alpha_-}$ (resp. $(\gamma_+, w_+) \in R_{\alpha_+}$) such that

$$\lim_{\tau\to\pm\infty}(u_a \circ \varphi_a)(\tau,t) = \gamma_\pm(t)$$

Moreover, $[u_*[\Sigma]]\#w_- \sim w_+$.
(10) We assume that $((\Sigma,(z_-,z_+,\vec z),a_1,a_2),u,\varphi)$ is stable in the sense of Definition 10.17 below.

Assume that $((\Sigma,(z_-,z_+,\vec z),a_1,a_2),u,\varphi)$ satisfies (1)–(9) above. The *extended automorphism group* $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec z),a_1,a_2),u,\varphi)$ of $((\Sigma,(z_-,z_+,\vec z),a_1,a_2)),u,\varphi)$ consists of map $v:\Sigma\to\Sigma$ such that it satisfies (1)(2)(5) of Definition 6.7, and (3) of Definition 6.7 for φ_a with $a \le a_1$ or $a_2 \le a$, and $\tau_a = 0$ if $a = a_1$ or $a = a_2$.

Definition 10.17. An object $((\Sigma,(z_-,z_+,\vec z),a_1,a_2),u,\varphi)$ satisfies (1)–(9) above is said to be *stable* if $\mathrm{Aut}^+((\Sigma,(z_-,z_+,\vec z),a_1,a_2),u,\varphi)$ is a finite group.

We can define the equivalence relation $\sim_2$ on the spaces $\widehat{\mathcal N}'_\ell(X,H^{*1*,\infty};\alpha_-,\alpha_+)$ and $\widehat{\mathcal N}'_\ell(X,H^{1*1,\infty};\alpha_-,\alpha_+)$ in the same way as in Definition 3.7 except we require $\tau_{a_1} = \tau_{a_1} = 0$ and require (3) only for a for which φ_a is defined. We put

$$\begin{aligned}\mathcal N'_\ell(X,H^{*1*,\infty};\alpha_-,\alpha_+) &= \widehat{\mathcal N}'_\ell(X,H^{*1*,\infty};\alpha_-,\alpha_+)/\sim_2,\\ \mathcal N'_\ell(X,H^{1*1,\infty};\alpha_-,\alpha_+) &= \widehat{\mathcal N}'_\ell(X,H^{1*1,\infty};\alpha_-,\alpha_+)/\sim_2 .\end{aligned} \tag{10.13}$$

As in the case of $T<\infty$, we have the moduli space

$$\mathcal N^\bullet_\ell(X,H^{1*1,\infty};\alpha_-,\alpha_+)$$

by replacing Condition (5) in Definition 10.16 by the following condition:

For $a \le a_1$, $u\circ\varphi_a$ satisfies

$$\frac{\partial(u_a\circ\varphi_a)}{\partial\tau} + J\left(\frac{\partial(u_a\circ\varphi_a)}{\partial t} - \mathfrak X_{H^{a,\infty}_{\tau,t}}\circ(u_a\circ\varphi_a)\right) = 0.$$

For $a \ge a_2$, $u\circ\varphi_a$ satisfies

$$\frac{\partial(u_a\circ\varphi_a)}{\partial\tau} + J\left(\frac{\partial(u_a\circ\varphi_a)}{\partial t} - \mathfrak X_{H^{a,\infty}_{\tau,t+t_0}}\circ(u_a\circ\varphi_a)\right) = 0,$$

for some $t_0 \in \mathbb R/\mathbb Z$.

When X is a point, we denote by

$$\mathcal N'_\ell(\text{source};*1*,\infty),\quad \mathcal N'_\ell(\text{source};1*1,\infty),\quad \mathcal N^\bullet_\ell(\text{source};1*1,\infty)$$

the spaces $\mathcal{N}'_\ell(\text{one point}, H^{*1*,\infty};\alpha_-,\alpha_+)$, $\mathcal{N}'_\ell(\text{one point}, H^{1*1,\infty};\alpha_-,\alpha_+)$ and $\mathcal{N}^\bullet_\ell(\text{one point}, H^{1*1,\infty};\alpha_-,\alpha_+)$, respectively.

Definition 10.18. We put

$$(10.8)\cup\mathcal{N}'_\ell(X,H^{*1*,\infty};\alpha_-,\alpha_+)=\mathcal{N}'_\ell(X,H^{*1*};\alpha_-,\alpha_+), \quad (10.14)$$
$$(10.9)\cup\mathcal{N}'_\ell(X,H^{1*1,\infty};\alpha_-,\alpha_+)=\mathcal{N}'_\ell(X,H^{1*1};\alpha_-,\alpha_+) \quad (10.15)$$

and

$$(10.10)\cup\mathcal{N}^\bullet_\ell(X,H^{1*1,\infty};\alpha_-,\alpha_+)=\mathcal{N}^\bullet_\ell(X,H^{1*1};\alpha_-,\alpha_+). \quad (10.16)$$

When X is a point, they are written as $\mathcal{N}'_\ell(\text{source};*1*)$, $\mathcal{N}'_\ell(\text{source};1*1)$ and $\mathcal{N}^\bullet_\ell(\text{source};1*1)$, respectively.

Proposition 10.19. (1) *We can define topologies on $\mathcal{N}'_\ell(X,H^{*1*};\alpha_-,\alpha_+)$ and $\mathcal{N}'_\ell(X,H^{1*1};\alpha_-,\alpha_+)$ so that they are compact and Hausdorff.*
(2) *We can define Kuranishi structures on them.*
(3) *We can define Kuranishi structures on the spaces $\mathcal{N}'_\ell(X,H^{*1*};\alpha_-,\alpha_+)^{\boxplus 1}$ and $\mathcal{N}'_\ell(X,H^{1*1};\alpha_-,\alpha_+)^{\boxplus 1}$.*
(4) *Together with other objects $\mathcal{N}'_\ell(X,H^{*1*};\alpha_-,\alpha_+)^{\boxplus 1}$ defines a homotopy between $\mathfrak{N}_{*(H,J)}\circ\mathfrak{N}_{(H,J)*}:\mathcal{F}^{\rm tri}_X\to\mathcal{F}^{\rm tri}_X$ and a morphism of energy loss 0, of which it will be an interpolation space.*
(5) *Together with other objects $\mathcal{N}'_\ell(X,H^{1*1};\alpha_-,\alpha_+)^{\boxplus 1}$ defines a homotopy between $\mathfrak{N}_{(H,J)*}\circ\mathfrak{N}_{*(H,J)}:\mathcal{F}_X(H,J)\to\mathcal{F}_X(H,J)$ and $\mathfrak{N}_{11}(\mathcal{J},H^{11})$, of which it will be an interpolation space.*

Proof. The proof is a straight forward generalization of that of Theorem 8.6 etc. by taking the points we discussed in the proof of Proposition 10.9 into account. So we only explain the point where the proof is different from that of Theorems 8.6 etc..

When we define the notion of symmetric stabilization, we put marked points to each unstable component[37] of the source curve which is either a bubble component or a mainstream component where we do not put a parametrization φ_a. We define a canonical marked point on each unstable mainstream component on which we define a parametrization φ_a as a part of the data and which is not a main component.

When we define obstruction bundle data, the neighborhood $\mathcal{V}(\mathfrak{x}_{\rm v}\cup\vec{w}_{\rm v}\cup\vec{w}_{\rm can,v})$ is an open subset of

(i) $\overset{\circ}{\mathcal{M}}{}^{\rm cl}_{\ell_{\rm v}}$ if $\Sigma_{\rm v}$ is a bubble component or a mainstream component on which we do not define a parametrization φ,
(ii) $\overset{\circ}{\mathcal{M}}_{\ell_{\rm v}}(\text{Source})$ if $\Sigma_{\rm v}$ is a mainstream component on which we define the parametrization $\varphi_{\rm v}$ and which is not a main component,
(iii) $\overset{\circ}{\mathcal{N}}_{\ell_{\rm v}}(\text{Source})$ if $\Sigma_{\rm v}$ is a main component.

[37] We call an irreducible component of the source curve *unstable* if the set of biholomorphic map of this component preserving all the marked and singular points on it is of infinite order.

We also include a codimension 2 submanifold $\mathcal{D}_i$ as in Definition 4.11 (7) for each i if w_i corresponds to an additional marked point which is not a canonical marked point.

We will define a map $\Phi_{\mathbf{p}}$ similar to (10.4) as follows.

Case 1: $\mathbf{p} = [((\Sigma,(z_-,z_+,\vec{z}),a_1,a_2),u,\varphi)] \in \mathcal{N}'_\ell(X,H^{*1*,\infty};\alpha_-,\alpha_+)$. Let k_1 be the number of mainstream components Σ_a with $a_1 < a < a_2$. Then the map $\Phi_{\mathbf{p}}$ is

$$\begin{aligned}\Phi_{\mathbf{p}} : \prod \mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec{w}_{\rm v} \cup \vec{w}_{\rm can,v}) \times (T_0,\infty]^{k_1+1} \times \textstyle\prod_{j=1}^{m+k_2} \left(((T_{0,j},\infty] \times S^1)/\sim\right)\\ \to \mathcal{N}'_\ell(\text{source};*1*,\infty).\end{aligned} \tag{10.17}$$

We note that k_1+1 is the number of transit points which lie between Σ_{a_1} and Σ_{a_2}. Moreover, k_2 is the number of other transits points and m is the number of non-transit singular points. The reason we have the S^1 factors for the transits points which do not lie between Σ_{a_1} and Σ_{a_2} is the same as the case of (10.4).

Case 2: $\mathbf{p} = [((\Sigma,(z_-,z_+,\vec{z}),a_0),u,\varphi),T] \in \mathcal{N}_\ell(X,H^{*1*,T};\alpha_-,\alpha_+) \times \{T_{\mathbf{p}}\}$ with $T_{\mathbf{p}} > 0$. The map $\Phi_{\mathbf{p}}$ is

$$\begin{aligned}\Phi_{\mathbf{p}} : \prod \mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec{w}_{\rm v} \cup \vec{w}_{\rm can,v}) \times \prod_{j=1}^{m+k} \left(((T_{0,j},\infty] \times S^1)/\sim\right) \times (T_{\mathbf{p}}-\epsilon, T_{\mathbf{p}}+\epsilon)\\ \to \mathcal{N}'_\ell(\text{source};*1*,\infty).\end{aligned} \tag{10.18}$$

Here k is the number of transit points and m is the number of non-transit singular points. We defined the parametrization of the mainstream only on the component Σ_{a_0}. This is the reason all the factors $(T_{0,j},\infty]$ come with the S^1 factors. Note in this case $\mathbf{p}$ is an interior point of our moduli space $\mathcal{N}'_\ell(\text{source};*1*,\infty)$.

Case 3: $\mathbf{p} = [((\Sigma,(z_-,z_+,\vec{z}),a_1,a_2),u,\varphi)] \in \mathcal{N}^\bullet_\ell(X,H^{1*1,\infty};\alpha_-,\alpha_+)$. The map $\Phi^\bullet_{\mathbf{p}}$ is

$$\begin{aligned}\Phi^\bullet_{\mathbf{p}} : \prod \mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec{w}_{\rm v} \cup \vec{w}_{\rm can,v}) \times (T_0,\infty]^{m_1+m_2} \times \prod_{j=1}^{m} \left(((T_{0,j},\infty] \times S^1)/\sim\right)\\ \times \left(\prod_{j=1}^{m_*} \left(((T_{0,j},\infty] \times S^1)/\sim\right)\right) \to \mathcal{N}^\bullet_\ell(\text{source};1*1,\infty).\end{aligned} \tag{10.19}$$

Here m_* is the number of transit points which lie between Σ_{a_1} and Σ_{a_2} and m_1 (resp. m_2) is the number of transit points which lie 'left' (resp. 'right') from Σ_{a_1} (resp. Σ_{a_2}). Also m is the number of non-transit singular points.

For $\mathbf{p} = [((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi)] \in \mathcal{N}'_\ell(X, H^{1*1,\infty}; \alpha_-, \alpha_+)$, we have

$$\Phi_{\mathbf{p}} : \prod \mathcal{V}(\mathfrak{x}_{\rm v} \cup \vec{w}_{\rm v} \cup \vec{w}_{\rm can,v}) \times (T_0, \infty]^{m_1+m_2} \times \prod_{j=1}^{m} \left(((T_{0,j}, \infty] \times S^1)/\sim\right)$$
$$\times \left(\prod_{j=1}^{m_*} \left(((T_{0,j}, \infty] \times S^1)/\sim\right)\right)' \to \mathcal{N}'_\ell(\text{source}; 1*1, \infty), \tag{10.20}$$

where

$$\left(\prod_{j=1}^{m_*} \left(((T_{0,j}, \infty] \times S^1)/\sim\right)\right)' \tag{10.21}$$

is the subset of $\prod_{j=1}^{m_*} \left(((T_{0,j}, \infty] \times S^1)/\sim\right)$ defined by

$$\theta_1 + \cdots + \theta_{m_*} = 0 \text{ in } S^1 = \mathbb{R}/\mathbb{Z}.$$

Here θ_j, $j = 1, \ldots, m_*$ are the coordinates of the S^1 factors. The way the factors appear in the left hand side of (10.20) is similar to other cases except the factor (10.21). We explain the way this factor appears.

Suppose all the $(T_{0,j}, \infty]$ components in this factor are finite. In this case we use this parameter together with $\theta_j \in S^1$ to glue Σ_{a_1}, Σ_{a_2} and the components Σ_a with $a_1 < a < a_2$. We then obtain a component which we write as Σ_{a_0}. This will be the main component of the resulting element of $\mathcal{N}'_\ell(\text{source}; *1*, \infty)$. (More precisely, it may be glued with other mainstream components.)

Note we defined the parametrizations $\varphi_{a_1}, \varphi_{a_2}$ but not defined parametri zations for other Σ_a with $a_1 < a < a_2$. We consider a parametrization φ_{a_0} : $\mathbb{R} \times S^1 \to \Sigma_{a_0} \backslash \{z_{a_0,-}, z_{a_0,+}\}$ such that $\lim_{\tau \to \pm\infty} \varphi_{a_0}(\tau, t) = z_{a_0,\pm}$.

We can identify $[-T_0, T_0] \times S^1 \subset \Sigma_{a_1} \subset \Sigma_{a_0}$. Let $\mathfrak{v}_1$ be this map. We also take an embedding $\mathfrak{v}_2$ from $[-T_0, T_0] \times S^1 \subset \Sigma_{a_2}$ to Σ_{a_0}. There exist t_1, t_2 such that

$$(\mathfrak{v}_j \circ \varphi_{a_j})(\tau, t) = \varphi_{a_0}(\tau + \tau_j, t + t_j)$$

for $j = 1, 2$. Note the choice of φ_{a_0} is not unique. Namely we may change it by the $\mathbb{R}$ action. We may choose φ_{a_0} so that $\tau_1 = -\tau_2$. Then $\tau_2 - \tau_1 + 1 = T$. (See (8.11).) Here T is the second factor in (10.9).

We consider t_j. We may choose the representative φ_{a_0} so that one of them, say t_1 to be 0. However, it is impossible to take the representative φ_{a_0} for which both of t_j are zero. The notation $'$ stands for the constraint that $t_1 = t_2$. Under this assumption we may choose both of t_1 and t_2 are zero. Note if we change the S^1 factor in $((T_{0,j}, \infty] \times S^1)/\sim$ by θ_j, then $t_2 - t_1$ changes by $\sum_{j=1}^{m_*} \theta_j$.

We remark that the point $[\infty, \ldots, \infty]$ in (10.21) is a boundary point of (10.21). In fact, if we remove $'$ then it will be an interior point. Because of the constraint $t_1 = t_2$ this point becomes the boundary points. This is consistent to the fact that $\mathcal{N}'_\ell(\text{source}; 1*1, \infty)$ lies at the part $T = \infty$ of $\mathcal{N}'_\ell(\text{source}; 1*1)$.

Note that there exists $\mathfrak{T}_0 > 0$ such that the closed neighborhoods $W(\mathbf{p}_c)$ of $\mathbf{p}_c \in \mathcal{N}'_\ell(X, H^{1*1,\infty}; \alpha_-, \alpha_+)$ cover $\cup_{T \ge \mathfrak{T}_0} \mathcal{N}_\ell(X, H^{1*1,T}; \alpha_-, \alpha_+) \times \{T\}$. For a sufficiently large $\mathfrak{T} > \mathfrak{T}_0$, the obstruction space for $\mathbf{q} \in \cup_{T \ge \mathfrak{T}_0} \mathcal{N}_\ell(X, H^{1*1,T}; \alpha_-, \alpha_+) \times \{T\}$ is the direct sum of $I_{\mathbf{p_c},\mathrm{v};\mathbf{q}}(E_{\mathbf{p_c},\mathrm{v}}(\mathfrak{y}))$ with $\mathbf{q} \in W(\mathbf{p}_c)$.

Case 4: $\mathbf{p} \in \mathcal{N}_\ell(X, H^{1*1,T}; \alpha_-, \alpha_+) \times \{T_\mathbf{p}\}$. This case is entirely the same as the case of (8.9). Using this map $\Phi_\mathbf{p}$, we define the notion of ϵ-closeness. Then we can define an obstruction space using them. The definition of transversal constraint is similar. We then take an appropriate closed finite covering of our moduli space by $\{W(\mathbf{p}_c)\}_c$ and use it to define a Kuranishi structure in the same way as in the proof of Theorems 8.6.

We note that the boundary of it has the form required by Proposition 10.19 (4)(5). In fact, the boundary of the space $\mathcal{N}'_\ell(X, H^{*1*}; \alpha_-, \alpha_+)$ is only at $T = 0$ and $T = \infty$. This is the consequence of the description of the domain of the maps (10.17), (10.18), especially of the fact that the domain of (10.18) has no boundary.

The boundary of the space $\mathcal{N}'_\ell(X, H^{1*1}; \alpha_-, \alpha_+)$ other than those which are at $T = 0$ and $T = \infty$ can be described using the space of connecting orbits of $\mathcal{F}_X(H, J)$. Moreover, the part of $\mathcal{N}'_\ell(X, H^{1*1}; \alpha_-, \alpha_+)$ which appears at the part $T = \infty$ is the union of various spaces

$$\begin{aligned}
&\mathcal{M}_{\ell_{1,1}}(X, H^1; \alpha_{1,0}, \alpha_{1,1})\, {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots \\
&\cdots\, {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\ell_{1,m_1}}(X, H^1; \alpha_{1,m_1-1}, \alpha_{1,m_1}) \\
&{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{N}'_{\ell'}(X, H^{1*}; \alpha_{1,m_1}, \alpha_{*,1}) \\
&{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}^{\mathrm{cl}}_{\ell_{*,1}+2}(X; \alpha_{*,1}, \alpha_{*,2})\, {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots \\
&{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}^{\mathrm{cl}}_{\ell_{*,m_*}+2}(X; \alpha_{*,m_*-1}, \alpha_{*,m_*}) \\
&{}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{N}'_{\ell''}(X, H^{*1}; \alpha_{*,m_*}, \alpha_{2,1}) \\
&\mathcal{M}_{\ell_{2,1}}(X, H^1; \alpha_{2,1}, \alpha_{2,2})\, {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \cdots \\
&\cdots\, {}_{\mathrm{ev}_+}\times_{\mathrm{ev}_-} \mathcal{M}_{\ell_{2,m_2+1}}(X, H^1; \alpha_{2,m_2}, \alpha_{2,m_2+1}).
\end{aligned} \tag{10.22}$$

Here $\mathcal{M}^{\mathrm{cl}}_\ell(X, \alpha)$ is the moduli space of stable maps of genus 0 with ℓ marked points and of homology class α.

The situation at $T = \infty$ is analogous to the space $\mathfrak{forget}^{-1}([\Sigma_0])$ in [17, Section 2.6, Lemmas 2.6.3, 2.6.27], which is examined in detail in [17, Section 4.6]. We remark that (10.22) lies in a codimension $m_1 + m_2 + 1$ locus. This is a consequence of the discussion on (10.20).

For $\mathcal{N}^\bullet_\ell(X, H^{1*1}; \alpha_-, \alpha_+)$, we perform the collar construction as well as the fattening of origins of complex smoothing parameters at transit points between Σ_{a_1} and Σ_{a_2} as in the proof of Proposition 10.9 to obtain the desired Kuranishi structure on $\mathcal{N}^\bullet_\ell(X, H^{1*1}; \alpha_-, \alpha_+)^{\boxplus 1}$.

We regard the part of $\mathcal{N}'_\ell(X, H^{1*1}; \alpha_-, \alpha_+)$ with $T \in [\mathfrak{T}, \infty]$ as a subspace of $\mathcal{N}^\bullet_\ell(X, H^{1*1}; \alpha_-, \alpha_+)$ defined by $\theta_1 + \cdots + \theta_{m_*} = 0$. For the construction of Kuranishi structures, we take a finite subset $\{\mathbf{p}_c\} \subset \mathcal{N}'_\ell(X, H^{1*1,T}; \alpha_-, \alpha_+)$ such that the union of Int $W(\mathbf{p}_c)$ contains $\mathcal{N}'_\ell(X, H^{1*1,T}; \alpha_-, \alpha_+)$, and a finite subset $\{\mathbf{p}_{c'}\} \subset \mathcal{N}^\bullet_\ell(X, H^{1*1}; \alpha_-, \alpha_+) \setminus \mathcal{N}'_\ell(X, H^{1*1}; \alpha_-, \alpha_+)$ such that $W(\mathbf{p}_{c'})$'s are disjoint from $\mathcal{N}'_\ell(X, H^{1*1}; \alpha_-, \alpha_+)$ and $\mathcal{N}^\bullet_\ell(X, H^{1*1}; \alpha_-, \alpha_+)$

is covered by Int $W(\mathbf{p}_c)$'s and Int $W(\mathbf{p}_{c'})$'s. By the construction, the Kuranishi structure on $\mathcal{N}^{\bullet}_{\ell}(X, H^{1*1,T}; \alpha_-, \alpha_+)$ is compatible with the fattening on $D_j(\delta)$ for a sufficiently small $\delta > 0$ (Definition 10.11).

The Kuranishi structure on $\mathcal{N}^{\bullet}_{\ell}(X, H^{1*1}; \alpha_-, \alpha_+)^{\boxplus 1}$ restricts to the one on the subspace defined by $\sum_{j=1}^{k_2} \theta_j = 0$ and $T \in [\mathfrak{T}, \infty]$. Our choice of $\mathbf{p}_c$, we see that the Kuranishi structure on $\mathcal{N}'_{\ell}(X, H^{1*1}; \alpha_-, \alpha_+)$ with $T \in [\mathfrak{T}, \infty]$ can be arranged to match with the part $T \in [0, \mathfrak{T}+1]$. Thus we obtain $\mathcal{N}'_{\ell}(X, H^{1*1}; \alpha_-, \alpha_+)^{\boxplus 1}$ with the desired Kuranishi structure.

We approximate $\mathcal{N}_{\ell}(X, H^{1*1,\infty}; \alpha_-, \alpha_+)^{\boxplus 1}$ in $\mathcal{N}^{\bullet}_{\ell}(X, H^{1*1}; \alpha_-, \alpha_+)^{\boxplus 1}$ using a family of smooth submanifolds $S^{(k)}_{\epsilon}$ of codimension 2 in $\prod_{j=1}^{k_2} D_j^2$ given below.

Let $\rho : [-1, 0] \to \mathbb{R}$ be a non-decreasing smooth function such that $\rho(t) = 3(t+1)/2$ for $-1 \le t \le -1 + \delta/2$ and $\rho(t) = \delta$ for $t \ge -1 + \delta$. Choose $\epsilon < \delta/10$ and define $S^{(k)}_{\epsilon}$ by $\prod_{j=1}^{k} \rho(t_j) = \epsilon\delta^{k-1}$ and $\sum_{j=1}^{k} \theta_j = 0$ in $\mathbb{R}/\mathbb{Z}$. Then $S^{(k)}_{\epsilon}$ approaches to $\cup_{j=1}^{k} D_1^2 \times \cdots \times \check{D}_j^2 \times \cdots \times D_k^2$.

The space $\mathcal{N}^{\bullet}_{\ell}(X, H^{1*1}; \alpha_-, \alpha_+)^{\boxplus 1}$ is obtained by gluing various

$$(10.22) \times [-1, 0]^{m_1+m_2} \times \prod_{j=1}^{m_*} D_j^2 \tag{10.23}$$

to $\mathcal{N}^{\bullet}_{\ell}(X, H^{1*1}; \alpha_-, \alpha_+)$.

We define $(\mathcal{N}^{\bullet}_{\ell}(X, H^{1*1}; \alpha_-, \alpha_+)^{\boxplus 1})_{\epsilon}$ by the union of $(10.22) \times \prod_{j=1}^{m_1+m_2} \{-1\} \times S^{(m_*)}_{\epsilon}$. In order to define the smooth correspondence (see [23, Section 9.5, (9.17)]), we need to take a CF-perturbation (see [23, Chapter 7]). Since the Kuranishi structure is $\mathfrak{C}$-collared as in Proposition 5.5 and Remark 5.6 and compatible with the fattening on $D_j^2(\delta)$, we can take a CF-perturbation, which is τ-collared for a small $\tau > 0$ with respect to $\mathfrak{C}$ (see [23, Lemma 17.40 (2)]). In the same way, we arrange the CF-perturbation such that it is compatible with the fattening, i.e., the product type, on $D_j^2(\delta/2)$. Then the contribution to the smooth correspondence from the part with $t_j \in [-1, -1+\delta/2]$ is zero and the smooth correspondence given by $(\mathcal{N}^{\bullet}_{\ell}(X, H^{1*1}; \alpha_-, \alpha_+)^{\boxplus 1})_{\epsilon}$ equipped with the CF-perturbation is equal to the smooth correspondence given by (10.22).

In fact, the composition of $\mathfrak{N}_{(H,J)*} \circ \mathfrak{N}_{*(H,J)}$ is given by $((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi)$ in Definition (10.14) with a choice of one of $z_{a,+} = z_{a+1,-}$, $a_1 \le a \le a_2 - 1$, i.e., a transit point between Σ_{a_1} and Σ_{a_2}. (This choice determines the output of $\mathfrak{N}_{*(H,J)}$ and the input of $\mathfrak{N}_{(H,J)*}$. Adding this choice in the data to the object, $((\Sigma, (z_-, z_+, \vec{z}), a_1, a_2), u, \varphi, a)$, $a \in \{a_1, \ldots, a_2 - 1\}$, the intersection of $\cup_{j=1}^{k} D_1^2 \times \cdots \times \check{D}_j^2 \times \cdots \times D_k^2$ in the fattening factor is resolved (normalization) and $(\mathcal{N}^{\bullet}_{\ell}(X, H^{1*1}; \alpha_-, \alpha_+)^{\boxplus 1})_{\epsilon}$ becomes the K-space giving the smooth correspondence for $\mathfrak{N}_{(H,J)*} \circ \mathfrak{N}_{*(H,J)}$. Since our choice of the CF-perturbation makes such intersections do not contribute to the smooth correspondence, we can also use $(\mathcal{N}^{\bullet}_{\ell}(X, H^{1*1}; \alpha_-, \alpha_+)^{\boxplus 1})_{\epsilon}$.

The proof of Proposition 10.19 is complete. □

The proof of Theorem 10.2 is now complete. □

Remark 10.20. In Theorem 10.2 we did not claim that the composition $\mathfrak{N}_{*(H,J)} \circ \mathfrak{N}_{(H,J)*} : \mathcal{F}_X^{\mathrm{tri}} \to \mathcal{F}_X^{\mathrm{tri}}$ is homotopic to the identity morphism. We can prove it. In fact, it follows from Claim 9.13, but we do not provide the technical details of the proof of Claim 9.13 in this article.

References

[1] Abouzaid, M., Blumberg, A.: Arnold conjecture and Morava K-theory. arXiv:2103.01507

[2] Arnol'd, V.: Sur une propriété topologique des applications globalement canoniques de la mécanique classique. C. R. Acad. Sci. Paris **261**, 3719–3722 (1965)

[3] Atiyah, M.F., Patodi, V.K., Singer, I.M.: Spectral asymmetry and Riemannian geometry: I. Math. Proc. Camb. Phil. Soc. **88**, 43–69 (1975)

[4] Atiyah, M.F., Patodi, V.K., Singer, I.M.: Spectral asymmetry and Riemannian geometry: III. Math. Proc. Camb. Phil. Soc. **79**, 71–99 (1976)

[5] Buser, P.: Geometry and spectra of compact Riemann surfaces. In: Progr. Math., vol. 106. Birkhäuser, Basel (1992)

[6] Chekanov, Y.: Hofer's symplectic energy and Lagrangian intersections, Contact and symplectic geometry (Cambridge, 1994), 296–306. In: Publ. Newton Inst., vol. 8. Cambridge Univ. Press, Cambridge (1996)

[7] Conley, C., Zehnder, E.: The Birkhoff–Lewis fixed point theorem and a conjecture of V.I. Arnol'd. Invent. Math. **73**, 33–49 (1983)

[8] Floer, A.: Morse theory for Lagrangian intersections. J. Differ. Geom. **28**, 513–547 (1988)

[9] Floer, A.: Symplectic fixed points and holomorphic spheres. Commun. Math. Phys. **120**, 575–611 (1989)

[10] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Lagrangian Intersection Floer Theory-anomaly and Obstruction, Part I, AMS/IP Studies in Advanced Math., vol. 46.1. International Press/Amer. Math. Soc. (2009). **MR2553465**

[11] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Lagrangian Intersection Floer Theory-anomaly and Obstruction, Part II, AMS/IP Studies in Advanced Math. vol. 46.2. International Press/Amer. Math. Soc. (2009). **MR2548482**

[12] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Lagrangian Floer theory on compact toric manifolds II: bulk deformations. Selecta Math. New Ser. **17**, 609–711 (2011)

[13] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Technical details on Kuranishi structure and virtual fundamental chain. arXiv:1209.4410

[14] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Displacement of polydisks and Lagrangian Floer theory. J. Symplectic Geom. **11**(2), 231–268 (2013)

[15] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Kuranishi structure, Pseudo-holomorphic curve, and Virtual fundamental chain: Part 1. arXiv:1503.07631

[16] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Exponential decay estimate and smoothness of the moduli space of pseudoholomorphic curves, to appear in Mem. Amer. Math. Soc. arXiv:1603.07026

[17] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Lagrangian Floer theory and mirror symmetry on compact toric manifolds. Astérisque **376**, vi+340 (2016)

[18] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Anti-symplectic involution and Floer cohomology. Geom. Topol. **21**, 1–106 (2017)

[19] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Kuranishi structure, Pseudo-holomorphic curve, and Virtual fundamental chain: Part 2. arXiv:1704.01848

[20] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Construction of Kuranishi structures on the moduli spaces of pseudo-holomorphic disks: I. Surv. Differ. Geom. **22**, 133–190 (2018)

[21] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Construction of Kuranishi structures on the moduli spaces of pseudo-holomorphic disks: II. arXiv:1808.06106

[22] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Spectral invariants with bulk, quasi-homomorphisms and Lagrangian Floer theory. Mem. Amer. Math. Soc. **260**(1254), x+266 (2019)

[23] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Kuranishi structures and virtual fundamental chains. In: Springer Monographs in Mathematics. Springer, Berlin (2020) . (**MR4179586. (This is an improved version of a part of the union of [15] and [19].)**)

[24] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Forgetful map and its applications to Gromov–Witten invariant and to Lagrangian Floer theory. **in preparation**

[25] Fukaya, K., Ono, K.: Arnold conjecture and Gromov–Witten invariant. Topology **38**(5), 933–1048 (1999)

[26] Hofer, H., Salamon, D.: Floer homology and Novikov rings. In: The Floer memorial Volume, Progr. Math., vol. 133, pp. 483–524. Basel, Birkhäuser (1995)

[27] Hummel, C.: Gromov's compactness theorem for pseudo-holomorphic curves. In: Progr. Math., vol. 151. Birkhäuser, Basel (1997)

[28] Li, J., Tian, G.: Virtual moduli cycles and Gromov–Witten invariants of general symplectic manifolds. In: Topics in Symplectic 4-Manifolds (Irvine, CA, First Int. Press Lect. Ser. 1 Internat. Press Cambridge, MA, USA 1998), pp. 47–83 (1996)

[29] Liu, G., Tian, G.: Floer homology and Arnold conjecture. J. Differ. Geom. **49**, 1–74 (1998)

[30] Ono, K.: On the Arnold conjecture for weakly monotone symplectic manifolds. Invent. Math. **119**, 519–537 (1995)

[31] Pardon, J.: An algebraic approach to virtual fundamental cycles on moduli spaces of J-holomorphic curves. Geom. Topol. **20**, 779–1034 (2016)

[32] Ratcliffe, J.: Foundations of Hyperbolic Manifolds, Graduate Texts in Mathematics, vol. 149. Springer, Berlin (1994)

[33] Ruan, Y.: Virtual neighborhood and pseudoholomorphic curve. Turkish J. Math. **23**, 161–231 (1999)

[34] Satake, I.: On a generalization of of the notion of manifold. Proc. Natl. Acad. Sci. USA **42**, 359–363 (1956)

[35] Schwarz, M.: A quantum cup length estimate for symplectic fixed points. Invent. Math. **133**, 353–397 (1998)

[36] Siebert, B.: Gromov–Witten invariants of general symplectic manifolds. arXiv:dg-ga/9608005

[37] Usher, M.: Deformed Hamiltonian Floer theory, capacity estimates and Calabi quasimorphisms. Geom. Topol. **15**(3), 1313–1417 (2011)

Kenji Fukaya
Simons Center for Geometry and Physics
State University of New York
Stony Brook NY11794-3636
USA
e-mail: `kfukaya@scgp.stonybrook.edu`

Yong-Geun Oh
Center for Geometry and Physics
Institute for Basic Sciences (IBS)
Pohang
Korea
e-mail: `yongoh1@postech.ac.kr`

and

Department of Mathematics
POSTECH
Pohang
Korea

Hiroshi Ohta
Graduate School of Mathematics
Nagoya University
Nagoya
Japan
e-mail: `ohta@math.nagoya-u.ac.jp`

Kaoru Ono
Research Institute for Mathematical Sciences
Kyoto University
Kyoto
Japan
e-mail: `ono@kurims.kyoto-u.ac.jp`

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:23
https://doi.org/10.1007/s11784-022-00946-9
Published online April 4, 2022

Journal of Fixed Point Theory and Applications

What does a vector field know about volume?

Hansjörg Geiges

To Claude Viterbo on the occasion of his 60th birthday.

Abstract. This note provides an affirmative answer to a question of Viterbo concerning the existence of nondiffeomorphic contact forms that share the same Reeb vector field. Starting from an observation by Croke–Kleiner and Abbondandolo that such contact forms define the same total volume, we discuss various related issues for the wider class of geodesible vector fields. In particular, we define an Euler class of a geodesible vector field in the associated basic cohomology and give a topological characterisation of vector fields with vanishing Euler class. We prove the theorems of Gauß–Bonnet and Poincaré–Hopf for closed, oriented 2-dimensional orbifolds using global surfaces of section and the volume determined by a geodesible vector field. This volume is computed for Seifert fibred 3-manifolds and for some transversely holomorphic flows.

Mathematics Subject Classification. 57R30, 37C10, 53C22, 53D35, 57R25, 58A10.

Keywords. Reeb vector field, geodesible vector field, contact form, basic cohomology, Seifert fibration, Euler class.

1. Introduction

This paper is concerned with a question about Reeb flows posed to me by Claude Viterbo: are there nondiffeomorphic contact forms with the same Reeb vector field? Viterbo's question was prompted by Alberto Abbondandolo's discovery of a miraculous identity on differential forms.

This work is part of a project in the SFB/TRR 191 'Symplectic Structures in Geometry, Algebra and Dynamics', funded by the DFG.

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

Lemma 1.1. *(Abbondandolo) Given two differential* 1*-forms* α, β *on the same manifold, the identity*

$$\alpha \wedge (\mathrm{d}\alpha)^n - \beta \wedge (\mathrm{d}\beta)^n = (\alpha - \beta) \wedge \sum_{j=0}^{n} (\mathrm{d}\alpha)^j \wedge (\mathrm{d}\beta)^{n-j} + \mathrm{d}\big(\alpha \wedge \beta \wedge \sum_{j=1}^{n-1} (\mathrm{d}\alpha)^j \wedge (\mathrm{d}\beta)^{n-1-j}\big) \tag{1}$$

holds for any $n \in \mathbb{N}_0$.

Identity (1), whose verification is straightforward, has the following striking consequence, which—as we learned in the meantime—has been observed earlier by Croke and Kleiner [10, Lemma 2.1]. They do not state identity (1), but give a quite similar proof.

Proposition 1.2. (Croke–Kleiner) *Let* X *be a nonsingular vector field on a closed, oriented manifold* M *of dimension* $2n+1$*. Let* α, β *be* 1*-forms on* M *that are invariant under the flow of* X *and satisfy*

$$\alpha(X) = \beta(X) = 1. \tag{2}$$

Then,

$$\int_M \alpha \wedge (\mathrm{d}\alpha)^n = \int_M \beta \wedge (\mathrm{d}\beta)^n. \tag{3}$$

Proof. Given (2), the invariance condition $L_X\alpha = L_X\beta = 0$ is equivalent to

$$i_X \mathrm{d}\alpha = i_X \mathrm{d}\beta = 0 \tag{4}$$

by the Cartan formula. Then, (3) is immediate from (1) and Stokes's theorem. □

In particular, this proposition says that any two contact forms on a closed, oriented manifold that share the same Reeb vector field give rise to volume forms that integrate to the same total volume. In other words, this total volume is determined by the Reeb vector field alone. Abbondandolo has raised the question whether one can compute this volume from a given Reeb vector field, not knowing a contact form it is associated with.

Remark 1.3. Croke and Kleiner used this proposition to conclude that two compact Riemannian manifolds with C^1-conjugate geodesic flows have the same volume [10, Proposition 1.2]. This follows by considering the canonical contact form on the unit cotangent bundle, whose Reeb vector field generates the cogeodesic flow [11, Theorem 1.5.2].

As we shall see, the existence of a 1-form α as in Proposition 1.2 is equivalent to the vector field X being geodesible (Definition 3.1, Proposition 3.3).

Definition 1.4. We write vol_X for the real number defined by (3) and call it the *volume of* X, even though $\alpha \wedge (\mathrm{d}\alpha)^n$ is not, in general, a volume form.

Much of this paper is a rumination on the consequences and ramifications of Proposition 1.2, leading us ultimately towards an affirmative answer to Viterbo's question (Theorem 10.1), which shows that Proposition 1.2 is indeed a nontrivial statement, even within the class of Reeb vector fields. We pay special attention to the cases where the geodesible vector field X generates an S^1-action, or where the flow of X admits a global surface of section. In these cases, one can compute vol_X and give it a geometric interpretation.

Along the way, we introduce the Euler class e_X of a geodesible vector field X in the basic cohomology of the foliation it determines, and we argue that Proposition 1.2 ought to be interpreted as a statement in basic cohomology (Proposition 5.6). These considerations will allow us to establish a criterion for the vanishing of e_X in terms of the existence of a transverse invariant foliation (Theorem 5.7). Geodesible vector fields X with $e_X = 0$ exist precisely on manifolds that fibre over S^1 (Corollary 5.8).

In Sect. 6, we compute vol_X for vector fields that define a Seifert fibration on a 3-manifold. This computation involves the use of global surfaces of section. With similar arguments we prove the theorems of Gauß–Bonnet and Poincaré–Hopf for closed, oriented 2-dimensional orbifolds in Sect. 7.

For certain geodesible vector fields X whose flow admits a transverse holomorphic structure, we can relate vol_X to the Bott invariant of that structure. This is the content of Section 8.

In Sect. 9, we derive a formula for vol_X when X admits a global surface of section. After presenting the answer to Viterbo's question in Sect. 10, we end the paper in Sect. 11 with a brief discussion of orbit equivalent geodesible vector fields.

2. Dimension three

In dimension three, the answer to Viterbo's question is negative.

Proposition 2.1. *Let α_0, α_1 be two contact forms on a closed 3-manifold M sharing the same Reeb vector field R. Then, α_0 and α_1 define the same orientation of M. Furthermore, there is an isotopy $(\psi_t)_{t\in[0,1]}$ of M, starting at $\psi_0 = \mathrm{id}_M$, such that $\psi_1^*\alpha_1 = \alpha_0$ and $(\psi_t^*)^{-1}\alpha_0$ is a contact form with Reeb vector field R for all $t \in [0,1]$.*

Proof. The fact that α_0 and α_1 define the same orientation of M follows from Proposition 1.2, since by (3), the two volume forms $\alpha_i \wedge \mathrm{d}\alpha_i$ must have the same sign.

Set $\alpha_t := (1-t)\alpha_0 + t\alpha_1$, $t \in [0,1]$. Since $\mathrm{d}\alpha_0$ and $\mathrm{d}\alpha_1$ restrict to nondegenerate 2-forms defining the same orientation on any tangent 2-plane field transverse to R, so does $\mathrm{d}\alpha_t$. It follows that α_t is likewise a contact form with Reeb vector field R. Now, apply the Moser trick [11, p. 60] to the equation

$$\psi_t^*\alpha_t = \alpha_0, \tag{5}$$

where we would like the isotopy (ψ_t) to be the flow of a time-dependent vector field $X_t \in \ker\alpha_t$. Under this last assumption, by differentiating (5), we find

$$\alpha_1 - \alpha_0 + i_{X_t}\mathrm{d}\alpha_t = 0,$$

which has a unique solution $X_t \in \ker\alpha_t$. □

Nonetheless, the question how to compute vol_R for the Reeb vector field R on a closed contact 3-manifold (M, α) is extremely interesting. Cristofaro-Gardiner, Hutchings and Ramos [9, Theorem 1.2] have established a deep connection between vol_R and embedded contact homology (ECH). For a contact 3-manifold (M, α) with nonzero contact ECH invariant and finite ECH capacities $c_k(M, \alpha)$, $k \in \mathbb{N}_0$, the volume of R can be computed as

$$\mathrm{vol}_R = \lim_{k\to\infty} \frac{c_k(M,\alpha)^2}{2k}.$$

Through this asymptotic formula, vol_R is determined in a subtle way by the periodic Reeb orbits and their actions.

3. Geodesible vector fields and taut foliations

As shown by Wadsley [31], for a nonsingular vector field X, the existence of a 1-form α satisfying conditions (2) and (4) is equivalent to X being geodesible. Here, we briefly recall the proof of this result, since it is essential to our discussion; see also [16,28]. Notice that vol_X is only defined for vector fields X on closed manifolds of odd dimension, but all the considerations about geodesible vector fields in this and the following two sections make sense, unless stated otherwise, for manifolds of arbitrary dimension.

Definition 3.1. (a) A nonsingular vector field X on a manifold M is called *geodesible* if there exists a Riemannian metric on M with respect to which X has unit length and the flow lines of X are geodesics.

(b) A 1-dimensional foliation $\mathcal{F}$ on a manifold M is called *taut* if there exists a Riemannian metric on M for which the leaves of $\mathcal{F}$ (suitably parametrised) are geodesics.

Lemma 3.2. *Let $\big(M, \langle\,.\,,\,.\,\rangle\big)$ be a Riemannian manifold with Levi-Civita connection ∇. Let X be a vector field of unit length, and set $\alpha = \langle X,\,.\,\rangle$. Then,*

$$L_X\alpha = \langle \nabla_X X,\,.\,\rangle.$$

Proof. The claimed identity is a pointwise statement. Locally, one can always extend a tangent vector $Y_p \in T_pM$ to an X-invariant vector field Y, i.e. a vector field satisfying $[X, Y] = 0$. Therefore, it suffices to verify the identity

$$(L_X\alpha)(Y) = \langle \nabla_X X, Y\rangle$$

for such X-invariant vector fields Y. Notice that ∇ being torsion-free then translates into $\nabla_X Y = \nabla_Y X$. Using the fact that the Lie derivative commutes with contraction, we compute

$$
\begin{aligned}
(L_X\alpha)(Y) &= L_X(\alpha(Y)) - \alpha(L_XY) \;=\; L_X(\alpha(Y))\\
&= X\langle X, Y\rangle\\
&= \langle \nabla_X X, Y\rangle + \langle X, \nabla_X Y\rangle\\
&= \langle \nabla_X X, Y\rangle + \langle X, \nabla_Y X\rangle\\
&= \langle \nabla_X X, Y\rangle + \frac{1}{2} Y\langle X, X\rangle\\
&= \langle \nabla_X X, Y\rangle.
\end{aligned}
$$

□

In the following proposition, the equivalence of (i) with (iv) is due to Sullivan [28], who gives an entirely geometric proof. A more formal proof is given in [30, Proposition 6.7]; the proof I give is a little more direct.

Proposition 3.3. *(Wadsley, Sullivan) Let X be a nonsingular vector field on a manifold M. Then, the following are equivalent:*

(i) *X is geodesible;*
(ii) *there exists a 1-form α on M with $\alpha(X) = 1$ and $L_X\alpha = 0$;*
(iii) *there exists a 1-form α on M with $\alpha(X) = 1$ and $i_X \mathrm{d}\alpha = 0$;*
(iv) *there is a hyperplane field η transverse to X and invariant under the flow of X.*

Proof. The equivalence of (ii) and (iii) is clear from the Cartan formula. We first prove the equivalence of (i) and (ii).

Assuming (i), we take $\langle\, .\, ,\, .\,\rangle$ to be the metric for which the flow lines of X are geodesics parametrised by arc length and set $\alpha = \langle X, \,.\,\rangle$. Then, $\nabla_X X = 0$, and (ii) follows from the lemma.

Conversely, given α as in (ii) we choose a metric $\langle\, .\, ,\, .\,\rangle$ on M with $\langle X, X\rangle = 1$ and $X \perp \ker\alpha$. Then, $\alpha = \langle X, \,.\,\rangle$, and the vanishing of $L_X\alpha$ implies, by the lemma, that $\nabla_X X = 0$.

Next, we show the equivalence of (ii) and (iv). Given (ii), the hyperplane field $\eta := \ker\alpha$ satisfies (iv). Conversely, given η as in (iv), define a 1-form α on M by the conditions $\alpha(X) = 1$ and $\ker\alpha = \eta$. Then, $i_X\mathrm{d}\alpha = L_X\alpha$, and the latter equals $f\alpha$ for some $f \in C^\infty(M)$ by the invariance of η. Thus, $i_X\mathrm{d}\alpha$ vanishes on η. Since $TM = \eta \oplus \langle X\rangle$, the Lie derivative $L_X\alpha = i_X\mathrm{d}\alpha$ vanishes identically. □

Example 3.4. The Reeb vector field of a contact form or a stable Hamiltonian structure [7] is geodesible.

The following characterisation of *oriented* taut 1-dimensional foliations, first observed in [28], is then immediate. We write $\mathcal{F} = \langle X\rangle$ with any nonsingular vector field X whose flow lines are the leaves of $\mathcal{F}$.

Proposition 3.5. *The oriented 1-dimensional foliation $\mathcal{F} = \langle X\rangle$ is taut if and only if there is a 1-form α on M with $\alpha(X) > 0$ and $i_X\mathrm{d}\alpha = 0$.*

Proof. If $\mathcal{F} = \langle X\rangle$ is taut, rescale X to a vector field of length 1 with respect to the metric that makes the leaves of $\mathcal{F}$ geodesics. Then, the existence of the desired 1-form α follows from the equivalence of (i) and (iii) in Proposition 3.3.

Conversely, given α, the rescaled vector field $X/\alpha(X)$, which likewise spans $\mathcal{F}$, satisfies (iii) in Proposition 3.3. □

Remark 3.6. (1) Alternatively, one can derive the equivalence of (i) and (iii) in Proposition 3.3 from the identity

$$i_X \mathrm{d}\alpha = \langle \nabla_X X, \,.\, \rangle - \mathrm{d}\big(\langle X, X\rangle/2\big),$$

where again $\alpha = \langle X, \,.\, \rangle$; this identity holds for any vector field X, see [7, Section 2.3].

(2) The main point of Sullivan's article [28] is a characterisation of taut foliations in terms of the absence of "tangent homologies". I refer to [16] for a beautiful discussion of Sullivan's theorem; there one can find examples of 1-dimensional oriented foliations that are not taut.

4. Basic cohomology

Here are the elementary notions of basic differential forms and basic cohomology associated with a foliation. I restrict attention to oriented 1-dimensional foliations $\mathcal{F} = \langle X \rangle$; for a more comprehensive treatment see [30, Chapter 4].

Definition 4.1. A differential form ω on $(M, \mathcal{F})$ is called *basic* if

$$i_X \omega = 0 \quad \text{and} \quad i_X \mathrm{d}\omega = 0.$$

Notice that this definition does not depend on the choice of vector field X spanning $\mathcal{F}$. We write $\Omega^k_{\mathrm{B}}(\mathcal{F})$ for the vector space of basic k-forms on $(M, \mathcal{F})$. The usual exterior differential d restricts to

$$\mathrm{d}_{\mathrm{B}} \colon\; \Omega^k_{\mathrm{B}} \longrightarrow \Omega^{k+1}_{\mathrm{B}},$$

and the *basic cohomology groups* $H^k_{\mathrm{B}}(\mathcal{F})$ are defined as the cohomology groups of the complex $\big(\Omega^\bullet_{\mathrm{B}}(\mathcal{F}), \mathrm{d}_{\mathrm{B}}\big)$. The cohomology class of a k-form $\omega \in \ker \mathrm{d}_{\mathrm{B}}$ is written as $[\omega]_{\mathrm{B}} \in H^k_{\mathrm{B}}(\mathcal{F})$.

The following definitions are motivated by Propositions 3.3 and 3.5. The notation $C_X, C_{\mathcal{F}}$ is chosen because the 1-form $\alpha = \langle X, \,.\, \rangle$ (with X of unit length) is the *characteristic form* of $\mathcal{F}$ [30, p. 69] with respect to the metric $\langle \,.\,, \,.\, \rangle$. We adapt this definition to the case of geodesible vector fields, where it is reasonable to consider only those metrics for which the flow lines of X are geodesics.

Definition 4.2. Let X be a geodesible vector field. Any 1-form α with $\alpha(X) = 1$ and $i_X \mathrm{d}\alpha = 0$ is called a *characteristic* 1-*form* of X. We write char_X for the space of these characteristic forms.

Definition 4.3. (a) Let X be a geodesible vector field on a manifold M. Set

$$\Omega^1_X := \big\{\alpha \in \Omega^1(M) \colon\; \alpha(X) = c \text{ for some } c \in \mathbb{R}^+,\; i_X \mathrm{d}\alpha = 0\big\}$$

and

$$C_X := \Omega^1_X/\!\sim,$$

where

$$\alpha \sim \beta :\Longleftrightarrow \alpha(X) = \beta(X).$$

The equivalence class of $\alpha \in \Omega^1_X$ is written as $[\alpha]_X \in C_X$. Obviously, there is a canonical identification of C_X with $\mathbb{R}^+$.

(b) Let $\mathcal{F} = \langle X \rangle$ be an oriented taut 1-dimensional foliation on M. Set

$$\Omega^1_{\mathcal{F}} := \left\{\alpha \in \Omega^1(M)\colon\ \alpha(X) > 0,\ i_X \mathrm{d}\alpha = 0\right\}$$

and

$$C_{\mathcal{F}} := \Omega^1_{\mathcal{F}}/\sim,$$

where the equivalence relation $\sim$ is defined as in (a). The equivalence class of $\alpha \in \Omega^1_{\mathcal{F}}$ is written as $[\alpha]_{\mathcal{F}}$. Notice that these definitions do not depend on the choice of X.

The assumptions on geodesibility and tautness, respectively, guarantee that we are not talking about empty sets.

The spaces Ω^1_X, char_X and $\Omega^1_{\mathcal{F}}$ are obviously convex. The proof of Proposition 3.3 shows that, for a geodesible vector field X, the map

$$\begin{array}{rcl} \mathrm{Met}_X & \longrightarrow & \mathrm{char}_X \\ \langle\,.\,,\,.\,\rangle & \longmapsto & \alpha = \langle X,\,.\,\rangle \end{array}$$

from the space Met_X of metrics for which X has unit length and geodesic flow lines is a Serre fibration with fibre the space of metrics on a hyperplane field transverse to X, which can be seen as follows. Given a family $\alpha_{q,t} \in \mathrm{char}_X$, where $t \in [0,1]$ and q varies in some parameter space, and a family of metrics $\langle\,.\,,\,.\,\rangle_{q,0}$ with $\langle X,\,.\,\rangle_{q,0} = \alpha_{q,0}$, one simply defines $\langle\,.\,,\,.\,\rangle_{q,t}$ by the following requirements:

(i) $\langle X, X\rangle_{q,t} = 1$;
(ii) $\ker \alpha_{q,t} \perp X$ with respect to $\langle\,.\,,\,.\,\rangle_{q,t}$;
(iii) $\langle\,.\,,\,.\,\rangle_{q,t}|_{\ker \alpha_{q,t}} = \langle\,.\,,\,.\,\rangle_{q,0}|_{\ker \alpha_{q,0}}$ under the identification of $\ker \alpha_{q,t}$ with $\ker \alpha_{q,0}$ given by projection along X.

Of course, this Serre fibration property is not terribly useful, since all spaces in question are contractible.

Proposition 4.4. *Let M be a closed, oriented manifold of dimension m.*

(a) Let X be a geodesible vector field on M. Set $\mathcal{F} = \langle X \rangle$. Then, the map

$$\begin{array}{ccccl} C_X & \times & H^{m-1}_{\mathrm{B}}(\mathcal{F}) & \longrightarrow & \mathbb{R} \\ \big([\alpha]_X & , & [\sigma]_{\mathrm{B}}\big) & \longmapsto & [\alpha]_X \bullet [\sigma]_{\mathrm{B}} := \int_M \alpha \wedge \sigma \end{array}$$

is well defined.

(b) Let $\mathcal{F}$ be an oriented taut 1-dimensional foliation on M. Then, the map

$$\begin{array}{ccccl} C_{\mathcal{F}} & \times & H^{m-1}_{\mathrm{B}}(\mathcal{F}) & \longrightarrow & \mathbb{R} \\ \big([\alpha]_{\mathcal{F}} & , & [\sigma]_{\mathrm{B}}\big) & \longmapsto & [\alpha]_{\mathcal{F}} \bullet [\sigma]_{\mathrm{B}} := \int_M \alpha \wedge \sigma \end{array}$$

is well defined.

Proof. We prove (b); the proof of (a) is completely analogous. Write $\mathcal{F} = \langle X \rangle$ with some nonsingular vector field X spanning $\mathcal{F}$.

(i) We have $i_X\sigma = 0$, since $\sigma \in \Omega^{m-1}_{\mathrm{B}}(\mathcal{F})$. Suppose $[\alpha]_{\mathcal{F}} = [\alpha']_{\mathcal{F}}$, which means that the function $\alpha(X) - \alpha'(X)$ is identically zero. It follows that the m-form $(\alpha - \alpha') \wedge \sigma$ vanishes identically, since its interior product with the nonsingular vector field X vanishes.

(ii) Suppose $[\sigma]_{\mathrm{B}} = [\sigma']_{\mathrm{B}} \in H^{m-1}_{\mathrm{B}}(\mathcal{F})$, that is, $\sigma - \sigma' = \mathrm{d}\tau$ for some $\tau \in \Omega^{m-2}_{\mathrm{B}}(\mathcal{F})$. Then,

$$\begin{aligned}\int_M \alpha \wedge (\sigma - \sigma') &= \int_M \alpha \wedge \mathrm{d}\tau\\ &= -\int_M \mathrm{d}(\alpha \wedge \tau) + \int_M \mathrm{d}\alpha \wedge \tau.\end{aligned}$$

The first summand vanishes by Stokes's theorem; the integrand of the second summand vanishes identically, since $i_X(\mathrm{d}\alpha \wedge \tau) = 0$. □

Observe that the maps defined in this proposition are positively homogeneous of degree 1 on the first factor, and linear in the second factor.

5. The Euler class of a geodesible vector field

Let X be a geodesible vector field on a manifold M and set $\mathcal{F} = \langle X \rangle$. Choose a characteristic 1-form α for X.

Lemma 5.1. *The basic cohomology class $e_X := -[\mathrm{d}\alpha]_{\mathrm{B}} \in H^2_{\mathrm{B}}(\mathcal{F})$ is determined by X.*

Proof. Let β be a further characteristic 1-form. Then, $\gamma := \alpha - \beta \in \Omega^1_{\mathrm{B}}(\mathcal{F})$, and $\mathrm{d}\alpha - \mathrm{d}\beta = \mathrm{d}\gamma = \mathrm{d}_{\mathrm{B}}\gamma$, hence $[\mathrm{d}\alpha]_{\mathrm{B}} = [\mathrm{d}\beta]_{\mathrm{B}}$. □

I do not know whether the following definition has been made before, but it is certainly a very natural one.

Definition 5.2. The class $e_X \in H^2_{\mathrm{B}}(\mathcal{F})$ is called the *Euler class* of the geodesible vector field X.

Example 5.3. (1) If the flow of X generates a principal S^1-action, where we think of S^1 as $\mathbb{R}/\mathbb{Z}$, then e_X can be naturally identified with the real Euler class $e \otimes \mathbb{R} \in H^2(M/S^1; \mathbb{R})$ of the S^1-bundle $M \to M/S^1$. Our definition accords with the usual sign convention, cf. [24, Section 6.2], [11, Section 7.2].

(2) If the flow of X generates a locally free S^1-action, then $H^\bullet_{\mathrm{B}}(\mathcal{F})$ may be thought of as the orbifold cohomology of the orbifold M/S^1, and e_X as the real Euler class of the S^1-orbibundle $M \to M/S^1$. We discuss examples of this kind in detail in Sects. 6 and 7. For more information on S^1-orbibundles in the general sense see [18].

We shall meet further examples in Section 9, where we discuss surfaces of section for the flow of X.

The next lemma is the generalisation of a result for connection 1-forms of principal S^1-bundles.

Lemma 5.4. *Let X be a geodesible vector field and $\omega \in \Omega^2_{\mathrm{B}}(\mathcal{F})$ a basic 2-form with $-[\omega]_{\mathrm{B}} = e_X$. Then, there is a characteristic 1-form β with $\mathrm{d}\beta = \omega$.*

Proof. Since $[\omega]_{\mathrm{B}} = [\mathrm{d}\alpha]_{\mathrm{B}}$, we find a basic 1-form $\gamma \in \Omega^1_{\mathrm{B}}(\mathcal{F})$ with $\omega = \mathrm{d}\alpha + \mathrm{d}\gamma$. Then, $\beta := \alpha + \gamma$ is the desired characteristic form. □

This lemma implies the following proposition.

Proposition 5.5. *A geodesible vector field X on a manifold M of dimension $2n+1$ is the Reeb vector field of a contact form if and only if the Euler class e_X has an odd-symplectic representative, i.e. if there is a closed basic 2-form $\omega \in \Omega^2_{\mathrm{B}}(M)$ with $-[\omega]_{\mathrm{B}} = e_X$ and $\omega^n \neq 0$.*

The following expression of the volume vol_X in terms of the Euler class is immediate from the definitions. This is the promised cohomological interpretation and generalisation of Proposition 1.2.

Proposition 5.6. *Let X be a geodesible vector field on a closed, oriented manifold M of dimension $2n+1$, and α a characteristic form for X. Then,*

$$\mathrm{vol}_X = (-1)^n [\alpha]_X \bullet e_X^n.$$

If X generates a free S^1-action, we have—with $e \in H^2(B;\mathbb{Z})$ denoting the Euler class of the fibration $M \to M/S^1 =: B$—

$$\mathrm{vol}_X = (-1)^n \langle e^n, [B]\rangle,$$

where $[B]$ denotes the fundamental class of B and $\langle\, .\, ,\, .\,\rangle$ the Kronecker pairing.

Here is a useful vanishing criterion for the Euler class. For the flow of X to be globally defined, we assume M to be closed.

Theorem 5.7. *The Euler class $e_X \in H^2_{\mathrm{B}}(\mathcal{F})$ of a geodesible vector field X on a closed manifold M vanishes if and only if X admits a transverse foliation $\mathcal{T}$ invariant under the flow of X.*

Proof. Suppose that $e_X = 0$. As before we write $\mathcal{F} = \langle X\rangle$. Choose a 1-form α with $\alpha(X) = 1$ and $i_X\mathrm{d}\alpha = 0$. Then, $[\mathrm{d}\alpha]_{\mathrm{B}} = -e_X = 0$, so there is a basic 1-form $\gamma \in \Omega^1_{\mathrm{B}}(\mathcal{F})$ with $\mathrm{d}\gamma = \mathrm{d}\alpha$. Then, $\beta := \alpha - \gamma$ is a closed 1-form with $\beta(X) = 1$. In particular, $\ker\beta$ defines a foliation $\mathcal{T}$ transverse to X, and $\mathcal{T}$ is invariant under the flow of X since $L_X\beta = \mathrm{d}(\beta(X)) + i_X\mathrm{d}\beta = 0$.

Conversely, let $\mathcal{T}$ be a transverse invariant foliation. Define a 1-form α by $\alpha(X) = 1$ and $\ker\alpha = T\mathcal{T}$, where $T\mathcal{T}$ denotes the distribution of tangent spaces to $\mathcal{T}$. Then, $i_X\mathrm{d}\alpha = L_X\alpha$, and the latter equals $f\alpha$ for some $f \in C^\infty(M)$ by the invariance of $\mathcal{T}$. This implies $\mathrm{d}\alpha(X,Y) = 0$ for $Y \in \Gamma(T\mathcal{T})$.

Given two (local) vector fields $Y_1, Y_2 \in \Gamma(T\mathcal{T})$, we compute

$$\mathrm{d}\alpha(Y_1,Y_2) = Y_1\alpha(Y_2) - Y_2\alpha(Y_1) - \alpha([Y_1,Y_2]) = 0.$$

Thus, we conclude that $\mathrm{d}\alpha = 0$, and hence $e_X = 0$. □

Corollary 5.8. *A closed manifold M admits a geodesible vector field X with $e_X = 0$ if and only if M fibres over S^1.*

Proof. If M admits a geodesible vector field with $e_X = 0$, the fact that M fibres over S^1 follows from the existence of a closed, nonsingular 1-form on M, established in the foregoing proof, and a result of Tischler [29], cf. [8, Section 9.3].

Conversely, a manifold M that fibres over S^1 always admits a geodesible vector field [16]. Such a manifold M can be written as $[0,1] \times F/(1,x) \sim (0,\psi(x))$, where F denotes the fibre and ψ the monodromy of the bundle. Let g_θ, $\theta \in [0,1]$ be any smooth family of metrics on F with $\psi^* g_0 = g_1$. Then, $d\theta^2 + g_\theta$ defines a metric on $[0,1] \times F$ for which the segments $[0,1] \times \{x\}$ are geodesics, and this metric descends to M.

Alternatively, let α be the pull-back of the 1-form $\mathrm{d}\theta$ under the bundle projection $M \to S^1$. Then, $\mathrm{d}\alpha = 0$, so any vector field X on M with $\alpha(X) = 1$, i.e. any lift of ∂_θ, is geodesible, and clearly $e_X = 0$. $\square$

Example 5.9. For Seifert fibred 3-manifolds (see the next section), the statement of Corollary 5.8 can be found in [26, Theorem 5.4].

6. Seifert fibred 3-manifolds

In this section, we take $M \to B$ to be a Seifert fibration of a closed, oriented 3-manifold M over a closed, oriented 2-dimensional orbifold B. Let X be the vector field whose flow defines an S^1-action on M with orbits equal to the Seifert fibres, where the minimal period of the regular fibres is assumed to be equal to 1. I refer to [15] and [17] for the basic terminology of Seifert fibrations.

Suppose the Seifert invariants of $M \to B$ are

$$\big(g; (\alpha_1, \beta_1), \ldots, (\alpha_n, \beta_n)\big),$$

where $g \in \mathbb{N}_0$ is the genus of B, and the (α_i, β_i), $i = 1, \ldots, n$, are pairs of coprime integers with $\alpha_i \neq 0$. Here the α_i give the multiplicities of the singular fibres; the pairs with $\alpha_i = 1$ do not correspond to singular fibres, but contribute to the Euler class of the fibration.

Concretely, M is recovered from these Seifert invariants as follows. Let B be the closed, oriented surface of genus g, and remove n disjoint discs to obtain

$$B_0 = B \setminus \operatorname{Int}\big(D_1^2 \sqcup \ldots \sqcup D_n^2\big).$$

Over this surface with boundary, we take the trivial S^1-bundle $M_0 = B_0 \times S^1 \to B_0$. Write the boundary ∂B_0 with the opposite of its natural orientation as

$$-\partial B_0 = S_1^1 \sqcup \ldots \sqcup S_n^1.$$

We write the fibre class of this trivial fibration as $h = \{*\} \times S^1$, and on ∂M_0 we consider the curves

$$q_i = S_i^1 \times \{0\},\ i = 1, \ldots, n;$$

recall that we think of the fibre S^1 as $\mathbb{R}/\mathbb{Z}$. The labels $h, q_1, \ldots, q_n$ should be read as isotopy classes of curves on ∂M_0.

Choose integers α_i', β_i', $i = 1, \dots, n$, such that

$$\begin{vmatrix} \alpha_i & \alpha_i' \\ \beta_i & \beta_i' \end{vmatrix} = 1.$$

Further, take n copies $V_i = D^2 \times S^1$ of a solid torus, where D^2 is the unit disc in $\mathbb{R}^2$, with respective meridian and longitude

$$\mu_i = \partial D^2 \times \{0\}, \quad \lambda_i = \{1\} \times S^1 \subset \partial V_i.$$

Then, glue the V_i to M_0 along the boundary, where ∂V_i is identified with the component $S_i^1 \times S^1$ of ∂M_0 via

$$h = -\alpha_i' \mu_i + \alpha_i \lambda_i, \quad q_i = \beta_i' \mu_i - \beta_i \lambda_i. \tag{6}$$

Notice that the fibration of M_0 given by the fibre class h extends to a fibration of V_i with the central fibre of multiplicity α_i. This is the description of $M \to B$ with the given Seifert invariants.

The Euler number e of the Seifert fibration with the given Seifert invariants, defined as the obstruction to the existence of a section (in the Seifert sense) [17, Section 3], is

$$e = -\sum_{i=1}^{n} \frac{\beta_i}{\alpha_i}.$$

We now want to use a global surface of section (in a slightly generalised sense) to derive this formula.

Proposition 6.1. *Let X be a vector field on a closed, oriented 3-manifold M defining a Seifert fibration of regular period 1 with invariants*

$$\big(g; (\alpha_1, \beta_1), \dots, (\alpha_n, \beta_n)\big).$$

Then,

$$\langle e_X, [B] \rangle = -\sum_{i=1}^{n} \frac{\beta_i}{\alpha_i},$$

where $\langle\, .\, ,\, .\, \rangle$ denotes the Kronecker pairing between $H^2_{\mathrm{B}}(\mathcal{F})$ and $H_2(B)$.

Recall that a *global surface of section* (s.o.s.) for the flow of X is an embedded compact surface $\Sigma \subset M$ whose boundary consists of orbits of X, whose interior $\mathrm{Int}(\Sigma)$ is transverse to X, and such that the flow line of X through any point not on $\partial\Sigma$ hits $\mathrm{Int}(\Sigma)$ in forward and backward time. We now describe such an s.o.s. for the situation at hand.

Proof of Proposition 6.1. In M_0, we can take $B_0 \times \{0\}$ as section. The boundary of this section consists of the curves $-q_i$, $i = 1, \dots, n$, which are identified with $-\beta_i' \mu_i + \beta_i \lambda_i$ on ∂V_i. Thus, by isotoping these respective curves radially towards the spine $\sigma_i = \{0\} \times S^1$ of V_i, we sweep out a surface Σ that is not quite an s.o.s. in the sense of the definition above, but which has the following properties:

– the inclusion $\Sigma \subset M$ is an embedding on $\mathrm{Int}(\Sigma)$;
– the boundary of Σ is made up of the curves σ_i, each covered β_i times;

– the interior $\mathrm{Int}(\Sigma)$ is intersected positively in a single point by each X-orbit different from the σ_i.

We now choose a specific connection 1-form α on $M \to B$, i.e. a characteristic 1-form for X. On $V_i = D^2 \times S^1$ we write $(r, 2\pi\phi)$ for polar coordinates on the D^2-factor, and $\theta \in S^1 = \mathbb{R}/\mathbb{Z}$. It follows from (6) that on V_i we may assume X to be given by $-\alpha_i'\partial_\phi + \alpha_i\partial_\theta$. So we choose the 1-form α equal to $\alpha = \mathrm{d}\theta/\alpha_i$ near the spine of V_i, and then extend arbitrarily as a connection form over M (using a partition of unity).

We then compute

$$\langle e_X, [B]\rangle = -\int_B \mathrm{d}\alpha = -\int_\Sigma \mathrm{d}\alpha = -\int_{\partial\Sigma} \alpha = -\sum_{i=1}^n \frac{\beta_i}{\alpha_i}.$$

Notice that the integral $\int_B \mathrm{d}\alpha$ is well defined, since $\mathrm{d}\alpha$ is a basic form. □

Remark 6.2. As in Proposition 5.5, one argues that if the Euler number e of the Seifert fibration is nonzero, one can choose α as a contact form (defining the correct orientation of M if $e < 0$, the opposite one if $e > 0$).

Corollary 6.3. *The volume* vol_X *of a vector field* X *defining a Seifert fibration on a closed, orientable* 3*-manifold, with the regular fibres having minimal period* 1*, equals minus the Euler number of that Seifert fibration. In particular, with* m *denoting the least common multiple of the multiplicities* $\alpha_1, \ldots, \alpha_n$*, we have that* $m \cdot \mathrm{vol}_X$ *is an integer.*

Proof. The value of the integral of $\alpha \wedge \mathrm{d}\alpha$ over M does not change when we remove the singular fibres of the Seifert fibration. But then the integral equals

$$\int_{\mathrm{Int}(\Sigma)\times S^1} \alpha \wedge \mathrm{d}\alpha = \int_\Sigma \mathrm{d}\alpha = -e.$$

□

Remark 6.4. The integrality statement has been observed in greater generality by Weinstein [32].

Example 6.5. The positive Hopf fibration

$$\begin{array}{ccccc} \mathbb{C}^2 \supset & S^3 & \longrightarrow & \mathbb{CP}^1 & = S^2 \\ & (z_1, z_2) & \longmapsto & [z_1 : z_2] & \end{array}$$

is given by the vector field $X = 2\pi(\partial_{\varphi_1} + \partial_{\varphi_2})$ of period 1, where $\varphi_1, \varphi_2 \in \mathbb{R}/2\pi\mathbb{Z}$. The corresponding connection 1-form is

$$\alpha = \frac{1}{2\pi}\big(r_1^2\,\mathrm{d}\varphi_1 + r_2^2\,\mathrm{d}\varphi_2\big).$$

With $r^2 = r_1^2 + r_2^2$, one computes

$$r\,\mathrm{d}r \wedge \alpha \wedge \mathrm{d}\alpha = \frac{1}{2\pi^2}(r_1^2 + r_2^2)\cdot(r_1\,\mathrm{d}r_1 \wedge \mathrm{d}\varphi_1 \wedge r_2\,\mathrm{d}r_2 \wedge \mathrm{d}\varphi_2).$$

So along the unit sphere $S^3 = \{r = 1\}$, the 3-form $\alpha \wedge \mathrm{d}\alpha$ restricts to the standard volume form up to a factor $1/2\pi^2$, hence

$$\int_{S^3} \alpha \wedge \mathrm{d}\alpha = \frac{1}{2\pi^2}\mathrm{Vol}(S^3) = 1.$$

A section of the Hopf fibration over $\mathbb{C} \cong \mathbb{CP}^1 \setminus \{[0:1]\}$ is defined by

$$r\mathrm{e}^{\mathrm{i}\varphi} \longmapsto \left[1 : r\mathrm{e}^{\mathrm{i}\varphi}\right] \longmapsto \left(\frac{1}{\sqrt{1+r^2}}, \frac{r\mathrm{e}^{\mathrm{i}\varphi}}{\sqrt{1+r^2}}\right).$$

Under this map, $\mathrm{d}\alpha$ pulls back to

$$\frac{1}{\pi} \cdot \frac{r}{(1+r^2)^2}\,\mathrm{d}r \wedge \mathrm{d}\varphi.$$

This yields

$$\int_{S^2} \mathrm{d}\alpha = \frac{1}{\pi}\int_0^{2\pi}\int_0^{\infty} \frac{r}{(1+r^2)^2}\,\mathrm{d}r\,\mathrm{d}\varphi = 1.$$

Thus, both computations confirm that the positive Hopf fibration has Euler number $e = -1$, see also [5, Lemma 2.2].

7. The theorems of Gauß–Bonnet and Poincaré–Hopf

In this section, we formulate and prove the theorems of Gauß–Bonnet and Poincaré–Hopf for oriented 2-dimensional orbifolds, using an s.o.s. argument as in the preceding section. Versions of these theorems for higher-dimensional orbifolds (including those with boundary) can be found in [25] and [27]. To avoid confusion with formulas found elsewhere, in this section, we follow the usual convention that the regular fibres in the unit tangent bundle of a 2-dimensional orbifold have length 2π.

Thus, let B be a closed, oriented 2-dimensional Riemannian orbifold with underlying surface of genus g and n cone points of multiplicities $\alpha_1, \ldots, \alpha_n$. There is a well-defined unit tangent bundle STB, cf. [13], which is a Seifert manifold with invariants

$$\big(g, (1, 2g-2), (\alpha_1, \alpha_1 - 1), \ldots, (\alpha_n, \alpha_n - 1)\big).$$

The orbifold Euler characteristic $\chi_{\mathrm{orb}}(B)$ is the Euler number of the Seifert fibration $\pi\colon STB \to B$, so by Proposition 6.1, we have

$$\chi_{\mathrm{orb}}(B) = 2 - 2g - n + \sum_{i=1}^{n} \frac{1}{\alpha_i}.$$

This formula can also be derived combinatorially, using the Riemann–Hurwitz formula for coverings, see [26, p. 427].

Just like the unit tangent bundle of a smooth surface, the unit tangent bundle STB of an orbifold admits a pair of Liouville–Cartan forms λ_1, λ_2 and a connection 1-form $\tilde{\alpha}$ satisfying the structure equations

$$\begin{aligned}
\mathrm{d}\lambda_1 &= -\lambda_2 \wedge \tilde{\alpha},\\
\mathrm{d}\lambda_2 &= -\tilde{\alpha} \wedge \lambda_1,\\
\mathrm{d}\tilde{\alpha} &= -(\pi^* K)\lambda_1 \wedge \lambda_2,
\end{aligned}$$

where K is the Gauß curvature of the Riemannian metric on B. See [3, Section 2.1] for the surface case, and [12, Section 7] for a discussion of Liouville–Cartan forms for orbifolds.

Theorem 7.1. (Gauß–Bonnet) *The total curvature of a closed, oriented* 2*-dimensional Riemannian orbifold* B *equals*

$$\int_B K\,\mathrm{d}A = 2\pi\chi_{\mathrm{orb}}(B).$$

Proof. The characteristic 1-form α for the vector field X that makes the regular fibres of STB of length 1 is $\alpha = \tilde{\alpha}/2\pi$. Therefore, with $e = \chi_{\mathrm{orb}}(B)$, we obtain

$$\int_B K\,\mathrm{d}A = -\int_\Sigma \mathrm{d}\tilde{\alpha} = -2\pi\int_\Sigma \mathrm{d}\alpha = 2\pi\chi_{\mathrm{orb}}(B),$$

where Σ is as in the proof of Proposition 6.1. □

Now, let Y be a vector field with isolated zeros on the orbifold B. To formulate the Poincaré–Hopf theorem we need to give a definition of the index $\mathrm{ind}_p Y$ in an orbifold singularity $p \in B$, cf. [25, Section 3.2]. First, let $p \in B$ be a smooth point where Y has a zero. Choose a small disc $D_\varepsilon(p) \subset B$ not containing other zeros of Y (and hence, as we shall see, in particular no orbifold points of B). Choose a trivialisation $TD_\varepsilon(p) \cong D_\varepsilon(p) \times \mathbb{R}^2$. Then, $\mathrm{ind}_p Y$ is the degree of the map $\partial D_\varepsilon(p) \to S^1$, $x \mapsto Y(x)/|Y(x)|$.

When the zero of Y happens to be an orbifold singularity $p_i \in B$ of order α_i, we consider a local description $\pi_{\alpha_i}\colon D^2 \to D^2/\mathbb{Z}_{\alpha_i} \cong D^2$ of the singularity, where the cyclic group $\mathbb{Z}_{\alpha_i}$ is generated by the rotation about $0 \in \mathbb{R}^2$ through an angle $2\pi/\alpha_i$.

We drop the index i for the time being; there should be little grounds for confusing α in the following discussion with the connection 1-form.

The fibre of STB over the singular point $p \in B$ has Seifert invariants $(\alpha, \beta = \alpha - 1)$, so we may take $\alpha' = \beta' = 1$. Then, cf. [15], the local description of the fibration $STB \to B$ near the orbifold point p is given by

$$\begin{array}{rccc}
\pi\colon & D^2 \times S^1 & \longrightarrow & D^2\\
 & \bigl(r\mathrm{e}^{\mathrm{i}\varphi}, \mathrm{e}^{\mathrm{i}\theta}\bigr) & \longmapsto & r\mathrm{e}^{\mathrm{i}(\alpha\varphi+\theta)},
\end{array}$$

where we identify p with $0 \in D^2$. Notice that the fibres of π are described by $\alpha\varphi + \theta = \mathrm{const.}$, or in parametric form as

$$t \longmapsto \bigl(\varphi(t), \theta(t)\bigr) = (\varphi_0 - t, \theta_0 + \alpha t).$$

This accords with (6).

Now, consider the following commutative diagram,

$$\begin{tikzcd}
D^2 \times S^1 \arrow[d, "\tilde{\pi}"'] & \ni & \left(r\mathrm{e}^{\mathrm{i}\tilde{\varphi}}, \mathrm{e}^{\mathrm{i}\tilde{\theta}}\right) \arrow[r, mapsto, "\tilde{\pi}_\alpha"] \arrow[d, mapsto] & \left(r\mathrm{e}^{\mathrm{i}\varphi}, \mathrm{e}^{\mathrm{i}\theta}\right) \arrow[d, mapsto] & \in & D^2 \times S^1 \arrow[d, "\pi"] \\
D^2 & \ni & r\mathrm{e}^{\mathrm{i}(\tilde{\varphi}+\tilde{\theta})} \arrow[r, mapsto, "\pi_\alpha"] & r\mathrm{e}^{\mathrm{i}(\alpha\varphi+\theta)} & \in & D^2,
\end{tikzcd}$$

where the quotient map π_α under the $\mathbb{Z}_\alpha$-action is given by

$$\pi_\alpha(r\mathrm{e}^{\mathrm{i}t}) = r\mathrm{e}^{\mathrm{i}\alpha t},$$

its lift $\tilde{\pi}_\alpha$ to the unit tangent bundle by

$$\tilde{\pi}_\alpha\colon\ (\tilde{\varphi}, \tilde{\theta}) \longmapsto (\varphi, \theta) = (\tilde{\varphi}, \alpha\tilde{\theta}).$$

Up to homotopy, the section $Y/|Y|$ of π over ∂D^2 is of the form

$$\big(\varphi(t), \theta(t)\big) = \big(2\pi k t, 2\pi(1-k\alpha)t\big), \quad t \in [0,1], \tag{7}$$

for some $k \in \mathbb{Z}$; notice that $\alpha\varphi(t) + \theta(t)$ goes from 0 to 2π as t goes from 0 to 1. The lift of the α-fold traversal of this curve under the map $\tilde{\pi}_\alpha$ is described by

$$\big(\tilde{\varphi}(t), \tilde{\theta}(t)\big) = \big(2\pi k t, 2\pi\frac{1-k\alpha}{\alpha}\,t\big), \quad t \in [0,\alpha]; \tag{8}$$

here $\tilde{\varphi}(t) + \tilde{\theta}(t) = \frac{2\pi}{\alpha}\,t$ goes from 0 to 2π as t goes from 0 to α.

The fibres of $\tilde{\pi}$ are described by $\tilde{\varphi}+\tilde{\theta} = \text{const.}$, and a single right-handed Dehn twist along a meridional disc of $D^2 \times S^1$,

$$(\tilde{\varphi}', \tilde{\theta}') := (\tilde{\varphi} + \tilde{\theta}, \tilde{\theta}),$$

will bring these fibres into the form $\tilde{\varphi}' = \tilde{\varphi}'_0$, and the curve (8) becomes

$$\big(\tilde{\varphi}'(t), \tilde{\theta}'(t)\big) = \Big(\frac{2\pi}{\alpha}\,t, 2\pi\frac{1-k\alpha}{\alpha}\,t\Big), \quad t \in [0,\alpha].$$

The index $\mathrm{ind}_p Y$ at an orbifold point of multiplicity α is defined as $\mathrm{ind}_{\tilde{p}}\widetilde{Y}/\alpha$, where $\tilde{p} = \pi_\alpha^{-1}(p)$ and $\widetilde{Y}$ is the lifted vector field. Our considerations show that, in dependence on $k \in \mathbb{Z}$, this index is

$$\mathrm{ind}_p Y = \frac{1}{\alpha} - k.$$

Notice that $k = 0$ corresponds to a rotationally symmetric source or sink of Y, which lifts to an identically looking zero of $\widetilde{Y}$. Also, for $\alpha > 1$ the term $1 - k\alpha$ in the θ-component of (7) never equals zero, no matter what $k \in \mathbb{Z}$, which means that orbifold points always must be zeros of Y.

Theorem 7.2. (Poincaré–Hopf) *Let Y be a vector field with isolated zeros on a closed, oriented 2-dimensional orbifold B. Then,*

$$\sum_{\substack{p \in B \\ Y(p)=0}} \mathrm{ind}_p Y = \chi_{\mathrm{orb}}(B).$$

Proof. The idea is simply to compute the Euler number $e = \chi_{\mathrm{orb}}(B)$ of the Seifert fibration $STB \to B$ with the help of an s.o.s. Σ^Y adapted to Y.

Outside small disc neighbourhoods of the zeros of Y we may normalise the vector field and regard it as a section Σ_0^Y of $STB \to B$ outside this set of discs in B. This surface Σ_0^Y extends to an s.o.s. Σ^Y of STB, with boundary components certain multiple covers of the fibres over the zeros p of Y, as in the proof of Proposition 6.1. The multiplicity of the covering is determined by the number of full turns the boundary component makes in the θ-direction. Notice that the orientation of the collection of circles $\pi(\partial\Sigma_0^Y)$ is the opposite of the orientation as boundaries of the removed discs. Thus, the multiplicity is $-\mathrm{ind}_p Y$ at a smooth point and, by (8), equal to $-(1-k_i\alpha_i) = -\alpha_i\,\mathrm{ind}_p Y$ at an orbifold point of order α_i, where $k_i \in \mathbb{Z}$ is the integer describing that particular zero of Y, see also Remark 7.3.

With a connection 1-form α corresponding to regular fibres having length 1 as in the proof of Proposition 6.1, that is, equal to $\mathrm{d}\theta/2\pi$ near the fibres over smooth zeros of Y, and equal to $\mathrm{d}\theta/2\pi\alpha_i$ over an orbifold point of order α_i, we have

$$\chi_{\mathrm{orb}}(B) = e = -\int_{\Sigma^Y} \mathrm{d}\alpha = -\int_{\partial\Sigma^Y} \alpha = \sum_{\substack{p\in B\\ Y(p)=0}} \mathrm{ind}_p Y.$$

□

Remark 7.3. It may be helpful to reformulate the first part of the proof in terms of meridians and longitudes, similar to the discussion of the topology of surfaces of section in [5].

First, consider a zero of Y at a smooth point $p \in B$. Let V be a tubular neighbourhood of the fibre ST_pB. Let μ be the meridian on ∂V, and λ the longitude determined by the parallel fibres. We orient λ as the fibres, and μ in such a way that (μ, λ) gives the positive orientation of ∂V. The component of $\partial\Sigma_0^Y$ on ∂V is $(-1, -\mathrm{ind}_p Y)$ in terms of the (μ, λ)-basis. Therefore, this component is isotopic to $-\mathrm{ind}_p Y$ times the spine of V. In addition, notice that the intersection number of the fibre $(0, 1)$ with $(-1, -\mathrm{ind}_p Y)$ is $+1$, which is consistent with Σ_0^Y being a section.

For a zero of Y at a singular point of order α_i, we take a neighbourhood V_i of $ST_{p_i}B$ with μ_i, λ_i as in Section 6. Now, the component of $\partial\Sigma_0^Y$ on ∂V_i is $(-k_i, -1+k_i\alpha_i)$ by (7), which is isotopic to $-1+k_i\alpha_i$ times the spine. Again, the intersection of the fibre $(-1, \alpha)$, see (6), with $(-k_i, -1+k_i\alpha_i)$ is $+1$.

8. Transversely holomorphic foliations and the Bott invariant

In [14] with Jesús Gonzalo, we proved a generalised Gauß–Bonnet theorem for transversely holomorphic 1-dimensional foliations on 3-manifolds. In certain situations, which I am going to describe now, this can be interpreted as a statement about vol_X for a vector field X whose flow defines such a foliation.

The following definition is from [12].

Definition 8.1. A pair of contact forms (ω_1, ω_2) on a closed, oriented 3-manifold M is called a *Cartan structure* if

$$\begin{aligned}\omega_1 \wedge \mathrm{d}\omega_1 = \omega_2 \wedge \mathrm{d}\omega_2 \neq 0\\ \omega_1 \wedge \mathrm{d}\omega_2 = \omega_2 \wedge \mathrm{d}\omega_1 = 0.\end{aligned}$$

Such structures exist in abundance, see [12, Theorem 1.2]. They are special cases of what we christened taut contact circles in that paper: any linear combination $\lambda_1\omega_1 + \lambda_2\omega_2$ with $(\lambda_1, \lambda_2) \in S^1 \subset \mathbb{R}^2$ is again a contact form defining the same volume form. The defining equations for a Cartan structure can be rephrased as saying that there is a uniquely defined nowhere vanishing 1-form α such that

$$\begin{aligned}\mathrm{d}\omega_1 &= \omega_2 \wedge \alpha,\\ \mathrm{d}\omega_2 &= \alpha \wedge \omega_1.\end{aligned}$$

In terms of the complex-valued 1-form $\omega_{\mathrm{c}} := \omega_1 + \mathrm{i}\,\omega_2$, these equations can be rewritten as

$$\mathrm{d}\omega_{\mathrm{c}} = \mathrm{i}\,\alpha \wedge \omega_{\mathrm{c}}.$$

Observe that

$$0 \neq \omega_1 \wedge \mathrm{d}\omega_1 = \omega_1 \wedge \omega_2 \wedge \alpha,$$

so α is nonzero on the common kernel of ω_1 and ω_2.

Lemma 8.2. *Let X be the vector field defined by $X \in \ker\omega_1 \cap \ker\omega_2$ and $\alpha(X) = 1$. Then, $i_X\mathrm{d}\alpha = 0$. Hence, by Proposition 3.3, X is geodesible.*

Proof. By taking the exterior derivative of the defining equations for α we find

$$0 = \mathrm{d}^2\omega_1 = \mathrm{d}\omega_2 \wedge \alpha - \omega_2 \wedge \mathrm{d}\alpha = -\omega_2 \wedge \mathrm{d}\alpha,$$

and similarly

$$0 = \mathrm{d}\alpha \wedge \omega_1.$$

This implies that $i_X\mathrm{d}\alpha$ must be a multiple both of ω_1 and ω_2, but these forms are pointwise linearly independent. □

The 1-form ω_{c} is formally integrable in the sense that $\omega_{\mathrm{c}} \wedge \mathrm{d}\omega_{\mathrm{c}} = 0$. In [14] it is shown that this is equivalent to saying that ω_{c} defines a transverse holomorphic structure for the 1-dimensional foliation defined by the flow of X.

In general, the formal integrability of a complex-valued 1-form ω_{c} only implies the existence of a (not uniquely defined) *complex-valued* 1-form α_{c} such that

$$\mathrm{d}\omega_c = \alpha_{\mathrm{c}} \wedge \omega_{\mathrm{c}}.$$

A Godbillon–Vey type argument shows that the complex number

$$\int_M \alpha_{\mathrm{c}} \wedge \mathrm{d}\alpha_{\mathrm{c}},$$

called the *Bott invariant*, is an invariant of the transversely holomorphic foliation that does not depend on the choice of ω_{c} and α_{c}.

The generalised Gauß–Bonnet theorem [14, Theorem 3.3] says that for transversely holomorphic foliations coming from a Cartan structure, this Bott invariant depends only on the 1-dimensional foliation defined by the common kernel flow, not on the specific transverse holomorphic structure. As observed before, for ω_{c} coming from a Cartan structure we can take $\alpha_{\mathrm{c}} = \mathrm{i}\alpha$. Thus, the generalised Gauß–Bonnet theorem from [14] can be rephrased as follows.

Theorem 8.3. *If the vector field X derives from a Cartan structure as described, then* vol_X *equals the negative of the Bott invariant of any transversely holomorphic structure on the foliation $\langle X\rangle$.*

The paper [14] contains examples which show this to be a nontrivial statement. There are instances of the generalised Gauß–Bonnet theorem where the transverse holomorphic structure is indeed not unique. In [14], one can also find a complete classification of the transversely holomorphic foliations on S^3, originally due (for all 3-manifolds) to Brunella and Ghys, and a computation of their Bott invariant.

9. Global surfaces of section

We now want to compute vol_X under the assumption that the geodesible vector field X admits a global surface of section $\Sigma \subset M$. For simplicity, we assume that M is a closed, oriented manifold of dimension 3, although our considerations extend in an obvious manner to global hypersurfaces of section in manifolds of higher odd dimension for an appropriate definition of that concept.

Given such an s.o.s., we can associate with each point $p \in \mathrm{Int}(\Sigma)$ its *return time* $\tau(p) \in \mathbb{R}^+$, i.e. the smallest positive real number with $\phi_{\tau(p)}(p) \in \mathrm{Int}(\Sigma)$, were ϕ_t denotes the flow of X.

Proposition 9.1. *Let σ be a basic 2-form on M that represents the Euler class e_X. Then,*

$$\mathrm{vol}_X = -\int_{\mathrm{Int}(\Sigma)} \tau\sigma,$$

where we interpret σ as a 2-form on the transversal $\mathrm{Int}(\Sigma)$ *for the flow of X.*

Proof. Let α be a characteristic 1-form of X. By Proposition 4.4, we have

$$\mathrm{vol}_X = \int_M \alpha \wedge \mathrm{d}\alpha = -\int_M \alpha \wedge \sigma.$$

To compute the integral on the right, we consider the injective immersion

$$\begin{array}{rccc} \Phi\colon & [0,1) \times \mathrm{Int}(\Sigma) & \longrightarrow & M \\ & (t \quad , \quad p) & \longmapsto & \phi_{t\tau(p)}(p). \end{array}$$

Since $T\Phi(\partial_t)$ is a multiple of X, and σ a basic differential form, we can compute

$$\begin{aligned}\int_M \alpha\wedge\sigma &= \int_{M\setminus\partial\Sigma} \alpha\wedge\sigma\\ &= \int_{[0,1)\times\mathrm{Int}(\Sigma)} \Phi^*(\alpha\wedge\sigma)\\ &= \int_{\mathrm{Int}(\Sigma)} \left(\int_0^1 \left(\Phi^*\alpha\right)_{(t,p)}(\partial_t)\,\mathrm{d}t\right)\sigma\\ &= \int_{\mathrm{Int}(\Sigma)} \tau\sigma.\end{aligned}$$

In the last line, we used that

$$\left(\Phi^*\alpha\right)_{(t,p)}(\partial_t) = \alpha_{\Phi(t,p)}\big(T\Phi(\partial_t)\big) = \alpha_{\Phi(t,p)}\big(\tau(p)X\big) = \tau(p).$$

Hence, $\mathrm{vol}_X = -\int_{\mathrm{Int}(\Sigma)} \tau\sigma$, as claimed. □

Example 9.2. On D^2 with polar coordinates (r,φ) we write $\lambda = r^2\,\mathrm{d}\varphi/2$ for the primitive 1-form of the standard area form $\omega = \mathrm{d}\lambda = r\,\mathrm{d}r\wedge\mathrm{d}\varphi$. On $\mathbb{R}/\mathbb{Z}\times D^2$, we consider the 1-form

$$\alpha = H\,\mathrm{d}\theta + \lambda,$$

where H is a smooth function of r^2. In the sequel, it will always be understood that H or its derivative H' is evaluated at r^2. Then,

$$\mathrm{d}\alpha = 2rH'\,\mathrm{d}r\wedge\mathrm{d}\theta + \omega$$

and

$$\alpha\wedge\mathrm{d}\alpha = (H - r^2H')\,\mathrm{d}\theta\wedge\omega.$$

We assume that $H - r^2H' > 0$; then α is a contact form. As discussed in [4], this 1-form descends to a contact form (still denoted α) on S^3, obtained from $S^1\times D^2$ by collapsing the circle action on the boundary $S^1\times\partial D^2$ generated by

$$\partial_\theta - 2H(1)\partial_\varphi \in \ker\alpha|_{T(S^1\times\partial D^2)}.$$

The Reeb vector field of α (on $S^1\times D^2$) is

$$X = \frac{\partial_\theta - 2H'\partial_\varphi}{H - r^2H'}. \tag{9}$$

Thus,

$$\mathrm{vol}_X = \int_{S^3}\alpha\wedge\mathrm{d}\alpha = \int_{S^1\times D^2}\alpha\wedge\mathrm{d}\alpha = \int_{D^2}(H - r^2H')\omega.$$

On the other hand, the disc $\{0\}\times D^2$ descends to an s.o.s. for the Reeb flow on S^3, and by (9) the return time is

$$\tau = H - r^2H'.$$

Therefore, we see that vol_X can likewise be computed as

$$\mathrm{vol}_X = \int_{D^2} \tau\sigma$$

with $\sigma = \mathrm{d}\alpha|_{TD^2}$ or any other 2-form that differs from $\mathrm{d}\alpha$ by the differential of a basic 1-form for X on S^3 (not just on $S^1 \times D^2$).

Remark 9.3. For an expression of vol_X in the preceding example in terms of the Calabi invariant of the return map on the s.o.s., see [2].

10. Contact forms with the same Reeb vector field

In this section, we present examples of nondiffeomorphic contact forms with the same Reeb vector field.

Theorem 10.1. *In any odd dimension ≥ 9, there is a closed manifold admitting a countably infinite family of contact forms that are pairwise nondiffeomorphic but share the same Reeb vector field.*

Proof. We construct these manifolds as Boothby–Wang bundles [6], [11, Section 7.2] over integral symplectic manifolds. Starting point for our construction are examples of symplectic manifolds, in any even dimension ≥ 8, with cohomologous but nondiffeomorphic symplectic forms, devised by McDuff [22]. In dimension eight, one begins with the manifold $S^2 \times T^2 \times S^2 \times S^2$ with the standard split symplectic form. We think of T^2 as $(\mathbb{R}/\mathbb{Z})^2$. One then twists this symplectic form by a diffeomorphism

$$(p_1; s_2, t_2; p_3; p_4) \longmapsto \big(p_1; s_2, t_2, \psi_k(p_1, t_2)(p_3); p_4\big),$$

where $\psi_k(p_1, t_2)\colon\ S^2 \to S^2$ is the rotation of S^2 about the axis determined by $\pm p_1$ through an angle $2\pi k t_2$. Finally, one takes the symplectic blow-up of these forms along $S^2 \times T^2 \times \{(p_3, p_4)\}$ with the same blow-up parameter (giving the 'size' of the blow-up) for all $k \in \mathbb{N}_0$.

The resulting symplectic forms ω_k on the blown-up manifold W are cohomologous and homotopic through (noncohomologous) symplectic forms, but they are pairwise nondiffeomorphic. By taking products with copies of S^2, one obtains similar examples in higher dimensions.

The cohomology class of the symplectic form on a manifold obtained as a blow-up has been computed in [21], and from there one sees that the blow-up can be chosen in such a way that this cohomology class is rational. Hence, after a constant rescaling, we may assume the symplectic forms ω_k to be integral, i.e. their de Rham cohomology class $[\omega_k]$ lies in the image of the inclusion $H^2(W;\mathbb{Z}) \subset H^2(W;\mathbb{R}) = H^2_{\mathrm{dR}}(W)$.

Now, choose a class $e \in H^2(W;\mathbb{Z})$ with $e \otimes \mathbb{R} = -[\omega_k]$, and let $\pi\colon\ M \to W$ be the S^1-bundle over W of Euler class e. One then finds, for each $k \in \mathbb{N}_0$, a connection 1-form α_k on M with curvature form ω_k, that is, $\mathrm{d}\alpha_k = \pi^*\omega_k$, see [11, Section 7.2]. (This normalisation corresponds to thinking of S^1 as $\mathbb{R}/\mathbb{Z}$.) Hence, the α_k are contact forms with Reeb vector field given by the unit tangent vector field along the fibres.

The α_k, $k \in \mathbb{N}_0$, are pairwise nondiffeomorphic, because any diffeomorphism between α_k and α_ℓ would preserve the Reeb vector field and hence descend to a diffeomorphism between ω_k and ω_ℓ. □

Remark 10.2. (1) I do not know whether the contact structures $\ker \alpha_k$ are diffeomorphic. They all have the same underlying almost contact structure.

(2) I hedge my bets concerning dimensions 5 and 7.

(3) Contact forms with all Reeb orbits closed and of the same minimal period are also called *Zoll contact forms* [1,2].

11. Orbit equivalence

A slighty weaker question than the one asked by Viterbo is the following: are there examples of contact forms with the same Reeb vector field up to scaling by a function? Or, put differently, one asks for nondiffeomorphic contact forms whose Reeb flows are smoothly orbit equivalent. For the more general class of geodesible vector fields, this problem is best phrased as follows: on a manifold M, is there a geodesible vector field X and a function $f \in C^\infty(M, \mathbb{R}^+)$ such that fX is likewise geodesible? Of course, one should exclude the trivial case of f being constant, where one simply rescales the metric by the inverse constant.

This is related, but not equivalent to the question about nontrivially geodesically equivalent metrics, where two Riemannian metrics share the same geodesics up to reparametrisation (so the geodesic flows are orbit equivalent), but one metric is not a constant multiple of the other.

Matveev [19] has shown that among closed, connected 3-manifolds, examples of nontrivially geodesically equivalent metrics exist only on lens spaces and Seifert manifolds with Euler number zero. See also [20] for a discussion of this phenomenon in the context of general relativity.

Our question asks about the nontrivial equivalence of two foliations by geodesible vector fields. In some sense, this is a weaker question; on the other hand, a nontrivial equivalence between two Riemannian metrics may well become trivial when restricted to any geodesic foliation.

Example 11.1. On the 2-torus $T^2 = (\mathbb{R}/\mathbb{Z})^2$, we consider the standard flat metric $g_1 = \mathrm{d}x_1^2 + \mathrm{d}x_2^2$ and a second flat metric $g_2 = \mathrm{d}x_1^2 + a\,\mathrm{d}x_2^2$ with $a \in \mathbb{R}^+ \setminus \{1\}$. Then, g_2 is not a constant multiple of g_1, but the two metrics are geodesically equivalent: the geodesics in both cases are the images of straight lines in $\mathbb{R}^2$ under the projection to T^2. A geodesic foliation is given by the straight lines of some constant slope, and along those parallel lines the unit vector fields for the two metrics differ by a constant.

Using an idea going back to Beltrami and explained in [19], we can exhibit a simple example of geodesically equivalent metrics on S^3 that give rise to a geodesible vector field admitting nontrivial rescalings into likewise geodesible vector fields. Here, by construction, the vector fields are diffeomorphic. As I shall explain, rescalings of geodesible vector fields that define an S^1-fibration will always be diffeomorphic.

Example 11.2. For $a_1, a_2 \in \mathbb{R}^+$, consider the linear map $A = A_{a_1,a_2}\colon\ (z_1, z_2) \mapsto (a_1 z_1, a_2 z_2)$ on $\mathbb{C}^2 = \mathbb{R}^4$. Then, define $\phi = \phi_{a_1,a_2}\colon\ S^3 \to S^3$ by $\phi(p) = A(p)/|A(p)|$. Let $g_{a_1,a_2} = \phi^* g_0$ be the pull-back of the round metric g_0 on S^3. Since ϕ takes great circles to great circles, the metric $\phi^* g_0$ is geodesically equivalent to g_0, nontrivially so unless $a_1 = a_2$.

A straightforward computation yields the following expression for g_{a_1,a_2}:

$$\begin{aligned} g_{a_1,a_2} = {} & \frac{a_1^2}{\Delta}\left(\mathrm{d}x_1^2 + \mathrm{d}y_1^2\right) + \frac{a_2^2}{\Delta}\left(\mathrm{d}x_2^2 + \mathrm{d}y_2^2\right) \\ & - \frac{a_1^4}{\Delta^2}\left(x_1\,\mathrm{d}x_1 + y_1\,\mathrm{d}y_1\right)^2 - \frac{a_2^4}{\Delta^2}\left(x_2\,\mathrm{d}x_2 + y_2\,\mathrm{d}y_2\right)^2 \\ & - \frac{2a_1^2 a_2^2}{\Delta^2}\left(x_1\,\mathrm{d}x_1 + y_1\,\mathrm{d}y_1\right)\left(x_2\,\mathrm{d}x_2 + y_2\,\mathrm{d}y_2\right), \end{aligned}$$

where we write

$$\Delta = \Delta_{a_1,a_2}(r_1, r_2) = a_1^2 r_1^2 + a_2^2 r_2^2.$$

Recall that in terms of polar coordinates we have $\mathrm{d}x_i^2 + \mathrm{d}y_i^2 = \mathrm{d}r_i^2 + r_i^2\,\mathrm{d}\varphi_i^2$ and $x_i\,\mathrm{d}x_i + y_i\,\mathrm{d}y_i = r_i\,\mathrm{d}r_i$.

The positive Hopf fibration is generated by $X_0 = \partial_{\varphi_1} + \partial_{\varphi_2}$. This vector field has constant length 1 with respect to all the metrics g_{a_1,a_2}, so from the viewpoint of geodesic foliations this yields nothing new. Also, the corresponding contact forms

$$\alpha_{a_1,a_2} = g_{a_1,a_2}(X_0, \,.\,) = \frac{a_1^2 r_1^2\,\mathrm{d}\varphi_1 + a_2^2 r_2^2\,\mathrm{d}\varphi_2}{\Delta}$$

all have X_0 as Reeb vector field, and so they are just diffeomorphic deformations of the standard contact form $\alpha_{1,1}$ by Proposition 2.1.

A more interesting choice is to take the great circle foliation generated by

$$X_1 = x_1\partial_{x_2} - x_2\partial_{x_1} + y_1\partial_{y_2} - y_2\partial_{y_1}.$$

We write $L = L_{a_1,a_2} = \left(g_{a_1,a_2}(X_1, X_1)\right)^{1/2}$ for the length of X_1 with respect to g_{a_1,a_2}. One computes

$$L^2 = \frac{a_1^2 r_2^2 + a_2^2 r_1^2}{\Delta} - \frac{(a_1^2 - a_2^2)^2}{\Delta^2}\,(x_1 x_2 + y_1 y_2)^2.$$

Thus, we have found the nontrivial family of geodesible vector fields $X_1/L_{a_1,a_2}$, with corresponding metric g_{a_1,a_2}, all generating the same foliation of S^3 by great circles. The corresponding 1-form

$$\alpha = \alpha_{a_1,a_2} = g_{a_1,a_2}(X_1/L_{a_1,a_2}, \,.\,)$$

can be computed explicitly as

$$\begin{aligned}L\alpha = &-\frac{a_1^2}{\Delta}\,(x_2\,\mathrm{d}x_1 + y_2\,\mathrm{d}y_1) + \frac{a_2^2}{\Delta}\,(x_1\,\mathrm{d}x_2 + y_1\,\mathrm{d}y_2)\\ &+\frac{a_1^4 - a_1^2a_2^2}{\Delta^2}\,(x_1x_2 + y_1y_2)\,(x_1\,\mathrm{d}x_1 + y_1\,\mathrm{d}y_1)\\ &-\frac{a_2^4 - a_1^2a_2^2}{\Delta^2}\,(x_1x_2 + y_1y_2)\,(x_2\,\mathrm{d}x_2 + y_2\,\mathrm{d}y_2).\end{aligned}$$

I did not check whether these are contact forms for all $a_1, a_2 \in \mathbb{R}^+$, but by the openness of the contact condition they certainly are for a_1, a_2 close to 1. Then, $X_1/L_{a_1,a_2}$ will be the Reeb vector field of α_{a_1,a_2}.

The following proposition gives a more systematic statement about rescalings of geodesible vector fields that define an S^1-fibration. This is essentially due to Wadsley [31] (in greater generality); for the case at hand it can be retraced to the work of Boothby and Wang [6].

Proposition 11.3. *Let X be a geodesible vector field on a closed manifold M such that the flow lines of X are the fibres of a principal S^1-bundle $M \to M/S^1$. Then, after a constant rescaling of X all orbits have (minimal) period 1, so that the flow of X defines the S^1-action. A rescaling fX of X is likewise geodesible if and only if all orbits have the same period. When this period is 1, the vector fields X and fX are diffeomorphic by a diffeomorphism that sends each fibre to itself and is isotopic to the identity via such diffeomorphisms.*

Proof. If X is geodesible, we find a 1-form α with $\alpha(X) = 1$ and $i_X\mathrm{d}\alpha = 0$ by Proposition 3.3. Then, [11, Lemmas 7.2.6 and 7.2.7], which fill a gap in [6], show that the orbits of X all have the same period. Notice that Lemma 7.2.7 in [11] is formulated for Reeb vector fields, but the proof only uses the property $i_X\mathrm{d}\alpha = 0$, not the nondegeneracy of $\mathrm{d}\alpha|_{\ker\alpha}$.

If fX is geodesible, the same argument applies. Conversely, if the orbits of fX all have the same period, then the flow of X defines an S^1-bundle structure, and any connection 1-form for this bundle is a characteristic 1-form for fX, which makes fX geodesible.

Now, suppose the period of fX equals 1. Given a local section $U \cong D^2$ of X, the flow of X defines a trivialisation $U \times S^1$ of the bundle, and f gives rise to a family of 1-periodic velocity functions $v_u\colon [0,1] \to \mathbb{R}^+$ with $\int_0^1 v_u(t)\,\mathrm{d}t = 1$ for every $u \in U$. (I refrain from writing v_u as a function on S^1, since the time parameter t should not be confused with the fibre parameter defined by the flow of X.) Let $\psi\colon D^2 \to [0,1]$ be a bump function equal to 1 on a disc of radius 1/2, say, and supported in the interior of D^2. Then,

$$\mu\bigl(\psi(u)v_u + 1 - \psi(u)\bigr) + 1 - \mu$$

defines for each $u \in U$ and $\mu \in [0,1]$ a 1-periodic velocity function $[0,1] \to \mathbb{R}^+$ of integral 1. This gives rise to an isotopy along fibres whose time-1 map sends X to fX on fibres where $\psi(u) = 1$, and which is stationary on fibres along which $f = 1$. This allows us to patch together such local isotopies to obtain the desired result. $\square$

The next corollary also applies to Example 11.2.

Corollary 11.4. *If R and fR are Reeb vector fields on a closed 3-manifold with all orbits periodic of the same period 1, then any corresponding contact forms are related via a fibre-preserving isotopy.*

Proof. This follows immediately by combining the last statement of Proposition 11.3 with Proposition 2.1.

Alternatively, one can give a direct proof, using a refinement of the proof of Proposition 2.1. Let α_0 be a contact form with Reeb vector field R, and α_1 a contact form for fR. Set $\alpha_t := (1-t)\alpha_0 + t\alpha_1$. We would like to find an isotopy $(\psi_t)_{t\in[0,1]}$ satisfying (5) as in the proof of Proposition 2.1.

The α_t are contact forms with Reeb vector field R_t proportional to R. We try to find an isotopy (ψ_t) generated by a vector field X_t of the form

$$X_t = h_t R_t + Y_t$$

with $Y_t \in \ker \alpha_t$. Differentiating (5), we find

$$\alpha_1 - \alpha_0 + \mathrm{d}h_t + i_{Y_t}\mathrm{d}\alpha_t = 0. \tag{10}$$

When we plug R into this equation, we find

$$f^{-1} - 1 + \mathrm{d}h_t(R) = 0. \tag{11}$$

The condition that the period of fR be 1 translates into f^{-1} integrating to 1 along any fibre of the S^1-bundle. This allows us to define a family of functions h_t satisfying (11), and then there is a unique vector field $Y_t \in \ker \alpha_t$ satisfying (10). Both $\alpha_1 - \alpha_0 + \mathrm{d}h_t$ and $\mathrm{d}\alpha_t$ are lifts of differential forms on the quotient surface M/S^1, hence the flow of Y_t preserves fibres. □

Remark 11.5. If R is the Reeb vector field of a contact form α (on a connected manifold M), then the rescaled vector field $f^{-1}R$ is never the Reeb vector field of $f\alpha$, unless the function f is constant, for the identity

$$0 = i_R\mathrm{d}(f\alpha) = i_R(\mathrm{d}f \wedge \alpha) = \mathrm{d}f(R)\alpha - \mathrm{d}f$$

implies that $\mathrm{d}f$ vanishes on all vectors tangent to the contact structure $\ker \alpha$, and by [11, Theorem 3.3.1] any two points in M can be joined by a curve tangent to the contact structure.

Acknowledgements

Special thanks go to Claude Viterbo for putting this question to me, and to Alberto Abbondandolo for showing me identity (1). I first learned about basic cohomology from a talk by Robert Wolak. I thank Peter Albers, Jesús Gonzalo and Kai Zehmisch for their collaboration on various projects that have contributed in many ways to the conception of this note, and Christian Lange for useful conversations on S^1-orbibundles. I also thank the referee for perceptive and detailed comments on the original manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

[1] Abbondandolo, A., Benedetti, G.: On the local systolic optimality of Zoll contact forms. arXiv:1912.04187

[2] Abbondandolo, A., Bramham, B., Hryniewicz, U. L., Salomão, P. A. S.: Sharp systolic inequalities for Reeb flows on the three-sphere. Invent. Math. **211**, 687–778

[3] Albers, P., Geiges, H., Zehmisch, K.: Reeb dynamics inspired by Katok's example in Finsler geometry. Math. Ann. **370**, 1883–1907 (2018)

[4] Albers, P., Geiges, H., Zehmisch, K.: Pseudorotations of the 2-disc and Reeb flows on the 3-sphere. Ergodic Theory Dynam. Systems **42**, 402–436 (2022)

[5] Albers, P., Geiges, H., Zehmisch, K.: A symplectic dynamics proof of the degree-genus formula. Arnold Math. J., to appear

[6] Boothby, W.M., Wang, H.C.: On contact manifolds. Ann. Math. **68**(2), 721–734 (1958)

[7] Cieliebak, K., Volkov, E.: First steps in stable Hamiltonian topology. J. Eur. Math. Soc. (JEMS) **17**, 321–404 (2015)

[8] Conlon, L.: Differentiable Manifolds, 2nd edn. Birkhäuser Adv. Texts Basler Lehrbücher, Birkhäuser, Basel (2001)

[9] Cristofaro-Gardiner, D., Hutchings, M., Ramos, V.: The asymptotics of ECH capacities. Invent. Math. **199**, 187–214 (2015)

[10] Croke, C.B., Kleiner, B.: Conjugacy and rigidity for manifolds with a parallel vector field. J. Differ. Geom. **39**, 659–680 (1994)

[11] Geiges, H.: An Introduction to Contact Topology, Cambridge Stud. Adv. Math. **109**, Cambridge University Press, Cambridge (2008)

[12] Geiges, H., Gonzalo Pérez, J.: Contact geometry and complex surfaces. Invent. Math. **121**, 147–209 (1995)

[13] Geiges, H., Gonzalo Pérez, J.: Generalised spin structures on 2-dimensional orbifolds. Osaka J. Math. **49**, 449–470 (2012)

[14] Geiges, H., Gonzalo Pérez, J.: Transversely holomorphic flows and contact circles on spherical 3-manifolds. Enseign. Math. **62**, 527–567 (2016)

[15] Geiges, H., Lange, C.: Seifert fibrations of lens spaces. Abh. Math. Semin. Univ. Hambg. **88**, 1–22 (2018)

[16] Gluck, H.: Dynamical Behavior of Geodesic Fields, in Global Theory of Dynamical Systems. Lecture Notes in Mathematics, vol. 819. Springer, Berlin (1980)

[17] Jankins, M., Neumann, W. D.: Lectures on Seifert Manifolds, Brandeis Lecture Notes **2**, Brandeis University, Waltham, MA (1983); available at http://www.math.columbia.edu/~neumann/preprints/

[18] Kegel, M., Lange, C.: A Boothby–Wang theorem for Besse contact manifolds. Arnold Math. J. **7**, 225–241 (2021)

[19] Matveev, V.S.: Three-dimensional manifolds having metrics with the same geodesics. Topology **42**, 1371–1395 (2003)

[20] Matveev, V.S.: Geodesically equivalent metrics in general relativity. J. Geom. Phys. **62**, 675–691 (2012)

[21] McDuff, D.: Examples of simply-connected symplectic non-Kählerian manifolds. J. Differ. Geom. **20**, 267–277 (1984)

[22] McDuff, D.: Examples of symplectic structures. Invent. Math. **89**, 13–36 (1987)

[23] McDuff, D., Salamon, D.: Introduction to Symplectic Topology. Oxf. Grad. Texts Math., 3rd edn. Oxford University Press, Oxford (2017)

[24] Morita, S.: Geometry of Differential Forms. Transl. Math. Monogr. **201**, American Mathematical Society, Providence, RI (2001)

[25] Satake, I.: The Gauss–Bonnet theorem for V-manifolds. J. Math. Soc. Jpn. **9**, 464–492 (1957)

[26] Scott, P.: The geometries of 3-manifolds. Bull. Lond. Math. Soc. **15**, 401–487 (1983)

[27] Seaton, C.: Two Gauss–Bonnet and Poincaré–Hopf theorems for orbifolds with boundary. Differ. Geom. Appl. **26**, 42–51 (2008)

[28] Sullivan, D.: A foliation of geodesics is characterized by having no "tangent homologies". J. Pure Appl. Algebra **13**, 101–104 (1978)

[29] Tischler, D.: On fibering certain foliated manifolds over S^1. Topology **9**, 153–154 (1970)

[30] Tondeur, P.: Geometry of Foliations. Monogr. Math. **90**, Birkhäuser, Basel (1997)

[31] Wadsley, A.W.: Geodesic foliations by circles. J. Differ. Geom. **10**, 541–549 (1975)

[32] Weinstein, A.: Symplectic V-manifolds, periodic orbits of Hamiltonian systems, and the volume of certain Riemannian manifolds. Commun. Pure Appl. Math. **30**, 265–271 (1977)

Hansjörg Geiges
Mathematisches Institut
Universität zu Köln
Weyertal 86–90
Köln 50931
Germany
e-mail: `geiges@math.uni-koeln.de`

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:37
https://doi.org/10.1007/s11784-022-00943-y
Published online April 12, 2022

Journal of Fixed Point Theory and Applications

On the symplectic fillings of standard real projective spaces

Paolo Ghiggini and Klaus Niederkrüger-Eid

Abstract. We prove, in a geometric way, that the standard contact structure on $\mathbb{RP}^{2n-1}$ is not Liouville fillable for $n \geq 3$ and odd. We also prove for all n that semipositive fillings of such contact structures are always simply connected. Finally, we give yet another proof of the Eliashberg–Floer–McDuff theorem on the diffeomorphism type of the symplectically aspherical fillings of the standard contact structure on S^{2n-1}.

Mathematics Subject Classification. 57R17.

Keywords. Real projective spaces, contact structures, Liouville fillings, J-holomorphic spheres.

1. Introduction

The standard contact structure ξ on S^{2n-1} is described in coordinates by the equation

$$\xi = \ker \sum_{j=1}^{n} (x_j\, dy_j - y_j\, dx_j).$$

Geometrically, ξ_p is the unique complex hyperplane in T_pS^{2n-1} for every $p \in S^{2n-1}$. The antipodal involution of S^{2n-1} preserves ξ, and therefore induces a contact structure on $\mathbb{RP}^{2n-1}$ which we still denote by ξ. The disc bundle of the line bundle $\mathcal{O}_{\mathbb{P}^{n-1}}(-2)$ on $\mathbb{CP}^{n-1}$ is a strong symplectic filling of $(\mathbb{RP}^{2n-1}, \xi)$. On the other hand, $\mathbb{RP}^{2n-1}$ cannot be the boundary of a $2n$-dimensional manifold with the homotopy type of an n-dimensional CW complex if $2n-1 \geq 5$; see [3, Section 6.2]. This implies that a real projective space of dimension at least 5 does not admit any Weinstein fillable contact structure. Our main result is the following.

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

Theorem 1.1. *The standard contact structure on $\mathbb{RP}^{2n-1}$ admits no symplectically aspherical fillings for $n > 1$ and odd. In particular, it is not Liouville fillable.*

These are the first examples of strongly but not Liouville fillable contact structures in high dimension. Examples in dimension three were given by the first author in [4] using Heegaard Floer homology. In contrast with the high dimensional situation, the standard contact structure on $\mathbb{RP}^3$ is the canonical contact structure on the unit cotangent bundle of S^2 and, therefore, is Weinstein fillable.

After a preliminary version of our result (originally for $\mathbb{RP}^5$ only) was announced, Zhou proved in [11] that $(\mathbb{RP}^{2n-1}, \xi)$ is not Liouville fillable if $n \neq 2^k$. He also proves similar nonfillability results for some other links of cyclic quotient singularities. Zhou's proof uses advanced properties of symplectic cohomology; in contrast our proof is more direct, as it relies on the analysis of how a certain moduli space of holomorphic spheres can break, in the spirit of McDuff's classification of symplectic fillings of $\mathbb{RP}^3$ in [7].

The strategy is the following. The standard contact structure ξ on $\mathbb{RP}^{2n-1}$ admits a contact form whose Reeb orbits are the fibres of the Hopf fibration $\mathbb{RP}^{2n-1} \to \mathbb{CP}^{n-1}$. If (W, ω) is a strong symplectic filling of $(\mathbb{RP}^{2n-1}, \xi)$, by a symplectic reduction of ∂W (informally speaking, by replacing ∂W with its quotient by the Reeb flow) we obtain a closed symplectic manifold $(\overline{W}, \overline{\omega})$ with a codimension two symplectic submanifold $W_\infty \cong \mathbb{CP}^{n-1}$ (corresponding to the quotient of ∂W) such that $\overline{W} \setminus W_\infty$ is symplectomorphic to $\mathrm{int}(W)$; that is, $\overline{\omega}|_{\overline{W} \setminus W_\infty} = \omega|_{\mathrm{int}(W)}$. The normal bundle of W_∞ is isomorphic to $\mathcal{O}_{\mathbb{P}^{n-1}}(2)$.

We fix a point and a hyperplane in W_∞, and we consider the moduli space of holomorphic spheres in $\overline{W}$ which are homotopic to a projective line and pass both through the point and the hyperplane. We prove by topological considerations that if the compactification of that moduli space contains only nodal curves with at most two irreducible components each of which intersect W_∞ nontrivially, then some of these nodal curves will be composed of two spheres that represent identical homology classes up to torsion. This implies in particular that the homology class of a projective line in W_∞ is the double of some homology classes in $\overline{W}$ up to torsion.

If n is odd this is a contradiction because the first Chern class of a line is $n + 2$; only at this step we use the hypothesis on the parity of n. This implies that there is either a nodal holomorphic sphere in $\overline{W}$ in the homology class of a line of W_∞ with at least three irreducible components or a nodal holomorphic sphere with an irreducible component which is disjoint from W_∞. Since a nodal sphere intersects W_∞ in exactly two points, in either case at least one irreducible component must lie entirely in $\mathrm{int}(W)$, which, therefore, is not symplectically aspherical.

If (W, ω) is not symplectically aspherical we lose control on the compactification of the moduli space, which is not surprising, given that $(\mathbb{RP}^{2n-1}, \xi)$ does admit spherical fillings. However, if W is semipositive (and maybe even more generally, using some abstract perturbation scheme), we still have enough control to be able to draw conclusions about the fundamental group of W.

Theorem 1.2. *If (W, ω) is a semipositive symplectic filling of $(\mathbb{RP}^{2n-1}, \xi)$, then W is simply connected.*

If we apply the same techniques to a symplectically aspherical filling of the standard contact structure on S^{2n-1}, we obtain that the filling must be diffeomorphic to the ball, a result originally due to Eliashberg, Floer and McDuff. This is, at least, the fifth proof, after the original one in [8], a very similar one in [10], the one in [5] using moduli spaces of holomorphic discs with boundary on a family of LOB's, and the one in [1]. The proof given here is close to the original one, but uses a different compactification of the filling and is slightly simpler.

2. The moduli space of lines

2.1. The smooth stratum

By the Weinstein neighbourhood theorem, W_∞ has a tubular neighbourhood that is symplectomorphic to a neighbourhood of the zero section in the total space of $\mathcal{O}_{\mathbb{P}^{n-1}}(2)$. Let $\mathcal{J}$ be the space of almost complex structures on $\overline{W}$ which are compatible with $\overline{\omega}$ and coincide with the natural (integrable) complex structure on $\mathcal{O}_{\mathbb{P}^{n-1}}(2)$ on a fixed neighbourhood of W_∞.

For any almost complex structure $J \in \mathcal{J}$, any line $\ell \subset \mathbb{CP}^{n-1} \cong W_\infty$ is a J-holomorphic sphere. Moreover,

$$T\overline{W}|_\ell \cong \mathcal{O}_{\mathbb{P}^1}(2) \oplus \underbrace{\mathcal{O}_{\mathbb{P}^1}(1) \oplus \cdots \oplus \mathcal{O}_{\mathbb{P}^1}(1)}_{n-2} \oplus \mathcal{O}_{\mathbb{P}^1}(2) \tag{1}$$

as *holomorphic* vector bundle, where the first $\mathcal{O}_{\mathbb{P}^1}(2)$ summand is the tangent bundle of ℓ, the $(n-2)$-many $\mathcal{O}_{\mathbb{P}^1}(1)$-summands correspond to the normal bundle of ℓ in $\mathbb{CP}^{n-1} \cong W_\infty$ and the last $\mathcal{O}_{\mathbb{P}^1}(2)$-summand is the normal bundle of W_∞ in $\overline{W}$ restricted to ℓ.

We fix a point $p_0 \in W_\infty$ and a hyperplane $H_\infty \cong \mathbb{CP}^{n-2} \subset W_\infty$ such that $p_0 \notin H_\infty$. We denote the moduli space of unparametrised J-holomorphic spheres in $\overline{W}$ that are homotopic to the lines in $W_\infty \cong \mathbb{CP}^{n-1}$ with pointwise constraints at p_0 and H_∞ by $\mathcal{M}(p_0, H_\infty)$. We also consider the moduli space $\mathcal{M}_z(p_0, H_\infty)$ of unparametrised J-holomorphic spheres as above with an extra free marked point z. There is a projection

$$\mathfrak{f} \colon \mathcal{M}_z(p_0, H_\infty) \to \mathcal{M}(p_0, H_\infty).$$

that forgets the marked point.

Lemma 2.1. *$\mathcal{M}(p_0, H_\infty)$ has expected dimension $2n-2$ and $\mathcal{M}_z(p_0, H_\infty)$ has expected dimension $2n$, where $\dim \overline{W} = 2n$.*

Proof. The decomposition (1) gives $\langle c_1(T\overline{W}), [\ell]\rangle = n + 2$. The expected dimension of $\mathcal{M}(p_0, H_\infty)$ is

$$\text{vir-dim}\, \mathcal{M}(p_0, H_\infty) = 2\langle c_1(T\overline{W}), [\ell]\rangle + 2n + 4 - 2n - 4 - 6 = 2n - 2.$$

The first two terms compute the index of the linearised Cauchy-Riemann operator, the third is the contribution of two extra marked points, the fourth and the fifth come from the condition that the marked points be mapped to p_0 and H_∞, and the last is the dimension of the biholomorphism group of the sphere. $\square$

The main reason for keeping the almost complex structure integrable near W_∞ is to have positivity of intersection between W_∞ and J-holomorphic spheres. This fact makes our moduli space particularly well behaved, as the following lemma shows.

Lemma 2.2. *All J-holomorphic spheres of $\mathcal{M}(p_0, H_\infty)$ are simply covered and are either lines in W_∞ or intersect W_∞ transversely in exactly two points.*

Proof. Since the algebraic intersection between W_∞ and ℓ is 2, positivity of intersection implies that a sphere of $\mathcal{M}(p_0, H_\infty)$ is either contained in W_∞, in which case it is a line and, therefore, simply covered, or it intersect W_∞ with total multiplicity two. Since the constraints force two distinct intersection points, positivity of intersection implies that they are the only ones and that they each have multiplicity one. $\square$

Proposition 2.3. *For a generic almost complex structure $J \in \mathcal{J}$, the moduli spaces $\mathcal{M}(p_0, H_\infty)$ and $\mathcal{M}_z(p_0, H_\infty)$ are smooth manifolds of dimension $2n-2$ and $2n$, respectively.*

Proof. The J-holomorphic spheres of $\mathcal{M}(p_0, H_\infty)$ which are contained in the neighbourhood of W_∞ where J is integrable correspond to holomorphic sections of $\mathcal{O}_{\mathbb{P}^{n-1}}$ and therefore admit a decomposition of the restriction of $T\overline{W}$ as in Eq. (1). Since the decomposition is into positive holomorphic line bundles, those spheres are Fredholm regular for every almost complex structures $J \in \mathcal{J}$, because the Cauchy-Riemann operator on a positive line bundle over $\mathbb{CP}^1$ is surjective by Serre duality; see [9, Lemma 3.3.1]

All other J-holomorphic spheres of $\mathcal{M}(p_0, H_\infty)$ are Fredholm regular for a generic $J \in \mathcal{J}$, because they are simply covered and intersect the region where J is generic. Moreover, the pointwise constraints cut out $\mathcal{M}(p_0, H_\infty)$ transversely for a generic J: for spheres near W_∞ this is an explicit computation, and for all other spheres of $\mathcal{M}(p_0, H_\infty)$ it follows from [9, Theorem 3.4.1] and [9, Remark 3.4.8]. Therefore $\mathcal{M}(p_0, H_\infty)$ is a smooth manifold of the dimension predicted by Lemma 2.1. The corresponding statements for $\mathcal{M}_z(p_0, H_\infty)$ follow from those for $\mathcal{M}(p_0, H_\infty)$. $\square$

2.2. The compactified moduli space

Let $\overline{\mathcal{M}}(p_0, H_\infty)$ and $\overline{\mathcal{M}}_z(p_0, H_\infty)$ be the Gromov compactifications of $\mathcal{M}(p_0, H_\infty)$ and $\mathcal{M}_z(p_0, H_\infty)$, respectively, and let

$$\bar{\mathfrak{f}} \colon \overline{\mathcal{M}}_z(p_0, H_\infty) \to \overline{\mathcal{M}}(p_0, H_\infty),$$

be the forgetful map. We denote $\mathcal{M}^{\mathrm{red}}(p_0, H_\infty) = \overline{\mathcal{M}}(p_0, H_\infty) \backslash \mathcal{M}(p_0, H_\infty)$ and $\mathcal{M}_z^{\mathrm{red}}(p_0, H_\infty) = \overline{\mathcal{M}}_z(p_0, H_\infty) \backslash \mathcal{M}_z(p_0, H_\infty)$.

Lemma 2.4. *If $\overline{W} \backslash W_\infty$ is symplectically aspherical, then every nodal sphere of $\mathcal{M}^{\mathrm{red}}(p_0, H_\infty)$ has exactly two irreducible components, one of which intersects*

W_∞ only at p_0 and the other one which intersects W_∞ only at a point of H_∞. Both components are simply covered and their intersection with W_∞ is transverse.

Proof. None of the irreducible components of nodal spheres in $\overline{\mathcal{M}}(p_0, H_\infty)$ is contained in W_∞ because any bubble component needs to have positive symplectic area strictly smaller than the symplectic area of ℓ, but the homology class of ℓ has the smallest positive symplectic area in $\mathbb{CP}^{n-1}$. Then, by positivity of intersection with W_∞ a nodal sphere must intersect W_∞ in at most two points. Moreover, if $\overline{W}\setminus W_\infty$ is symplectically aspherical, every irreducible component must intersect W_∞. This implies that there are exactly two irreducible components and the intersection of each with W_∞ has multiplicity one. Therefore, both components are simply covered. □

This lemma implies that we have enough topological control on the nodal curves to show that they have smooth moduli spaces.

Lemma 2.5. *The moduli space $\mathcal{M}^{\mathrm{red}}(p_0, H_\infty)$ is a smooth manifold of dimension $2n-4$. The forgetful map*

$$\mathfrak{f}^{\mathrm{red}}\colon \mathcal{M}_z^{\mathrm{red}}(p_0, H_\infty) \to \mathcal{M}^{\mathrm{red}}(p_0, H_\infty),$$

is a locally trivial fibration with fibre $S^2 \vee S^2$.

Proof. By Lemma 2.4, the irreducible components of the nodal spheres of the moduli space $\mathcal{M}^{\mathrm{red}}(p_0, H_\infty)$ are simply covered and intersect the region where the almost complex structure can be made generic. Then, the statement follows from [9, Theorem 6.2.6]. □

What we have shown so far about $\overline{\mathcal{M}}_z(p_0, H_\infty)$ is enough to show that the image of the evaluation is a pseudocycle in $\overline{W}$ (see [9, Section 6.5]). While pseudocycles are good enough for certain degree arguments, as for example in Sect. 4, in the proof of our main theorem, we will need a differentiable structure on the compactified moduli spaces. The reason is that our proof is based on studying the properties of the homology class $[\mathrm{ev}^{-1}(\ell)]$ in $\overline{\mathcal{M}}_z(p_0, H_\infty)$, and we have found no better way to show that $\mathrm{ev}^{-1}(\ell)$ is well-behaved than by the implicit function theorem. Since there is no reason to expect that ℓ can be made disjoint from the image of $\mathcal{M}_z^{\mathrm{red}}(p_0, H_\infty)$ by the evaluation map, we need a smooth structure on the whole compactified moduli space, and not only in its irreducible part. Luckily our moduli space is simple enough that standard gluing theory (as explained, for example, in [9]) already produces a smooth structure.

In the rest of the section, we will sketch the construction of a C^1-structure, which in turn can be promoted to a smooth structure by a classical result in differential topology; see for example [6, Section 2.2, Theorem 2.10].

To endow $\overline{\mathcal{M}}(p_0, H_\infty)$ and $\overline{\mathcal{M}}_z(p_0, H_\infty)$ with the structure of a C^1-manifold, we exhibit them as union of two open manifolds which are patched together by a diffeomorphism of class C^1. The two patches are the irreducible strata $\mathcal{M}(p_0, H_\infty)$ and $\mathcal{M}_z(p_0, H_\infty)$ on one side, and suitable fibrations over the reducible strata $\mathcal{M}^{\mathrm{red}}(p_0, H_\infty)$ and $\mathcal{M}_z^{\mathrm{red}}(p_0, H_\infty)$ on the other hand. The fibres of those fibrations are, roughly speaking, spaces of gluing parameters. We are going to define them momentarily.

The first step in our construction is to introduce local gauge fixing conditions to simplify the presentation of the moduli spaces, so that they almost become spaces of parametrised J-holomorphic spheres. Identify the neighbourhood of W_∞ in $\overline{W}$ with a neighbourhood of the 0-section of $\mathcal{O}_{\mathbb{P}^{n-1}}(2)$ as already discussed above. The hyperplane H_∞ is the 0-set of a holomorphic section σ in $\mathcal{O}_{\mathbb{P}^{n-1}}(1)$, and it follows that σ^2 is a section of $\mathcal{O}_{\mathbb{P}^{n-1}}(2)$ that has a zero of order two along H_∞. Multiplying σ^2 with a small constant, we can assume that its image lies in an arbitrarily small neighbourhood of the 0-section. Its graph is a J-holomorphic hypersurface in $\overline{W}$ that we will call $\widetilde{W}_\infty$. In particular $\widetilde{W}_\infty \cap W_\infty = H_\infty$ and $T\widetilde{W}_\infty|_{H_\infty} = TW_\infty|_{H_\infty}$. Then, every sphere of $\mathcal{M}(p_0, H_\infty)$ which is not contained in W_∞ intersects $\widetilde{W}_\infty$ in two points: one in H_∞ and one in $\widetilde{W}_\infty \setminus H_\infty$.

Let $\widetilde{\mathcal{M}}(p_0, H_\infty)$ be the open subset of $\mathcal{M}(p_0, H_\infty)$ consisting of those spheres which are not contained in W_∞ and $\widetilde{\mathcal{M}}_z(p_0, H_\infty)$ the corresponding open subset of $\mathcal{M}_z(p_0, H_\infty)$. By the discussion in the previous paragraph, we can fix a parametrisation for every element in $\widetilde{\mathcal{M}}(p_0, H_\infty)$ identifying this moduli space with the set of J-holomorphic maps $u\colon S^2 \to \overline{W}$ whose image is homotopic to ℓ but not contained in W_∞, and such that $u(0) = p_0$, $u(1) \in \widetilde{W}_\infty \setminus H_\infty$ and $u(\infty) \in H_\infty$.

We also denote by $\widetilde{\mathcal{M}}^{\mathrm{red}}(p_0, H_\infty)$ the set of pairs of J-holomorphic maps (u^0, u^∞), with $u^0, u^\infty\colon S^2 \to \overline{W}$, such that

$$\begin{aligned} &u^0(0) = p_0\,, && u^0(\infty) = u^\infty(0)\,, && u^\infty(\infty) \in H_\infty\,, \\ &u^0(1) \in \widetilde{W}_\infty \setminus H_\infty\,, && |d_\infty u^\infty| = 1 && \end{aligned}$$

and the image of the "connected sum map" $u^0 \# u^\infty\colon S^2 \# S^2 \cong S^2 \to \overline{W}$ is homotopic to ℓ. Here $|d_\infty u^\infty|$ denotes the norm of the differential of u^∞ at $\infty \in S^2$ computed with respect to the round metric on S^2 and the metric induced by J and $\overline{\omega}$ on $\overline{W}$. Only the component u^0 meets $\widetilde{W}_\infty \setminus H_\infty$, because u^1 already intersects $\widetilde{W}_\infty$ in H_∞.

The group of complex numbers of modulus 1 acts on $\widetilde{\mathcal{M}}^{\mathrm{red}}(p_0, H_\infty)$ by

$$\theta \cdot (u^0, u^\infty) = \bigl(u^0, u^\infty(\theta^{-1} \cdot)\bigr),$$

and the quotient by this action is $\mathcal{M}^{\mathrm{red}}(p_0, H_\infty)$. This implies that the projection $\widetilde{\mathcal{M}}^{\mathrm{red}}(p_0, H_\infty) \to \mathcal{M}^{\mathrm{red}}(p_0, H_\infty)$ is a principal S^1-bundle which need not be trivial.

The second step is to define the fibrations over $\mathcal{M}^{\mathrm{red}}(p_0, H_\infty)$ and $\mathcal{M}_z^{\mathrm{red}}(p_0, H_\infty)$ which give one of the two patches. Let $\pi\colon S^2 \times S^2 \dashrightarrow S^2$ be the rational map $\pi(x, y) = y/x$, which is not defined at the points $(0, 0)$ and (∞, ∞). If we make S^1 act on $S^2 \times S^2$ by $\theta \cdot (x, y) = (x, \theta y)$ and on S^2 by $\theta \cdot w = \theta w$, then π is S^1-equivariant. Let $\mathfrak{X}$ be the smooth variety obtained by blowing up $S^2 \times S^2$ at $(0, 0)$ and (∞, ∞). The action of S^1 on $S^2 \times S^2$ induces an action on $\mathfrak{X}$, and π extends to a smooth S^1-equivariant map $\pi\colon \mathfrak{X} \to S^2$.

We denote by D_ϵ the disc with centre in 0 and radius ϵ in $\mathbb{C} \subset S^2$ and $\mathfrak{X}_\epsilon = \pi^{-1}(D_\epsilon)$. We define $E_\epsilon = \widetilde{\mathcal{M}}^{\mathrm{red}}(p_0, H_\infty) \times_{S^1} D_\epsilon$ and $X_\epsilon = \widetilde{\mathcal{M}}^{\mathrm{red}}(p_0, H_\infty)$ $\times_{S^1} \mathfrak{X}_\epsilon$. Both E_ϵ and X_ϵ are fibre bundles over $\mathcal{M}^{\mathrm{red}}(p_0, H_\infty)$ and there is a bundle map

$$\begin{tikzcd} X_\epsilon \arrow[rr, "\varpi"] \arrow[dr] & & E_\epsilon \arrow[dl] \\ & \mathcal{M}^{\mathrm{red}}(p_0, H_\infty). & \end{tikzcd}$$

The zero section E_0 of E is, of course, diffeomorphic to $\mathcal{M}^{\mathrm{red}}(p_0, H_\infty)$. Let us denote $\mathfrak{X}_0 = \pi^{-1}(0)$ and $X_0 = \varpi^{-1}(E_0)$. We observe that X_0 is diffeomorphic to $\mathcal{M}_z^{\mathrm{red}}(p_0, H_\infty)$. In fact $X_0 = \widetilde{\mathcal{M}}^{\mathrm{red}}(p_0, H_\infty) \times_{S^1} \mathfrak{X}_0$ and $\mathfrak{X}_0 = \{y = 0\} \cup \{x = \infty\}$, so we can identify the sphere $\{y = 0\}$ with the domain of u^0 and the sphere $\{x = \infty\}$ with the domain of u^∞.

The third, and last, step is the definition of the gluing maps between the two patches. Let $\dot{E}_\epsilon$ denote E_ϵ with the zero section removed, and $\dot{X}_\epsilon$ the preimage of $\dot{E}_\epsilon$. We can identify $\dot{E}_\epsilon \cong [\epsilon^{-1}, +\infty) \times \widetilde{\mathcal{M}}^{\mathrm{red}}(p_0, H_\infty)$, and therefore, standard gluing theory (see for example [9, Chapter 10]) yields a C^1-embeddings $\mathfrak{g} \colon \dot{E}_\epsilon \to \widetilde{\mathcal{M}}(p_0, H_\infty)$: if $(r, (u^0, u^\infty)) \in [\epsilon^{-1}, +\infty) \times \widetilde{\mathcal{M}}^{\mathrm{red}}$ $(p_0, H_\infty) \cong \dot{E}_\epsilon$, then $\mathfrak{g}(r, (u^0, u^\infty)) \in \widetilde{\mathcal{M}}(p_0, H_\infty)$ is the J-holomorphic sphere obtained by gluing u^0 and u^∞ with gluing parameter $R = \sqrt{r}$; see [9, Section 10.1].

We can also define a C^1-embedding $\mathfrak{G} \colon \dot{X}_\epsilon \to \widetilde{\mathcal{M}}_z(p_0, H_\infty)$ such that the diagram

$$\begin{tikzcd} \dot{X}_\epsilon \arrow[r, "\varpi"] \arrow[d, "\mathfrak{G}"'] & \dot{E}_\epsilon \arrow[d, "\mathfrak{g}"] \\ \widetilde{\mathcal{M}}_z(p_0, H_\infty) \arrow[r, "\mathfrak{f}"] & \widetilde{\mathcal{M}}(p_0, H_\infty) \end{tikzcd},$$

commutes. To define $\mathfrak{G}$ it is enough to identify $\varpi^{-1}(e)$ with the domain of $\mathfrak{g}(e)$ for all $e \in \dot{E}_\epsilon$. We recall that $\mathfrak{g}(e)$ is obtained by deforming a preglued map $\mathfrak{p}(e)$ with the same domain, so it is enough to identify (in a smooth way) $\varpi^{-1}(e)$ with the domain of the preglued map $\mathfrak{p}(e)$, whose construction we sketch now. Denote $e = [((u^0, u^\infty), t)]$ with $t \in D_\epsilon$ and $(u^0, u^\infty) \in \widetilde{\mathcal{M}}^{\mathrm{red}}(p_0, H_\infty)$, and $S_t = \pi^{-1}(t)$. We define $\widetilde{\mathfrak{p}}(u^0, u^\infty, t) \colon S_t \to \overline{W}$ by

$$\widetilde{\mathfrak{p}}(u^0, u^\infty, t)(x, y) = \begin{cases} u^0(x) & \text{if } |x| < \frac{1}{2\sqrt{|t|}}, \\ u^\infty(y) & \text{if } |x| > \frac{2}{\sqrt{|t|}}, \end{cases}$$

and in the region $\left\{(x, y) \in S_t : \frac{1}{2\sqrt{|t|}} \leq |x| \leq \frac{2}{\sqrt{|t|}}\right\}$ we interpolate between u^0 and u^∞ while remaining close to $u^0(\infty) = u^\infty(0)$. It is possible to chose the interpolation compatibly with the S^1-actions on $\widetilde{\mathcal{M}}^{\mathrm{red}}(p_0, H_\infty)$ and $\mathfrak{X}_\epsilon$ so that the map $\widetilde{\mathfrak{p}}(u^0, u^\infty, t)$ induces a well defined map $\mathfrak{p}(e) \colon \varpi^{-1}(e) \to \overline{W}$.

If we choose a representative (u^0, u^∞, t) of e where t is real positive and we define $R = \frac{1}{\sqrt{t}}$, we see that the pregluing $\mathfrak{p}(e)$ is the same as the pregluing defined in [9, Section 10:2], up to a holomorphic change of coordinates and the introduction of a constant δ, which is necessary for the gluing estimates, but does not change in any significant way the geometric picture we have described.

Combining [9, Theorem 6.2.6] with the discussion above we obtain the following structural result for the moduli spaces we are interested in.

Proposition 2.6. *The moduli spaces $\overline{\mathcal{M}}(p_0, H_\infty)$ and $\overline{\mathcal{M}}_z(p_0, H_\infty)$ are closed and orientable C^1-manifolds and there is a C^1-map*

$$\bar{\mathfrak{f}}\colon \overline{\mathcal{M}}_z(p_0, H_\infty) \to \overline{\mathcal{M}}(p_0, H_\infty),$$

which forgets the marked point.

While $\overline{\mathcal{M}}(p_0, H_\infty)$ is not a priori connected, since we have not ruled out that a J-holomorphic sphere could be homotopic to a line $\ell \subset W_\infty$ but not homotopic through J-holomorphic spheres, we can assume without loss of generality that $\overline{\mathcal{M}}(p_0, H_\infty)$ is connected by restricting our attention to the connected component which contains a line $\ell \subset W_\infty$.

3. Proof of the main theorem

3.1. Degree of the evaluation map

Let $\mathrm{ev}\colon \overline{\mathcal{M}}_z(p_0, H_\infty) \to \overline{W}$ be the evaluation map at the free marked point.

Lemma 3.1. *There is an open subset $U \subset \overline{W}$ such that every J-holomorphic sphere of $\overline{\mathcal{M}}(p_0, H_\infty)$ passing through a point of U belongs to $\mathcal{M}(p_0, H_\infty)$ and its image is contained in the neighbourhood of W_∞ on which J is integrable.*

Proof. Choose a point $q_0 \in W_\infty \setminus H_\infty$ such that $q_0 \neq p_0$. The unique line ℓ_0 in W_∞ passing through p_0 and q_0 also intersects H_∞, and therefore determines an element of $\mathcal{M}(p_0, H_\infty)$. Moreover, any sphere of $\mathcal{M}(p_0, H_\infty)$ passing through q_0 intersects W_∞ in three points, and therefore must be contained in it, so it is equal to ℓ_0.

Since none of the nodal spheres passes through q_0, and since $\mathcal{M}_z^{\mathrm{red}}(p_0, H_\infty)$ is compact, there is a neighbourhood U of q_0 in $\overline{W}$ such that $\mathrm{ev}^{-1}(U) \subset \mathcal{M}_z(p_0, H_\infty)$.

After possibly reducing the size of U, we can assume that every J-holomorphic sphere of $\mathcal{M}(p_0, H_\infty)$ passing through U is contained in the neighbourhood of W_∞ on which J is integrable. Suppose on the contrary that there is a sequence $[u_n]$ of elements of $\mathcal{M}(p_0, H_\infty)$ and a sequence of points $q_n \in \overline{W}$ converging to q_0 such that the image of u_n contains q_n, but is not contained in some fixed neighbourhood of W_∞. Then, by Gromov compactness, there is a subsequence of $[u_n]$ converging to a (possibly nodal) J-holomorphic sphere of $\overline{\mathcal{M}}(p_0, H_\infty)$ passing through q_0 and not contained in the fixed neighbourhood of W_∞. This is a contradiction, because the only element of $\overline{\mathcal{M}}(p_0, H_\infty)$ passing through q_0 is ℓ_0, which is contained in W_∞. □

Lemma 3.2. *The evaluation map* $\mathrm{ev}\colon \overline{\mathcal{M}}_z(p_0, H_\infty) \to \overline{W}$ *has degree one.*

Proof. Let U be the neighbourhood defined in Lemma 3.1. We will show that $\#\,\mathrm{ev}^{-1}(q) = 1$ for every $q \in U$.

Since all J-holomorphic spheres passing through U are contained in the neighbourhood where J is integrable, we can pretend we are working in the total space of $\mathcal{O}_{\mathbb{P}^{n-1}}(2)$. Given $q \in U$, let $\overline{q}$ be its projection to $\mathbb{CP}^{n-1} \cong W_\infty$. Any J-holomorphic sphere of $\mathcal{M}(p_0, H_\infty)$ passing through q projects to the unique line ℓ_q in W_∞ passing through p_0 and $\overline{q}$. The sphere itself corresponds then to a section of $\mathcal{O}_{\mathbb{P}^{n-1}}(2)|_{\ell_q} \cong \mathcal{O}_{\mathbb{P}^1}(2)$ which vanishes at p_0 and at $p_\infty = \ell_q \cap H_\infty$. The space of sections of $\mathcal{O}_{\mathbb{P}^1}(2)$ vanishing at p_0 and p_∞ has complex dimension one, and thus there is a unique such section for any point q in the fibre of $\mathcal{O}_{\mathbb{P}^1}(2)$ over $\overline{q}$.

This shows that $\#\,\mathrm{ev}^{-1}(q) = 1$ for every $q \in U$, and since U is open, by Sard's theorem it contains a regular value of the evaluation map. This proves that ev has degree one. □

It is important to have a degree one map because such maps induce surjections in homology. More generally, we have the following lemma.

Lemma 3.3. *Let $f\colon X \to Y$ be a smooth map between closed oriented smooth manifolds of the same dimension. Assume that f has degree d, and let $S \subset Y$ be a compact, oriented k-dimensional submanifold that is transverse to f.*

Then it follows that $S' := f^{-1}(S)$ has an induced orientation and, with that orientation, we have the equality

$$f_*\big([S']\big) = d\,[S],$$

in $H_k(Y;\mathbb{Z})$.

Proof. A submanifold S is transverse to a map f if, for every $y \in S$ and $x \in f^{-1}(y)$ we have $T_yS \oplus d_xf(T_xX) = T_yY$. This property implies that

- $S' = f^{-1}(S)$ is a compact submanifold of X, and
- df defines an isomorphism between the normal bundle of S' and the normal bundle of S.

The orientations of S and Y determine an orientation of the normal bundle of S. This in turn induces an orientation of the normal bundle of S' via df which, combined with the orientation of X, induces the orientation of S'.

Let $f_S\colon S' \to S$ be the restriction of f. The condition on the normal bundles implies that the regular values of f_S are also regular values of f. If y is a regular value of f_S, then

$$\deg(f_S) = \sum_{x \in f_S^{-1}(y)} \mathrm{sign}(d_xf_S)$$

$$\deg(f) = \sum_{x \in f^{-1}(y)} \mathrm{sign}(d_xf).$$

Since $f_S^{-1}(y) = f^{-1}(y)$ by the definition of f_S and $\mathrm{sign}(d_xf_S) = \mathrm{sign}(d_xf)$ because df is an orientation preserving isomorphism between the normal bundles, we obtain $\deg(f_S) = \deg(f) = d$.

Now we consider the commutative diagram

$$\begin{tikzcd} H_k(S';\mathbb{Z}) \arrow[r, "(f_S)_*"] \arrow[d] & H_k(S;\mathbb{Z}) \arrow[d] \\ H_k(X;\mathbb{Z}) \arrow[r, "f_*"] & H_k(Y;\mathbb{Z}) \end{tikzcd},$$

where the vertical arrows are induced by the inclusions. The fundamental class of S' is mapped by $(f_S)_*$ to $\deg(f_S)$ times the fundamental class of S. The homology classes $[S']$ and $[S]$ are the images of the fundamental classes of S' and S in $H_k(X;\mathbb{Z})$ and $H_k(Y;\mathbb{Z})$, respectively, and therefore $f_*[S'] = \deg(f_S)\,[S] = d\,[S]$. □

3.2. Decomposition of the line

The following lemma is a warm up which illustrates how to derive topological implications from Lemma 3.2.

Lemma 3.4. *The moduli space $\mathcal{M}_z(p_0, H_\infty)$ is not compact.*

Proof. The moduli space $\mathcal{M}_z(p_0, H_\infty)$ is an S^2-bundle over $\mathcal{M}(p_0, H_\infty)$ with two distinguished sections $\mathrm{ev}^{-1}(p_0)$ and $\mathrm{ev}^{-1}(H_\infty)$. Then $\mathcal{M}_z(p_0, H_\infty) \setminus \mathrm{ev}^{-1}$ (H_∞) retracts onto $\mathrm{ev}^{-1}(p_0)$. This implies that

$$\mathrm{ev}_*\colon H_k\big(\mathcal{M}_z(p_0, H_\infty) \setminus \mathrm{ev}^{-1}(H_\infty);\mathbb{Z}\big) \to H_k(\overline{W};\mathbb{Z}),$$

is trivial whenever $k > 0$.

Let $\ell \subset \overline{W}$ be an embedded sphere which is homologous to a line in W_∞ but disjoint from H_∞. It is possible to find such a sphere, because H_∞ has codimension 4 in $\overline{W}$, but, in general, ℓ will not be holomorphic. We perturb ℓ to be transverse to the evaluation map and we denote $\mathrm{ev}^{-1}(\ell)$ by ℓ'. If $\mathcal{M}_z(p_0, H_\infty)$ is compact, $\mathrm{ev}_*([\ell']) = [\ell]$ by Lemma 3.3. Since $\ell \cap H_\infty = \emptyset$, it follows that ℓ' does not intersect $\mathrm{ev}^{-1}(H_\infty)$. The previous paragraph implies then that $[\ell] = \mathrm{ev}_*([\ell']) = 0$. This is a contradiction, because ℓ is homologous to a symplectic sphere, and therefore, is nontrivial in homology. □

Lemma 3.4 tells us thus that $\mathcal{M}^{\mathrm{red}}(p_0, H_\infty)$ is nonempty. We decompose it into connected components

$$\mathcal{M}^{\mathrm{red}}(p_0, H_\infty) = \mathcal{M}^{(1)}(p_0, H_\infty) \sqcup \cdots \sqcup \mathcal{M}^{(N)}(p_0, H_\infty),$$

and, correspondingly, we decompose the moduli space with a free marked point into connected components

$$\mathcal{M}_z^{\mathrm{red}}(p_0, H_\infty) = \mathcal{M}_z^{(1)}(p_0, H_\infty) \sqcup \cdots \sqcup \mathcal{M}_z^{(N)}(p_0, H_\infty).$$

Each $\mathcal{M}_z^{(i)}(p_0, H_\infty)$ is an $S^2 \vee S^2$-bundle over $\mathcal{M}^{(i)}(p_0, H_\infty)$ with three distinguished sections: one, denoted $\mathcal{S}_0^{(i)}$, where the free marked point is mapped to p_0, one, denoted $\mathcal{S}_\infty^{(i)}$, where the free marked point is mapped to H_∞, and one, denoted $\mathcal{S}_{\mathfrak{n}}^{(i)}$, where the free marked point lies on the node.[1] Therefore, we can see each $\mathcal{M}_z^{(i)}(p_0, H_\infty)$ as the union of two sphere bundles $\mathcal{N}_0^{(i)}$ and

[1]Strictly speaking ghost bubbles appear in these three cases and we tacitly contract them. We ignore this technical complication as it has no topological consequence.

$\mathcal{N}_\infty^{(i)}$ over $\mathcal{M}^{(i)}(p_0, H_\infty)$ glued together along the section $\mathcal{S}_\mathfrak{n}^{(i)}$. An element of $\mathcal{M}_z^{(i)}(p_0, H_\infty)$ belongs to $\mathcal{N}_0^{(i)}$ when the free marked point lies in the domain of the irreducible component passing through p_0, and to $\mathcal{N}_\infty^{(i)}$ when the free marked point lies in the domain of the irreducible component passing through H_∞. We denote the homology classes representing the fibres of $\mathcal{N}_0^{(i)}$ and of $\mathcal{N}_\infty^{(i)}$ by $A_0^{(i)}$ and by $A_\infty^{(i)}$, respectively.

Our aim is to show that there is a nodal curve in the compactification of $\mathcal{M}(p_0, H_\infty)$ that is composed of two holomorphic spheres representing homology classes which are equal up to torsion.

A nodal curve in $\mathcal{M}^{(i)}(p_0, H_\infty)$ is composed of two holomorphic spheres that are fibres of $\mathcal{N}_*^{(i)}$ for $* = 0$ or $* = \infty$ so that $\mathrm{ev}_*\big(A_0^{(i)}\big) + \mathrm{ev}_*\big(A_\infty^{(i)}\big) = [\ell]$ in $H_2(\overline{W}; \mathbb{Z})$.

The pull-back of the symplectic form $\overline{\omega}$ is cohomologically nontrivial on the fibres of $\mathcal{N}_*^{(i)}$ for any $i = 1, \dots, N$ and $* \in \{0, \infty\}$. Therefore, by the Leray-Hirsch Theorem (see [2, Theorem 5.11] for its cohomological form),

$$H_2(\mathcal{N}_*^{(i)}; \mathbb{Z}) \cong H_2\big(\mathcal{S}_*^{(i)}; \mathbb{Z}\big) \oplus H_2(S^2; \mathbb{Z}) \cong H_2\big(\mathcal{S}_\mathfrak{n}^{(i)}; \mathbb{Z}\big) \oplus H_2(S^2; \mathbb{Z}), \quad (2)$$

where the summand $H_2(S^2; \mathbb{Z})$ is generated by a fibre of $\mathcal{N}_*^{(i)}$.

In the next two lemmas, we use the Leray-Hirsch Theorem to gain homological information on the evaluation maps, and on the components of the nodal curves. Given homology classes A and B of complementary degrees (in the same manifold), we denote by $A \cdot B$ their intersection product. If A and B are represented by closed, oriented submanifolds which intersect transversely, $A \cdot B$ is the algebraic count of intersection points.

Lemma 3.5. *The map*

$$(\mathrm{ev}\,|_{\mathcal{S}_\mathfrak{n}^{(i)}})_* \colon H_2\big(\mathcal{S}_\mathfrak{n}^{(i)}; \mathbb{Z}\big) \to H_2(\overline{W}; \mathbb{Z}),$$

is trivial for every $i = 1, \dots, N$.

Proof. By Eq. (2) every class $c \in H_2(\mathcal{S}_\mathfrak{n}^{(i)}; \mathbb{Z})$ can be written as the sum of a class in $H_2(\mathcal{S}_0^{(i)}; \mathbb{Z})$ and a multiple of the class of the fibre. Since $\mathcal{S}_0^{(i)}$ is mapped to p_0, we obtain $(\mathrm{ev}\,|_{\mathcal{S}_0^{(i)}})_* = 0$, and thus $\mathrm{ev}_*(c) = k\ \mathrm{ev}_*\big(A_0^{(i)}\big)$. By Lemma 2.4 $\mathrm{ev}\big(\mathcal{S}_\mathfrak{n}^{(i)}\big) \cap W_\infty = \emptyset$ while $\mathrm{ev}_*\big(A_0^{(i)}\big) \cdot [W_\infty] = 1$ so that

$$0 = \mathrm{ev}_*(c) \cdot [W_\infty] = k\ \mathrm{ev}_*\big(A_0^{(i)}\big) \cdot [W_\infty] = k.$$

□

Let $\mathrm{ev}_\infty^{(i)} \colon \mathcal{S}_\infty^{(i)} \to H_\infty$ be the restriction of $\mathrm{ev} \colon \overline{\mathcal{M}}_z(p_0, H_\infty) \to \overline{W}$.

Lemma 3.6. *If* $\deg(\mathrm{ev}_\infty^{(i)}) \neq 0$, *then* $[\ell] = 2\,\mathrm{ev}_*\big([A_\infty^{(i)}]\big)$ *modulo torsion.*

Proof. Let ℓ be now a line that lies in H_∞ and perturb it (inside H_∞) to make it transverse to $\mathrm{ev}_\infty^{(i)} \colon \mathcal{S}_\infty^{(i)} \to H_\infty$. Then, $\ell_i' := (\mathrm{ev}_\infty^{(i)})^{-1}(\ell)$ is a smooth submanifold of $\mathcal{S}_\infty^{(i)}$ which, by Lemma 3.3, satisfies $\big(\mathrm{ev}_\infty^{(i)}\big)_*([\ell_i']) = \kappa_i\,[\ell]$ with $\kappa_i := \deg\big(\mathrm{ev}_\infty^{(i)}\big)$. Using that ev and $\mathrm{ev}_\infty^{(i)}$ commute with the corresponding inclusions, we also obtain $\mathrm{ev}_*([\ell_i']) = \kappa_i\,[\ell]$.

According to Equation (2), we can represent $[\ell_i']$ as

$$[\ell_i'] = d\,A_\infty^{(i)} + c,$$

for some $d \in \mathbb{Z}$ and $c \in H_2(\mathcal{S}_\mathfrak{n}^{(i)}; \mathbb{Z})$. This combined with Lemma 3.5 shows that $d\ \mathrm{ev}_*\big(A_\infty^{(i)}\big) = \kappa_i\,[\ell]$, and by intersecting with W_∞ we obtain $d = 2\kappa_i$ so that

$$\kappa_i\,\big(2\ \mathrm{ev}_*\big(A_\infty^{(i)}\big) - [\ell]\big) = 0.$$

□

The next step is to show that $\deg(\mathrm{ev}_\infty^{(i)}) \neq 0$ for at least one $i \in \{1, \ldots, N\}$. This will be the goal of the next lemmas.

Lemma 3.7. *Let X be a compact oriented n-dimensional manifold containing two closed oriented submanifolds S and Y. Suppose that $\dim S + \dim Y = n$, that $\dim Y \geq 2$, and that S and Y intersect transversely.*

Then, there is an oriented submanifold S' that is homologous to S and that intersects Y transversely in exactly $\big|[S]\cdot[Y]\big|$ many points.

More precisely, we can choose an arbitrarily small neighbourhood of Y such that S and S' agree outside this neighbourhood. Furthermore, given a compact subset $Y' \subset X$ that is disjoint from S, and that intersects Y in a codimension 2 submanifold, we can additionally assume that S' is also disjoint from Y'.

Proof. We obtain S' by attaching certain 1-handles to S.

If the number of intersection points of S and Y does not agree with $\big|[S]\cdot[Y]\big|$, then there needs to be a pair of intersection points $\{x_-, x_+\} \subset S \cap Y$ of opposite sign. Choose an embedded path γ in Y with end points x_- and x_+ that avoids any other intersection point in $S \cap Y$ and also $Y' \cap Y$, if such a Y' has been chosen.

Identify a tubular neighbourhood of Y with the normal bundle of Y, and assume that S corresponds in this neighbourhood to the fibres of the normal bundle over the points in $S \cap Y$. Note that the normal bundle is naturally oriented by the orientations of Y and X.

The restriction of the disk bundle over γ is a solid cylinder $D^k \times [0,1]$ such that $D^k \times \{0,1\}$ is a neighbourhood of $\{x_-, x_+\}$ in S. The solid cylinder is naturally oriented, and the orientation of S at $\{x_-, x_+\}$ is equal to the boundary orientation of $D^k \times [0,1]$.

Remove $D^k \times \{0,1\}$ from S, and glue instead the tube $(\partial D^k) \times [0,1]$ along the boundary of the two holes that we have created in S (abstractly this corresponds to performing an index 1 surgery). This yields, after smoothing, an oriented closed manifold S' that agrees outside the chosen tubular neighbourhood of Y with S. We can do this construction also avoiding Y' if necessary. Note that $S' \cap Y = (S \cap Y) \setminus \{x_-, x_+\}$, and that S and S' are homologous, because $[S'] - [S] = \big[\partial(D^k \times [0,1])\big]$.

By repeating this construction as often as necessary, we can cancel all pairs of intersection points of opposite sign until all points in $S' \cap Y$ have the same sign. This then implies as desired that $\#\big(S \cap Y\big) = \big|[S]\cdot[Y]\big|$. □

Lemma 3.8. *Let ℓ be a surface in $\overline{W}$ that is transverse to the evaluation map* ev*. Denote the oriented submanifold* $\mathrm{ev}^{-1}(\ell)$ *in* $\overline{\mathcal{M}}_z(p_0, H_\infty)$ *by* ℓ'*.*

If the intersection product $[\ell'] \cdot [\mathcal{N}_\infty^{(i)}]$ *is trivial for all* $i = 1, \dots, N$*, then it follows that* ℓ *is null-homologous.*

Proof. Generically, ℓ is disjoint from H_∞, so we may assume that ℓ' does not intersect $\mathrm{ev}^{-1}(H_\infty)$, and after a further perturbation we can assume that ℓ' is transverse to $\mathcal{N}_\infty^{(i)}$ without changing the homology class of ℓ'.

If $[\ell'] \cdot [\mathcal{N}_\infty^{(i)}] = 0$, we can apply Lemma 3.7 to find a surface ℓ'' in $\overline{\mathcal{M}}_z(p_0, H_\infty)$ that is homologous to ℓ' and that does not have any intersection points either with $\mathcal{N}_\infty^{(i)}$ or with $\mathrm{ev}^{-1}(H_\infty)$. Furthermore, since this modification has been performed in an arbitrarily small neighbourhood of $\mathcal{N}_\infty^{(i)}$, we may assume that we have not created any new intersection points with one of the other components $\mathcal{N}_\infty^{(j)}$ for $j \neq i$.

Thus, if the intersection product $[\ell'] \cdot [\mathcal{N}_\infty^{(i)}]$ is trivial for all $i = 1, \dots, N$, we obtain by successively applying this construction for each i a surface ℓ'' in $\overline{\mathcal{M}}_z(p_0, H_\infty)$ with $[\ell''] = [\ell']$ that does not intersect any of the $\mathcal{N}_\infty^{(i)}$ or $\mathrm{ev}^{-1}(H_\infty)$.

We then have that $\mathrm{ev}_*([\ell'']) = 0$ as in the proof of Lemma 3.4, because $\overline{\mathcal{M}}_z(p_0, H_\infty) \setminus \left(\bigcup_i \mathcal{N}_\infty^{(i)} \cup \mathrm{ev}^{-1}(H_\infty)\right)$ retracts to $\mathrm{ev}^{-1}(p_0)$, but due to Lemma 3.3 we see that $\mathrm{ev}_*([\ell']) = [\ell]$. Since $[\ell''] = [\ell']$, it follows that $[\ell] = 0$. □

Lemma 3.9. *Let $\ell \subset \overline{W}$ be a surface that is transverse to the evaluation map* ev *and that represents the homology class of a line in* W_∞*. Then, it follows for* $\ell' = \mathrm{ev}^{-1}(\ell)$ *that*

$$[\ell'] \cdot \left[\mathcal{N}_\infty^{(i)}\right] = \deg(\mathrm{ev}_\infty^{(i)}).$$

Proof. Let $y \in H_\infty$ be a regular value of $\mathrm{ev}_\infty^{(i)}$ for all $i = 1, \dots, N$, and let ℓ_0 be a line in W_∞ intersecting H_∞ transversely at y. It follows that ℓ_0 is transverse to $\mathrm{ev}|_{\mathcal{N}_\infty^{(i)}}$ at y, that is, for every $x \in \mathrm{ev}^{-1}(y) \cap \mathcal{N}_\infty^{(i)}$ we have

$$T_y \ell_0 \oplus d_x \,\mathrm{ev}\left(T_x \mathcal{N}_\infty^{(i)}\right) = T_y \overline{W}, \tag{3}$$

because the nodal J-holomorphic spheres in $\overline{\mathcal{M}}(p_0, H_\infty)$ are all transverse to W_∞.

By construction $\left(\mathrm{ev}_\infty^{(i)}\right)^{-1}(y) = \left(\mathrm{ev}|_{\mathcal{N}_\infty^{(i)}}\right)^{-1}(\ell_0)$. If $x \in \left(\mathrm{ev}|_{\mathcal{N}_\infty^{(i)}}\right)^{-1}(\ell_0)$, we define $\mathrm{sign}(x) = +1$ if the equality of Equation (3) preserves the orientation, and $\mathrm{sign}(x) = -1$ otherwise. Then $\mathrm{sign}(x) = \mathrm{sign}(d_x \,\mathrm{ev}_\infty^{(i)})$ because $d_x \,\mathrm{ev}$ is complex linear in the extra direction $T_x \mathcal{N}_\infty^{(i)} / T_x \mathcal{S}_\infty^{(i)}$.

Now let ℓ be a small perturbation of ℓ_0 which is transverse to ev. By Equation (3), we can assume that the perturbation is supported away from y and that no new intersection points between $\mathrm{ev}^{-1}(\ell)$ and $\mathcal{N}_\infty^{(i)}$ are created. Then

$$[\mathcal{N}_\infty^{(i)}]\cdot[\mathrm{ev}^{-1}(\ell)] = \sum_{x\in\left(\mathrm{ev}|_{\mathcal{N}_\infty^{(i)}}\right)^{-1}(\ell_0)} \mathrm{sign}(x) = \sum_{x\in\left(\mathrm{ev}_\infty^{(i)}\right)^{-1}(y)} \mathrm{sign}(d_x\,\mathrm{ev}_\infty^{(i)}) = \deg(\mathrm{ev}_\infty^{(i)}).$$

□

Lemma 3.10. *There exists an* $i\in\{1,\dots,N\}$ *such that* $\deg(\mathrm{ev}_\infty^{(i)})\neq 0$.

Proof. Let ℓ be a surface in $\overline{W}$ that is transverse to the evaluation map ev and that represents the homology class of a line in W_∞. Denote $\mathrm{ev}^{-1}(\ell)$ by ℓ'. By Lemma 3.9, $\deg(\mathrm{ev}_\infty^{(i)}) = [\ell']\cdot\left[\mathcal{N}_\infty^{(i)}\right]$. Thus, if $\deg(\mathrm{ev}_\infty^{(i)})$ were 0 for every $i=1,\dots,N$, it would follow from Lemma 3.8 that $[\ell]=0$. But this is impossible because the symplectic form evaluates positively on $[\ell]$. □

After all this preparation, the proof of Theorem 1.1 is falling at our feet like a ripe fruit.

Proof. (Proof of Theorem 1.1) By Lemma 3.10 there exists an $i\in\{1,\dots,N\}$ such that $\deg(\mathrm{ev}_\infty^{(i)})\neq 0$. Then, by Lemma 3.6, $[\ell]=2\,\mathrm{ev}_*\big([A_\infty^{(i)}]\big)$ modulo torsion. If we evaluate the first Chern class of $T\overline{W}$ on $[\ell]$ we obtain

$$n+2 = \big\langle c_1(T\overline{W}),[\ell]\big\rangle = 2\,\big\langle c_1(T\overline{W}),\mathrm{ev}_*\big([A_\infty^{(i)}]\big)\big\rangle,$$

which is a contradiction when n is odd. □

4. Fundamental group of semipositive fillings

In this section, let (W,ω) be a semipositive filling of $(\mathbb{RP}^{2n-1},\xi)$. We recall that (W,ω) is semipositive if every class A in the image of the Hurewicz homomorphism $\pi_2(W)\to H_2(W;\mathbb{Z})$ satisfying the conditions $\langle\omega,A\rangle>0$ and $\langle c_1(TW),A\rangle\geq 3-n$ also satisfies $\langle c_1(TW),A\rangle\geq 0$. See [9, Definition 6.4.1].

We use the same compactification $(\overline{W},\overline{\omega})$ and the same set of almost complex structures $\mathcal{J}$ as in the previous sections, but now that (W,ω) does not need to be symplectically aspherical we cannot assume anymore that $\overline{\mathcal{M}}(p_0,H_\infty)$ is a manifold or that its elements have no irreducible component contained completely inside $\overline{W}\setminus W_\infty$. The irreducible components which intersect W_∞ must be simply covered because the intersections are simple, and therefore are Fredholm regular for a generic almost complex structure $J\in\mathcal{J}$, but the irreducible components which are contained in $\overline{W}\setminus W_\infty$ can be multiply covered. However, according to [9, Theorem 6.6.1], the image of $\mathcal{M}^{\mathrm{red}}(p_0,H_\infty)$ under the evaluation map is contained in the union of images of finitely many compact codimension two smooth manifolds for a generic $J\in\mathcal{J}$ because the irreducible components intersecting W_∞ are Fredholm regular and the irreducible components contained in $\overline{W}\setminus W_\infty$ are controlled by semipositivity. In particular, $\overline{W}\setminus\mathrm{ev}(\mathcal{M}^{\mathrm{red}}(p_0,H_\infty))$ is open, dense and connected. Moreover the restriction of the evaluation map

$$\mathrm{ev}\colon \overline{\mathcal{M}}(p_0,H_\infty)\setminus\mathrm{ev}^{-1}\big(\mathrm{ev}(\mathcal{M}^{\mathrm{red}}(p_0,H_\infty))\big)\to\overline{W}\setminus\mathrm{ev}\big(\mathcal{M}^{\mathrm{red}}(p_0,H_\infty)\big),$$

is proper by Gromov compactness, and therefore its degree is well defined. Then, Lemma 3.2 can be rephrased as follows.

Lemma 4.1. *If* (W, ω) *is semipositive and* $y \in \overline{W} \setminus \mathrm{ev}(\mathcal{M}^{\mathrm{red}}(p_0, H_\infty))$ *is a regular value of* ev, *then*

$$\sum_{x \in \mathrm{ev}^{-1}(y)} \mathrm{sign}(d_x \, \mathrm{ev}) = 1.$$

In particular, $\mathrm{ev} \colon \overline{\mathcal{M}}_z(p_0, H_\infty) \to \overline{W}$ *is surjective.*

If we apply the argument of Lemma 3.4 to a 1-dimensional submanifold of W we obtain the following result.

Lemma 4.2. *If* (W, ω) *is a semipositive symplectic filling of* $(\mathbb{RP}^{2n-1}, \xi)$*, then the inclusion* $\iota \colon \mathbb{RP}^{2n-1} \to W$ *induces a surjective map* $\iota_* \colon \pi_1\big(\mathbb{RP}^{2n-1}\big) \to \pi_1(W)$*.*

Proof. Recall that $\overline{W} \setminus W_\infty$ is equal to $W \setminus \partial W$. Instead of proving that $\pi_1(\partial W)$ maps surjectively onto $\pi_1(W)$, we can equivalently show that for a sufficiently small neighbourhood U_ϵ of W_∞, $\pi_1(U_\epsilon \setminus W_\infty)$ is surjective in $\pi_1(\overline{W} \setminus W_\infty)$.

Choose a base point b for $\pi_1(\overline{W} \setminus W_\infty)$ that lies in the neighbourhood U of Lemma 3.2, and use $b' = \mathrm{ev}^{-1}(b)$ as the base point for $\pi_1(\mathcal{M}_z(p_0, H_\infty))$.

We can represent every element of $\pi_1\big(\overline{W} \setminus W_\infty\big)$ by a smooth embedding

$$\gamma \colon S^1 \hookrightarrow W,$$

that avoids $\mathrm{ev}\big(\mathcal{M}^{\mathrm{red}}(p_0, H_\infty)\big)$ by a codimension argument and that is transverse to the evaluation map. Using the fact that ev is a diffeomorphism of U onto its image and arguing as in point (i) of the proof of [5, Lemma2.3] we obtain a loop $\Gamma \colon S^1 \to \mathcal{M}_z(p_0, H_\infty)$ such that $\Gamma(1) = b'$ and $\mathrm{ev}_*([\Gamma]) = [\gamma]$ in $\pi_1(\overline{W} \setminus W_\infty)$. Furthermore Γ does not intersect any singular stratum or $\mathrm{ev}^{-1}(W_\infty)$.

We can isotope $\mathcal{M}_z(p_0, H_\infty) \setminus \mathrm{ev}^{-1}(H_\infty)$ into an arbitrarily small neighbourhood of $\mathrm{ev}^{-1}(p_0)$ by pushing the marked point in every holomorphic sphere from ∞ towards 0. This isotopy restricts to $\mathcal{M}_z(p_0, H_\infty) \setminus \mathrm{ev}^{-1}(W_\infty)$, so that Γ is homotopic in $\mathcal{M}_z(p_0, H_\infty) \setminus \mathrm{ev}^{-1}(W_\infty)$ to a loop in a neighbourhood of $\mathrm{ev}^{-1}(p_0)$.

Then, it follows that $[\gamma]$ is homotopic in $\overline{W} \setminus W_\infty$ to a loop that lies in an arbitrarily small neighbourhood of p_0 and $\pi_1(U_\epsilon \setminus W_\infty) \to \pi_1\big(\overline{W} \setminus W_\infty\big)$ is surjective. □

Combining this with the argument found in [3, Section 6.2] we obtain the main result of this section.

Theorem 4.3. *Any semipositive symplectic filling of* $(\mathbb{RP}^{2n-1}, \xi)$ *is simply connected.*

Proof. Let (W, ω) be a semipositive symplectic filling of $(\mathbb{RP}^{2n-1}, \xi)$. By Lemma 4.2 the map $\pi_1(\partial W) \to \pi_1(W)$ induced by the inclusion $\iota \colon \partial W \hookrightarrow W$ is surjective so that $\pi_1(W)$ is either trivial or isomorphic to $\mathbb{Z}/2\mathbb{Z}$.

In the latter case, ι induces an isomorphism between the fundamental groups, and thus

$$\iota^* \colon H^1(W; \mathbb{Z}/2\mathbb{Z}) \to H^1\big(\partial W; \mathbb{Z}/2\mathbb{Z}\big)$$

is also an isomorphism. Let $\alpha \in H^1(W;\mathbb{Z}/2\mathbb{Z})$ be the nontrivial element. Then $\iota^*\alpha \in H^1\big(\partial W;\mathbb{Z}/2\mathbb{Z}\big)$ is also nontrivial and, since $H^*(\mathbb{RP}^{2n-1};\mathbb{Z}/2\mathbb{Z})$ is generated as an algebra by the nontrivial element of degree one, $(\iota^*\alpha)^{2n-1}$ is the nontrivial element of $H^{2n-1}\big(\partial W;\mathbb{Z}/2\mathbb{Z}\big)$.

By the naturality of the cup product $(\iota^*\alpha)^{2n-1} = \iota^*(\alpha^{2n-1})$. However

$$\iota_*\colon H_{2n-1}\big(\partial W;\mathbb{Z}/2\mathbb{Z}\big) \to H_{2n-1}(W;\mathbb{Z}/2\mathbb{Z}),$$

is trivial, and consequently $\iota^*\colon H^{2n-1}(W;\mathbb{Z}/2\mathbb{Z}) \to H^{2n-1}\big(\partial W;\mathbb{Z}/2\mathbb{Z}\big)$ is also trivial by duality because we are working over a field. This contradicts $\iota^*(\alpha^{2n-1}) \neq 0$ and therefore shows that W is simply connected. □

5. Yet another proof of the Eliashberg-Floer-McDuff theorem

In this section we apply the constructions of this article to the symplectic fillings of the standard contact structure ξ on S^{2n-1}. This will lead to small changes in the meaning of the notation. If (W,ω) is a symplectic filling of (S^{2n-1},ξ) and we perform symplectic reduction of its boundary, we obtain a closed symplectic manifold $(\overline{W},\overline{\omega})$ with a codimension two symplectic submanifold $W_\infty \cong \mathbb{CP}^{n-1}$ whose normal bundle is isomorphic to $\mathcal{O}_{\mathbb{P}^{n-1}}(1)$. We choose an almost complex structure J on $\overline{W}$ which is integrable near W_∞ and generic elsewhere. Let $p_0 \in W_\infty$ be a point; we denote by $\mathcal{M}(p_0)$ the moduli space of unparametrised J-holomorphic spheres in $\overline{W}$ that are homotopic to a line in W_∞ and pass through p_0. If ℓ is a line in W_∞, then

$$T\overline{W}|_\ell \cong \mathcal{O}_{\mathbb{P}^1}(2) \oplus \underbrace{\mathcal{O}_{\mathbb{P}^1}(1) \oplus \cdots \oplus \mathcal{O}_{\mathbb{P}^1}(1)}_{n-1}.$$

Since $[\ell]\cdot[W_\infty] = 1$ all elements of $\mathcal{M}(p_0)$ are simply covered, and therefore $\mathcal{M}(p_0)$ is a smooth manifold by the analogue of Proposition 2.3. Let $\mathcal{M}_z(p_0)$ is the moduli space obtained by adding a free marked point to the elements of $\mathcal{M}(p_0)$. A Riemann-Roch calculation gives $\dim \mathcal{M}(p_0) = 2n-2$ and $\dim \mathcal{M}_z(p_0) = 2n$.

Lemma 5.1. *If (W,ω) is symplectically aspherical, then $\mathcal{M}_z(p_0)$ is compact.*

Proof. As the algebraic intersection between a line with W_∞ is one, any nodal J-holomorphic curve representing the homology class of a line must have an irreducible component in $\overline{W}\setminus W_\infty \cong W$. □

Lemma 3.2 still holds with the minimal necessary modifications, and therefore, the evaluation map $\mathrm{ev}\colon \mathcal{M}_z(p_0) \to \overline{W}$ has degree one.

Lemma 5.2. *If (W,ω) is a symplectically aspherical filling of (S^{2n-1},ξ), then $H_*(W;\mathbb{Z}) = 0$ for $* > 0$.*

Proof. The moduli space $\mathcal{M}_z(p_0)$ is an S^2-bundle over $\mathcal{M}(p_0)$ and $\mathrm{ev}^{-1}(p_0)$ is a section. Let $\widetilde{W}_\infty$ be a J-holomorphic hypersurface of $\overline{W}$ contained in the neighbourhood of W_∞ where J is integrable and obtained as the graph of a section of $\mathcal{O}_{\mathbb{P}^{n-1}}(1)$. We choose $\widetilde{W}_\infty$ such that $p_0 \notin \widetilde{W}_\infty$: then $\mathrm{ev}^{-1}(\widetilde{W}_\infty)$ is a section of $\mathcal{M}_z(p_0)$ which is disjoint from $\mathrm{ev}^{-1}(p_0)$. The map

$$\mathrm{ev}_*\colon H_*\big(\mathcal{M}_z(p_0) \setminus \mathrm{ev}^{-1}(\widetilde{W}_\infty); \mathbb{Z}\big) \to H_*\big(\overline{W} \setminus \widetilde{W}_\infty; \mathbb{Z}\big) \cong H_*(W; \mathbb{Z}), \quad (4)$$

is surjective by Lemma 3.3.

That lemma, strictly speaking, is about homology classes represented by submanifolds, but there are several ways to extend it to general homology classes.

On the other hand $\mathcal{M}_z(p_0) \setminus \mathrm{ev}^{-1}(\widetilde{W}_\infty)$ retracts onto $\mathrm{ev}^{-1}(p_0)$, and therefore the map (4) is trivial for $* > 0$. □

The proof of Lemma 4.2 works with the obvious modifications more or less unchanged for fillings of (S^{2n-1}, ξ), and therefore W is simply connected. Then the h-cobordism theorem implies the following corollary.

Corollary 5.3. *(Eliashberg–Floer–McDuff; see [8, Theorem 1.5]) If (W, ω) is a symplectically aspherical filling of (S^{2n-1}, ξ), then W is diffeomorphic to the ball D^{2n}.*

Funding The first author was partially supported by the ANR grant ANR-16-CE40-017 “Quantact”, the Simons Foundation and the Centre de Recherches Mathématiques, through the Simons-CRM scholar-in-residence program, and the Grant KAW 2019.0531 from the Knut and Alice Wallenberg Foundation. The second author was partially supported by the ANR Grant ANR-16-CE40-017 “Quantact”. The first author is grateful to the Mittag-Leffler Institute for its hospitality during the program “Knots, Strings, Symplectic Geometry and Dualities” and to the department of Mathematics of the University of Uppsala for hospitality in the Spring semester of 2021. The authors would also like to thank the referee for having carefully read the article.

References

[1] Barth, K., Geiges, H., Zehmisch, K.: The diffeomorphism type of symplectic fillings. J. Symplectic Geom. **17**(4), 929–971 (2019)

[2] Bott, R., Tu, L.W.: Differential Forms in Algebraic topology, Graduate Texts in Mathematics, vol. 82. Springer-Verlag, New York-Berlin (1982)

[3] Eliashberg, Y., Kim, S.S., Polterovich, L.: Geometry of contact transformations and domains: orderability versus squeezing. Geom. Topol. **10**, 1635–1747 (2006)

[4] Ghiggini, P.: Strongly fillable contact 3-manifolds without Stein fillings. Geom. Topol. **9**, 1677–1687 (2005)

[5] Ghiggini, P., Niederkrüger, K., Wendl, C.: Subcritical contact surgeries and the topology of symplectic fillings. J. Éc. polytech. Math. **3**, 163–208 (2016)

[6] Hirsch, M.: *Differential topology*, Graduate Texts in Mathematics, vol. 33, Springer-Verlag, New York, Corrected reprint of the 1976 original (1994)

[7] McDuff, D.: The structure of rational and ruled symplectic 4-manifolds. J. Am. Math. Soc. **3**(3), 679–712 (1990)

[8] McDuff, D.: Symplectic manifolds with contact type boundaries. Invent. Math. **103**(3), 651–671 (1991)

[9] McDuff, D., Salamon, D.: J-holomorphic curves and symplectic topology, second ed., American Mathematical Society Colloquium Publications, vol. 52, American Mathematical Society, Providence, RI, (2012)

[10] Oancea, A., Viterbo, C.: On the topology of fillings of contact manifolds and applications. Comment. Math. Helv. **87**(1), 41–69 (2012)

[11] Zhou, Z.: $(\mathbb{RP}^{2n-1}, \xi_{std})$ is not exactly fillable for $n \neq 2^k$. Geom. Topol. **25**(6), 3013–3052 (2021)

Paolo Ghiggini
Nantes Université, CNRS, Laboratoire de Mathématiques Jean Leray
LMJL, UMR 6629, F-44000 Nantes
France
e-mail: `paolo.ghiggini@univ-nantes.fr`

Klaus Niederkrüger-Eid
Institut Camille Jordan
Université Claude Bernard Lyon 1
43 boulevard du 11 novembre 1918
69622 Villeurbanne Cedex
France
e-mail: `niederkruger@math.univ-lyon1.fr`

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:35
https://doi.org/10.1007/s11784-022-00948-7
Published online April 12, 2022

Journal of Fixed Point Theory and Applications

On curves with the Poritsky property

Alexey Glutsyuk

To Professor Claude Viterbo on the occasion of his 60th birthday.

Abstract. Reflection in planar billiard acts on oriented lines. For a given closed convex planar curve γ, the string construction yields a one-parameter family Γ_p of nested billiard tables containing γ for which γ is a *caustic:* the reflection from Γ_p sends each tangent line to γ to a line tangent to γ. The reflections from Γ_p act on the corresponding tangency points, inducing a family of *string diffeomorphisms* $\mathcal{T}_p : \gamma \to \gamma$. We say that γ has the *string Poritsky property,* if it admits a parameter t (called the *Poritsky string length*) in which all the transformations $\mathcal{T}_p$ with small p are translations $t \mapsto t + c_p$. These definitions also make sense for germs of curves γ. The Poritsky property is closely related to the famous Birkhoff Conjecture. Each conic has the string Poritsky property. Conversely, *each germ of planar curve with the Poritsky property is a conic* (Poritsky, 1950). In the present paper, we extend this result of Poritsky to curves on surfaces of constant curvature and to outer billiards on all these surfaces. For curves with the Poritsky property on a surface with arbitrary Riemannian metric, we prove the following two results: *1) the Poritsky string length coincides with Lazutkin parameter up to additive and multiplicative constants; (2) a germ of C^5-smooth curve with the Poritsky property is uniquely determined by its 4-jet.* In the Euclidean case, the latter statement follows from the above-mentioned Poritsky's result on conics.

Mathematics Subject Classification. 37J35, 37D50.

CNRS, France (UMR 5669 (UMPA, ENS de Lyon) and UMI 2615 (Interdisciplinary Scientific Center J.-V.Poncelet))
HSE University, Moscow, Russian Federation.

The author is partially supported by Laboratory of Dynamical Systems and Applications, HSE University, of the Ministry of science and higher education of the RF grant ag. No 075-15-2019-1931.

Supported by part by RFBR grants 16-01-00748 and 16-01-00766.

Partially supported by RFBR and JSPS (research project 19-51-50005)

This material is based upon work supported by the National Science Foundation under Grant No. DMS-1440140, while the author was in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the fall semester 2018

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

Keywords. Billiard, caustic, Riemannian manifold, geodesics, Poritsky property and pararameter.

Contents

1. Introduction and main results

Consider the billiard in a bounded planar domain $\Omega \subset \mathbb{R}^2$ with a strictly convex smooth boundary. The billiard dynamics T acts on the space of oriented lines intersecting Ω. Namely, let L be an oriented line intersecting Ω, and let A be its last point (in the sense of orientation) of its intersection with $\partial\Omega$. By definition, $T(L)$ is the image of the line L under the symmetry with respect to the tangent line $T_A\partial\Omega$, being oriented from the point A inside the domain Ω. A curve $\gamma \subset \mathbb{R}^2$ is a *caustic* of the billiard Ω if each line tangent to γ is reflected from the boundary $\partial\Omega$ again to a line tangent to γ; in other words, if the curve formed by oriented lines tangent to γ is invariant under the billiard transformation T. In what follows, we consider only *smooth caustics* (in particular, without cusps).

It is well known that each planar billiard with sufficiently smooth strictly convex boundary has a Cantor family of caustics [15]. An analogous statement for outer billiards was proved in [2]. Every elliptic billiard is *Birkhoff caustic integrable,* that is, an inner neighborhood of its boundary is foliated by closed caustics. The famous Birkhoff Conjecture states the converse: the only Birkhoff caustic integrable planar billiards are ellipses. The Birkhoff Conjecture together with its extension to billiards on surfaces of constant curvature and its outer and projective billiard versions (due to Sergei Tabachnikov) are big open problems, see, e.g., [8,9,13,25] and references therein for history and related results. Recently, a Riemannian generalization of the Birkhoff Conjecture was suggested in [10, conjecture 1.2].

It is well known that each smooth convex planar curve γ is a caustic for a family of billiards $\Omega = \Omega_p$, $p \in \mathbb{R}_+$, whose boundaries $\Gamma = \Gamma_p = \partial\Omega_p$ are given by the p-th string constructions, see [24, p. 73]. Namely, let $|\gamma|$ denote the length of the curve γ. Take an arbitrary number $p > 0$ and a string of length $p + |\gamma|$ enveloping the curve γ. Let us put a pencil between the curve γ and the string, and let us push it out of γ to a position such that the string that envelopes γ and the pencil becomes tight. Then, let us move the pencil around the curve γ so that the string remains tight. Moving the pencil in this way draws a convex curve Γ_p that is called the *p-th string construction,* see Fig. 1.

For every $A \in \gamma$ by G_A, we denote the line tangent to γ at A. If γ is oriented by a vector in $T_A\gamma$, then we orient G_A by the same vector. The billiard reflection T_p from the curve Γ_p acts on the oriented lines tangent to γ. It induces the mapping $\mathcal{T}_p : \gamma \to \gamma$ acting on tangency points and called *string diffeomorphism.* It sends each point $A \in \gamma$ to the point of tangency of the curve γ with the line $T_p(G_A)$.

Consider the special case, where γ is an ellipse. Then, for every $p > 0$, the curve Γ_p given by the p-th string construction is an ellipse confocal to γ. Every ellipse γ admits a canonical bijective parametrization by the circle $S^1 = \mathbb{R}/2\pi\mathbb{Z}$ equipped with a parameter t such that for every $p > 0$ small

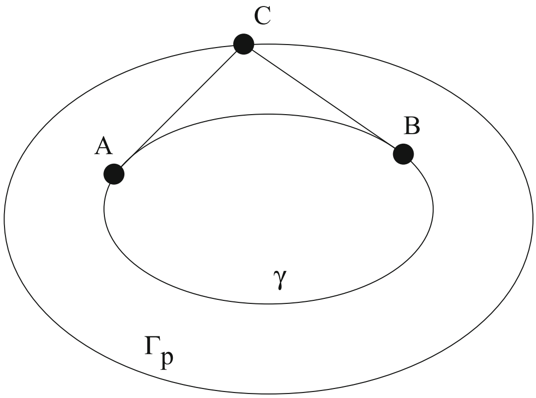


Figure 1. The string construction

enough one has $\mathcal{T}_p(t) = t + c_p$, $c_p = c_p(\gamma)$, see [24, the discussion before corollary 4.5]. The property of existence of the above parametrization will be called the *string Poritsky property,* and the parameter t will be called the *Poritsky–Lazutkin string length.*

In his seminal paper [21], Hillel Poritsky proved the Birkhoff Conjecture under the additional assumption called the *Graves* (or *evolution*) *property*: for every two nested caustics γ_λ, γ_μ of the billiard in question the smaller caustic γ_λ is also a caustic of the billiard in the bigger caustic γ_μ. His beautiful geometric proof was based on his remarkable theorem stating that in the Euclidean plane only conics have the string Poritsky property [21, section 7].

In the present paper, we extend this theorem by Poritsky to billiards on simply connected complete surfaces of constant curvature (Sects. 1.1 and 4) and prove its version for outer billiards and area construction on these surfaces (Sects. 1.2 and 5). All the results of the present paper will be stated and proved for germs of curves, and thus, in Sect. 1.1 (1.2), we state the definitions of the Poritsky string (area) property for germs. We also study the Poritsky property on arbitrary surfaces equipped with a Riemannian metric. In this general case, we show that the Poritsky string length coincides with the Lazutkin parameter

$$t_L(s) = \int_{s_0}^{s} \kappa^{\frac{2}{3}}(\zeta) d\zeta \tag{1.1}$$

introduced in [15, formula (1.3)], up to multiplicative and additive constants (Theorem 1.15 in Sect. 1.3, proved in Sect. 6). Here, κ is the geodesic curvature. This explains the name “Poritsky–Lazutkin length”.

Recall that the billiard ball map acting on the space of oriented geodesics preserves the canonical symplectic form (see the background material in Sect. 7.1). The above-mentioned Theorem 1.15 concerns the family of reflections from the string curves Γ_p, which is a family of symplectomorphisms having a common invariant curve: the curve of geodesics tangent to γ. In Sect. 7, we extend Theorem 1.15 to a more general class of symplectic maps: families of the so-called "weakly billiard-like maps" with a converging family of invariant curves (Theorem 7.10 stated in Sect. 7.2).

In [15], for a given curve $\gamma \subset \mathbb{R}^2$, Lazutkin introduced remarkable coordinates (x, y) on the space of oriented lines, in which the billiard ball map given by reflection from the curve γ takes the form

$$(x, y) \mapsto (x + y + o(y), y + o(y^2));$$

the x-axis coincides with the set of the lines tangent to γ;

$$x = t_L(s) \text{ on the } x-\text{axis up to multiplicative and additive constants.}$$

For the proof of Theorem 7.10, we use analogous coordinates for weakly billiard-like maps and prove Lemma 7.13 on asymptotic behavior of orbits in these coordinates (Sect. 7.3). We retrieve Theorem 1.15 (for C^6-smooth curves) from Theorem 7.10 at the end of Sect. 7.

Melrose and Marvizi [16] have shown that the billiard ball map given by a C^∞-smooth curve coincides with a unit time flow map of appropriately "time-rescaled" smooth Hamiltonian vector field, up to a flat correction.

For curves on arbitrary surface equipped with a C^6-smooth Riemannian metric, we show that a C^5-smooth germ of curve with the string Poritsky property is uniquely determined by its 4-jet (Theorem 1.19 stated in Sect. 1.4 and proved in Sect. 8). This extends similar property of planar conics.

Theorem 1.3 in Sect. 1.1 (proved in Sect. 3) states that if a metric and a germ of curve γ are both C^k, then the string curve foliation is tangent to a $C^{[\frac{k}{2}]-1}$-smooth line field on the closed concave side from γ.

In Sect. 2, we present a Riemannian-geometric background material on normal coordinates, equivalent definitions of geodesic curvature, etc. used in the proofs of main results.

1.1. The Poritsky property for string construction

Let Σ be a two-dimensional surface equipped with a Riemannian metric. Let $\gamma \subset \Sigma$ be a smooth curve (a germ of smooth curve at a point $O \in \Sigma$). We consider it to be *convex:* its geodesic curvature should be non-zero. For every given two points $A, B \in \gamma$ close enough by C_{AB} we will denote the unique point (close to them) of intersection of the geodesics G_A and G_B tangent to γ at A and B, respectively. (Its existence will be proved in Sect. 2.1.) Set

$$\begin{aligned} &\lambda(A, B) := \text{ the length of the arc } AB \text{ of the curve } \gamma, \\ &L(A, B) := |AC_{AB}| + |BC_{AB}| - \lambda(A, B). \end{aligned} \tag{1.2}$$

Here, for $X, Y \in \Sigma$ close enough and lying in a compact subset in Σ by $|XY|$, we denote the length of small geodesic segment connecting X and Y.

Definition 1.1 (equivalent definition of string construction). Let $\gamma \subset \Sigma$ be a germ of curve with non-zero geodesic curvature. For every $p \in \mathbb{R}_+$ small enough, the subset

$$\Gamma_p := \{C_{AB} \mid L(A, B) = p\} \subset \Sigma$$

is called the *p-th string construction,* see [24, p.73].

Remark 1.2. For every $p > 0$ small enough Γ_p is a well-defined smooth curve, we set $\Gamma_0 = \gamma$. The curve γ is a caustic for the billiard transformation acting by reflection from the curve Γ_p: a geodesic tangent to γ is reflected from the curve Γ_p to a geodesic tangent to γ [24, theorem 5.1]. In Sect. 3, we will prove the following theorem.

Theorem 1.3. *Let $k \geq 2$, Σ be a C^{k+1}-smooth surface equipped with a C^k-smooth Riemannian metric, and let $\gamma \subset \Sigma$ be a germ of C^k-smooth curve at $O \in \Sigma$ with positive geodesic curvature. Let $\mathcal{U} \subset \Sigma$ denote a small domain adjacent to γ from the concave side. For every $C \in \mathcal{U}$, let $\Lambda(C) \subset T_C\Sigma$ denote the exterior bisector of the angle formed by the two geodesics through C that are tangent to γ. Then, the following statements hold.*

1. *The one-dimensional subspaces $\Lambda(C)$ form a germ at O of line field Λ that is C^{k-1}-smooth on $\mathcal{U}$ and $C^{r(k)}$-smooth on $\overline{\mathcal{U}}$,*
$$r(k) = \left[\frac{k}{2}\right] - 1.$$
2. *The string curves Γ_p are tangent to Λ and $C^{r(k)+1}$-smooth. Their $(r(k)+1)$-jets at base points C depend continuously on $C \in \overline{\mathcal{U}}$.*

Definition 1.4. We say that a germ of oriented curve $\gamma \subset \Sigma$ with non-zero geodesic curvature has the *string Poritsky property,* if it admits a C^1-smooth parametrization by a parameter t (called the *Poritsky–Lazutkin string length*) such that for every $p > 0$ small enough there exists a $c = c_p > 0$ such that for every pair $B, A \in \gamma$ ordered by orientation with $L(A, B) = p$ one has $t(A) - t(B) = c_p$.

Example 1.5. It is classically known that

(i) For every planar conic $\gamma \subset \mathbb{R}^2$ and every $p > 0$, the p-th string construction Γ_p is a conic confocal to γ;
(ii) All the conics confocal to γ and lying inside a given string construction conic Γ_p are caustics of the billiard inside the conic Γ_p;
(iii) Each conic has the string Poritsky property [21, section 7], [24, p.58];
(iv) Conversely, *each planar curve with the string Poritsky property is a conic*, by a theorem of Poritsky [21, section 7].

Two results of the present paper extend statement (iv) to billiards on surfaces of constant curvature (by adapting Poritsky's arguments from [21, section 7]) and to outer billiards on the latter surfaces. To state them, let us recall the notion of a conic on a surface of constant curvature.

Without loss of generality we consider simply connected complete surfaces Σ of constant curvature 0, ± 1 and realize each of them in its standard model in the space $\mathbb{R}^3_{(x_1,x_2,x_3)}$ equipped with appropriate quadratic form

$$< Qx, x >, \ Q \in \{\mathrm{diag}(1,1,0), \mathrm{diag}(1,1,\pm 1)\}, \ < x, x >= x_1^2 + x_2^2 + x_3^2.$$

- Euclidean plane: $\Sigma = \{x_3 = 1\}$, $Q = \mathrm{diag}(1,1,0)$.
- The unit sphere: $\Sigma = \{x_1^2 + x_2^2 + x_3^2 = 1\}$, $Q = Id$.
- The hyperbolic plane: $\Sigma = \{x_1^2 + x_2^2 - x_3^2 = -1\} \cap \{x_3 > 0\}$, $Q = \mathrm{diag}(1,1,-1)$.

The metric of constant curvature on Σ is induced by the quadratic form $< Qx, x >$. The *geodesics* on Σ are its intersections with two-dimensional vector subspaces in $\mathbb{R}^3$. The *conics* on Σ are its intersections with quadrics $\{< Cx, x >= 0\} \subset \mathbb{R}^3$, where C is a real symmetric 3×3-matrix, see [12,29].

Proposition 1.6. *On every surface of constant curvature each conic has the string Poritsky property.*

Theorem 1.7. *Conversely, on every surface of constant curvature each germ of C^2-smooth curve with the string Poritsky property is a conic.*

Proposition 1.6 and Theorem 1.7 will be proved in Sect. 4.

Remark 1.8. In the case, when the surface under question is Euclidean plane, Proposition 1.6 was proved in [21, formula (7.1)], and Theorem 1.7 was proved in [21, section 7].

1.2. The Poritsky property for outer billiards and area construction

Let $\gamma \subset \mathbb{R}^2$ be a smooth strictly convex closed curve. Let $\mathcal{U}$ be the exterior connected component of the complement $\mathbb{R}^2 \setminus \gamma$. Recall that the *outer billiard map* $T : \mathcal{U} \to \mathcal{U}$ associated to the curve γ acts as follows. Take a point $A \in \mathcal{U}$. There are two tangent lines to γ through A. Let L_A denote the right tangent line (that is, the image of the line L_A under a small clockwise rotation around the point A is disjoint from the curve γ). Let $B \in \gamma$ denote its tangency point. By definition, the image $T(A)$ is the point of the line L_A that is central-symmetric to A with respect to the point B.

It is well known that if γ is an ellipse, then the corresponding outer billiard map is *integrable:* that is, an exterior neighborhood of the curve γ is foliated by invariant closed curves for the outer billiard map so that γ is a leaf of this foliation. The analogue of Birkhoff Conjecture for the outer billiards, which was suggested by Tabachnikov [25, p. 101], states the converse: if γ generates an integrable outer billiard, then it is an ellipse. Its polynomially integrable version was studied in [25] and recently solved in [11]. For a survey on outer billiards, see [23,24,28] and references therein.

For a given strictly convex smooth closed curve Γ, there exists a one-parametric family of curves γ_p such that γ_p lies in the interior component Ω of the complement $\mathbb{R}^2 \setminus \Gamma$, and the curve Γ is invariant under the outer billiard map T_p generated by γ_p. The curves γ_p are given by the following *area construction* analogous to the string construction. Let $\mathcal{A}$ denote the area of the domain Ω. For every oriented line ℓ intersecting Γ, let $\Omega_-(\ell)$ denote the connected component of the complement $\Omega \setminus \ell$ for which ℓ is a negatively oriented part of boundary. Let now L be a class of parallel and co-directed oriented lines. For every $p > 0$, $p < \frac{1}{2}\mathcal{A}$, let L_p denote the oriented line representing L that intersects Γ and such that $Area(\Omega_-(L_p)) = p$. For every given p, the lines L_p corresponding to different classes L form a one-parameter family parametrized by the circle: the azimuth of the line is the parameter. Let γ_p denote the enveloping curve of the latter family, and let T_p denote the outer billiard map generated by γ_p. It is well known that the curve Γ is T_p-invariant for every p as above [24, corollary 9.5]. See Fig. 2.

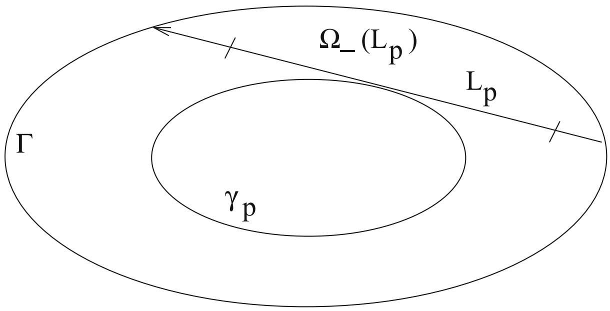


FIGURE 2. The area construction: $Area(\Omega_-(L_p)) \equiv p$

Remark 1.9. For every $p > 0$ small enough, the curve γ_p given by the area construction is smooth. But for big p, it may have singularities (e.g., cusps).

For Γ being an ellipse, all the γ_ps are ellipses homothetic to Γ with respect to its center. In this case, there exists a parametrization of the curve Γ by circle $S^1 = \mathbb{R}/2\pi\mathbb{Z}$ with parameter t in which $T_p : \Gamma \to \Gamma$ is a translation $t \mapsto t + c_p$ for every p. This follows from the area-preserving property of outer billiards, see [27, corollary 1.2], and T_q-invariance of the ellipse γ_p for $q > p$, analogously to the arguments in [21, section 7], [24, the discussion before corollary 4.5]. Similar statements hold for all conics, as in loc. cit.

In our paper, we prove the converse statement given by the following theorem, which will be stated in local context, for germs of smooth curves. To state it, let us introduce the following definition.

Definition 1.10. Let Σ be a surface with a smooth Riemannian metric, $O \in \Sigma$. Let $\Gamma \subset \Sigma$ be a germ of smooth strictly convex curve at a point O (i.e., with positive geodesic curvature). Let $U \subset \Sigma$ be a disk centered at O that is split by Γ into two components. One of these components is convex; let us denote it by V. Consider the curves γ_p given by the above area construction with $p > 0$ small enough and lines replaced by geodesics. The curves γ_p form a germ at O of foliation in the domain V, and its boundary curve $\Gamma = \gamma_0$ is a leaf of this foliation. We say that the curve Γ has the *area Poritsky property,* if it admits a local C^1-smooth parametrization by a parameter t called the *area Poritsky parameter* such that for every $p > 0$ small enough the mapping $T_p : \Gamma \to \Gamma$ is a translation $t \mapsto t + c_p$ in the coordinate t.

Proposition 1.11 (see [26, lemma 3] for the hyperbolic case; [27, lemma 5.1] for planar conics). *On every surface of constant curvature each conic has the area Poritsky property.*

Theorem 1.12 *Conversely, on every surface of constant curvature each germ of C^2-smooth curve with the area Poritsky property is a conic*[1].

[1]Earlier, in 2018, Tabachnikov proved Theorem 1.12 for planar C^5-smooth curves using a different, analytic method (showing constance of affine curvature). After the present paper was submitted, Arnold and Tabachnikov extended Tabachnikov's analytic proof to

Remark 1.13 (Tabachnikov) The area Poritsky property for conics on the sphere follows from their string Poritsky property and the fact that the spherical outer billiards are dual to the spherical Birkhoff billiards [23, subsection 4.1, lemma 5]: the duality is given by orthogonal polarity. Analogous duality holds on hyperbolic plane realized as the half-pseudo-sphere of radius -1 in 3-dimensional Minkowski space [6, section 2, remark 2]. As it was noticed in [3, end of section 3], in the spherical case the area Poritsky property is dual to the string Poritsky property. Therefore, in the spherical case Theorems 1.12 and 1.7 are dual and hence, equivalent.

1.3. Coincidence of the Poritsky and Lazutkin lengths

Everywhere in the subsection Σ is a two-dimensional surface equipped with a C^4-smooth Riemannian metric.

Definition 1.14 Let $\gamma \subset \Sigma$ be a C^2-smooth curve, let s be its natural length parameter. Let $\kappa(s)$ denote its geodesic curvature. Fix a point in γ, let s_0 denote the corresponding length parameter value. The parameter

$$t_L := \int_{s_0}^{s} \kappa^{\frac{2}{3}}(\zeta) d\zeta \tag{1.3}$$

is called the *Lazutkin parameter.* See [15, formula (1.3)].

Theorem 1.15 *Let $\gamma \subset \Sigma$ be a germ of C^3-smooth curve with positive geodesic curvature κ and the string Poritsky property. Then, its Poritsky string length parameter t coincides with the Lazutkin parameter (1.3) up to additive and multiplicative constants. That is, up to constant factor, one has*

$$\frac{\mathrm{d}t}{\mathrm{d}s} = \kappa^{\frac{2}{3}}(s). \tag{1.4}$$

A proof of Theorem 1.15 will be presented in Sect. 6. It is based on the following theorem on asymptotics of the function $L(A, B)$ and its corollaries on string diffeomorphisms, also proved in the same section.

Theorem 1.16 *Let $\gamma \subset \Sigma$ be a C^3-smooth curve with positive geodesic curvature. For every $A \in \gamma$ let s_A denote the corresponding natural length parameter value. Let $L(A, B)$ denote the quantity defined in (1.2). One has*

$$L(A, B) = \frac{\kappa^2(A)}{12}|s_A - s_B|^3(1 + o(1)), \tag{1.5}$$

uniformly, as $s_A - s_B \to 0$ so that A and B remain in a compact subarc in γ. Asymptotic (1.5) is also uniform in the metric running through a closed bounded subset in the space of C^4-smooth Riemannian metrics.

Corollary 1.17 *Let $\gamma \subset \Sigma$ be a germ of C^3-smooth curve with positive geodesic curvature. For every small $p > 0$ let $\mathcal{T}_p : \gamma \to \gamma$ denote the corresponding string diffeomorphism (induced by reflection of geodesics tangent to γ from*

---

spherical and hyperbolic cases of Theorem 1.12, see [3]. Our proof of Theorem 1.12 given in Sect. 5 is geometric, analogous to Poritsky's arguments from [21, section 7] (which were given for Birkhoff planar billiards and string construction).

the string curve Γ_p *and acting on the tangency points). For every points* B *and* Q *lying in a compact subarc* $\widehat{\gamma} \Subset \gamma$*, one has*

$$\kappa^{\frac{2}{3}}(B)\lambda(B,\mathcal{T}_p(B)) \simeq \kappa^{\frac{2}{3}}(Q)\lambda(Q,\mathcal{T}_p(Q)), \text{ as } p \to 0, \tag{1.6}$$

uniformly in $B, Q \in \widehat{\gamma}$.

Corollary 1.18 *In the conditions of Corollary 1.17, one has*

$$\kappa^{\frac{2}{3}}(B)\lambda(B,\mathcal{T}_p(B)) \simeq \kappa^{\frac{2}{3}}(\mathcal{T}_p^m(B))\lambda(\mathcal{T}_p^m(B),\mathcal{T}_p^{m+1}(B)), \text{ as } p \to 0, \tag{1.7}$$

uniformly in $B \in \widehat{\gamma}$ *and those* $m \in \mathbb{N}$ *for which* $\mathcal{T}_p^m(B) \in \widehat{\gamma}$.

A symplectic generalization of Theorem 1.15 to families of the so-called weakly billiard-like maps of string type will be presented in Sect. 7.

1.4. Unique determination by 4-jet

The next theorem is a Riemannian generalization of the classical fact stating that each planar conic is uniquely determined by its 4-jet at some its point.

Theorem 1.19 *Let* Σ *be a surface equipped with a* C^6*-smooth Riemannian metric. A* C^5*-smooth germ of curve with the string Poritsky property is uniquely determined by its 4-jet.*

Theorem 1.19 will be proved in Sect. 8.

Remark 1.20 In the case, when Σ is the Euclidean plane, the statement of Theorem 1.19 follows from Poritsky's result [21, section 7] (see statement (iv) of Example 1.5). Similarly, in the case, when Σ is a surface of constant curvature, the statement of Theorem 1.19 follows from Theorem 1.7.

2. Background material from Riemannian geometry

We consider curves γ with positive geodesic curvature on an oriented surface Σ equipped with a Riemannian metric. In Sect. 2.1, we recall the notion of normal coordinates. We state and prove equivalence of different definitions of geodesic curvature. One of these definitions deals with geodesics tangent to γ at close points A and B and the asymptotics of angle between them at their intersection point C. In the same subsection, we prove existence of two geodesics tangent to γ through every point C close to γ and lying on the concave side from γ; the corresponding tangency points will be denoted by $A = A(C)$ and $B = B(C)$. We also prove an asymptotic formula for derivative of azimuth of a vector tangent to a geodesic (Proposition 2.7). In Sect. 2.2, we prove formulas for the derivatives $\frac{dA}{dC}$, $\frac{dB}{dC}$, which will be used in the proofs of Theorems 1.3, 1.7, 1.15. In Sect. 2.3, we consider a pair of geodesics issued from the same point A and their points $G(s)$, $Z(s)$ lying at a given distance s to A. We give an asymptotic formula for difference of azimuths of their tangent vectors at $G(s)$ and $Z(s)$, as $s \to 0$. We will use it in the proof of Theorem 1.19. In Sect. 2.4 we state and prove some asymptotic formulas relating sides and angles of small geodesic and curvilinear triangles, which will be used in the proofs of Theorems 1.15, 1.16 and 1.19.

2.1. Normal coordinates and equivalent definitions of geodesic curvature

Let Σ be a two-dimensional surface equipped with a C^3-smooth Riemannian metric g. Let $O \in \Sigma$. Let γ be a C^2-smooth germ of curve at O parametrized by its natural length parameter. Recall that its geodesic curvature $\kappa = \kappa(O)$ equals the norm of the covariant derivative $\nabla_{\dot{\gamma}}\dot{\gamma}$. In the Euclidean case, it coincides with the inverse of the osculating circle radius.

Consider the exponential chart $\exp : v \mapsto \exp(v)$ parametrizing a neighborhood of the point O by a neighborhood of zero in the tangent plane $T_O\Sigma$. We introduce orthogonal linear coordinates (x, y) on $T_O\Sigma$, which together with the exponential chart, induce *normal coordinates* centered at O, also denoted by (x, y), on a neighborhood of the point O. It is well known that in normal coordinates the metric has the same 1-jet at O, as the standard Euclidean metric (we then say that its *1-jet is trivial* at O.) Its Christoffel symbols vanish at O.

Remark 2.1 Let the surface Σ be C^{k+2}-smooth and the metric be C^{k+1}-smooth. Then, normal coordinates are C^k-smooth. This follows from theorem on dependence of solution of differential equation on initial condition (applied to the equation of geodesics) and C^k-smoothness of the Christoffel symbols. In normal coordinates the metric is C^{k-1}-smooth. Thus, each C^k-smooth curve is represented by a C^k-smooth curve in normal coordinates.

Proposition 2.2 *For every curve γ as above its geodesic curvature $\kappa(O)$ equals its Euclidean geodesic curvature $\kappa_e(O)$ in normal coordinates centered at O. If the normal coordinates (x, y) are chosen so that the x-axis is tangent to γ, then γ is the graph of a germ of function:*

$$\gamma = \{y = f(x)\}, \quad f(x) = \pm\frac{\kappa(O)}{2}x^2 + o(x^2), \ \textit{as } x \to 0. \tag{2.1}$$

Proof Proposition 2.2 follows from definition and vanishing of the Christoffel symbols at O in normal coordinates.

Proposition 2.3 *Let the germ $(\gamma, O) \subset \Sigma$ be the same as at the beginning of the subsection, and let γ have positive geodesic curvature. Let $\mathcal{U} \subset \Sigma$ be a small domain adjacent to γ from the concave side: γ is its concave boundary. Let $\widehat{\gamma} \subset \gamma$ be a compact subset: an arc with boundary. For every $C \in \mathcal{U}$ close enough to $\widehat{\gamma}$, there exist exactly two geodesics through C tangent to γ. In what follows, we denote their tangency points with γ by $A = A(C)$ and $B = B(C)$ so that AC is the right geodesic through C tangent to γ.*

Proof The statement of the proposition is obvious in the Euclidean case. The non-Euclidean case is reduced to the Euclidean case by considering a point $C \in \mathcal{U}$ close to $\widehat{\gamma}$ and normal coordinates (x_C, y_C) centered at C so that their family depends smoothly on C. In these coordinates, the curves $\gamma = \gamma_C$ depend smoothly on C and are strictly convex in the Euclidean sense, by Proposition 2.2. The geodesics through C are lines. This together with the statement of Proposition 2.3 in the Euclidean case implies its statement in the non-Euclidean case.

Let us consider that Σ is a Riemannian disk centered at O, the curve γ splits Σ into two open parts, and γ has positive geodesic curvature. For every point $A \in \gamma$ the geodesic tangent to γ at A will be denoted by G_A.

Proposition 2.4 *Taking the disk Σ small enough, one can achieve that for every $A \in \gamma$ the curve γ lies in the closure of one and the same component of the complement $\Sigma \setminus G_A$, $\gamma \cap G_A = \{A\}$.*

Proposition 2.4 follows its Euclidean version and Proposition 2.2.

Proposition 2.5 *For every two points $A, B \in \gamma$ close enough to O, the geodesics G_A and G_B intersect at a unique point $C = C_{AB} \in \mathcal{U}$ close to O.*

Proof Let H denote the geodesic through B orthogonal to $T_B\gamma$. It intersects the geodesic G_A at some point $P(A, B) \in \mathcal{U}$. The geodesic G_B separates $P(A, B)$ from the punctured curve $\gamma \setminus \{B\}$, by construction and Proposition 2.4. Therefore, G_B intersects the interval $(A, P(A, B))$ of the geodesic G_A. Uniqueness of intersection point of two geodesics in a domain with small diameter is classical. This proves the proposition.

Proposition 2.6 *For every $A, B \in \gamma$ close enough to O, let $C = C_{AB}$ denote the point of intersection $G_A \cap G_B$. Let $\alpha(A, B)$ denote the acute angle between the geodesics G_A and G_B at C, and let $\lambda(A, B)$ denote the length of the arc AB of the curve γ. The geodesic curvature $\kappa(O)$ of the curve γ at O can be found from any of the two following limits:*

$$\kappa(O) = \lim_{A,B\to O} \frac{\alpha(A, B)}{\lambda(A, B)}; \tag{2.2}$$

$$\kappa(O) = \lim_{A,B\to O} 2\frac{\mathrm{dist}(B, G_A)}{\lambda(A, B)^2}. \tag{2.3}$$

Proof In the Euclidean case, formulas (2.2) and (2.3) are classical. Their non-Euclidean versions follow by applying the Euclidean versions in normal coordinates centered, respectively, at C and A (or at the point in G_A closest to B), as in the proof of Proposition 2.3.

For every point $A \in \Sigma$ lying in a chart (x, y), e.g., a normal chart centered at O, and every tangent vector $v \in T_A\Sigma$ set

$$\mathrm{az}(v) := \text{ the azimuth of the vector } v: \text{ its Euclidean angle with the } x-\text{axis},$$

i.e., the angle in the Euclidean metric in the coordinates (x, y). The azimuth of an oriented one-dimensional subspace in $T_A\Sigma$ is defined analogously.

Proposition 2.7 *Let the metric on Σ be C^4-smooth. Let $A \in \Sigma$ be a point close to O and $\alpha(s)$ be a geodesic through A parametrized by the natural length parameter s, $\alpha(0) = A$.*

1) Let $\kappa_e(s)$ denote the Euclidean curvature of the geodesic α as a planar curve in normal chart (x, y) centered at O. For every $\varepsilon > 0$ small enough,

$$\kappa_e(s) = O(\mathrm{dist}(\alpha, O)), \text{ as } A \to O, \text{ uniformly on } \{|s| \leq \varepsilon\}, \tag{2.4}$$
$$\mathrm{dist}(\alpha, O) := \text{ the distance of the geodesic } \alpha \text{ to the point } O.$$

2) Set $v(s)=\dot{\alpha}(s)$. One has

$$\frac{\mathrm{d}\,\mathrm{az}(v(s))}{\mathrm{d}s}=O(\mathrm{dist}(\alpha,O))=O(\angle(v(0),AO)\,\mathrm{dist}(A,O))\ \text{as } A\to O, \quad (2.5)$$

uniformly on the set $\{|s|\leq\varepsilon\}$. The latter angle in (2.5) is the Riemannian angle between the vector $v(0)$ and the Euclidean line AO.

Proof In the coordinates (x,y), the geodesics are solutions of the second order ordinary differential equation saying that $\ddot{\alpha}$ equals a quadratic form in $\dot{\alpha}$ with coefficients equal to appropriate Christoffel symbols of the metric g (which vanish at O), and $|\dot{\alpha}|=1$ in the metric g. The derivative in (2.5) is expressed in terms of the Christoffel symbols. This derivative taken along a geodesic α through O vanishes identically on α, since each geodesic through O is a straight line in normal coordinates. Therefore, if we move the geodesic through O out of O by a small distance δ, then the derivative in (2.5) will change by an amount of order δ: the Christoffel symbols are C^1-smooth, since the metric is C^4-smooth (hence, C^2-smooth in normal coordinates). This implies the first equality in (2.5). The second equality follows from the fact that the geodesics through A issued in the direction of the vectors $\overrightarrow{AO}$ and $v(0)$ are, respectively, the line AO and α, hence, $\mathrm{dist}(\alpha,O)=O(\angle(v(0),AO)\,\mathrm{dist}(A,O))$. This proves (2.5).

Let s_e denote the Euclidean natural parameter of the curve α, with respect to the standard Euclidean metric in the chart (x,y). Recall that $\kappa_e(s)=\frac{d\,\mathrm{az}(v(s))}{ds_e}$. For $\varepsilon>0$ small enough and A close enough to O, the ratio $\frac{ds_e}{ds}$ is uniformly bounded on $\{|s|\leq\varepsilon\}$. This together with (2.5) implies (2.4). The proposition is proved.

2.2. Angular derivative of exponential mapping and the derivatives $\frac{\mathrm{d}A}{\mathrm{d}C}$, $\frac{\mathrm{d}B}{\mathrm{d}C}$

In the proof of main results, we will use an explicit formula for the derivatives of the functions $A(C)$ and $B(C)$ from Proposition 2.3. To state it, let us introduce the following auxiliary functions. For every $x\in\Sigma$ set,

$$\psi(x,r):=\frac{1}{2\pi}\ (\text{the length of circle of radius } r \text{ centered at } x).$$

Consider the polar coordinates (r,ϕ) on the Euclidean plane $T_x\Sigma$. For every unit vector $v\in T_x\Sigma$, $|v|=1$ (identified with the corresponding angle coordinate ϕ) and every $r>0$ let $\Psi(x,v,r)$ denote $\frac{1}{r}$ times the module of derivative in ϕ of the exponential mapping at the point rv:

$$\Psi(x,v,r):=r^{-1}\left|\frac{\partial\exp}{\partial\phi}(rv)\right|. \quad (2.6)$$

Proposition 2.8 (see [7] in the hyperbolic case). *Let Σ be a complete simply connected Riemannian surface of constant curvature. Then,*

$$r\Psi(x,v,r)=\psi(x,r)=\psi(r)=\begin{cases} r, & \text{if } \Sigma \text{ is Euclidean plane,}\\ \sin r, & \text{if } \Sigma \text{ is unit sphere,}\\ \sinh r, & \text{if } \Sigma \text{ is hyperbolic plane.}\end{cases} \quad (2.7)$$

Proof The left equality in (2.7) and independence of x and v follow from homogeneity. Let us prove the right equality: formula for the function $\psi(r)$. In the planar case this formula is obvious.

(a) Spherical case. Without loss of generality, let us place the center $x = O$ of the circle under question to the north pole $(0, 0, 1)$ in the Euclidean coordinates (x_1, x_2, x_3) on the ambient space. Since each geodesic is a big circle of length 2π and due to symmetry, without loss of generality we consider that $0 < r \leq \frac{\pi}{2}$. Then, the disk in Σ centered at O of radius r is 1-to-1 projected to the disk of radius $\sin r$ in the coordinate (x_1, x_2)-plane. The length of its boundary equals the Euclidean length $2\pi \sin r$ of its projection.

(b) Case of hyperbolic plane. We consider the hyperbolic plane in the model of unit disk equipped with the metric $\frac{2|dz|}{1-|z|^2}$ in the complex coordinate z. For every $R > 0$, $R < 1$, the Euclidean circle $\{|z| = R\}$ of radius R is a hyperbolic circle of radius

$$r = \int_0^R \frac{2ds}{1 - s^2} = \log \left| \frac{1+R}{1-R} \right|.$$

The hyperbolic length of the same circle equals $L = \frac{4\pi R}{1-R^2}$. Substituting the former formula to the latter, one yields

$$R = \frac{e^r - 1}{e^r + 1}, \ L = 2\pi \sinh r,$$

and finishes the proof of the proposition.

Proposition 2.9 *Let $\gamma \subset \Sigma$ be a germ of C^2-smooth curve. Let s be the length parameter on γ orienting it positively as a boundary of a convex domain. Let $\mathcal{U} \subset \Sigma$ be a small domain adjacent to γ from the concave side, see Proposition 2.3. For every $C \in \mathcal{U}$, let $A(C)$, $B(C) \in \gamma$ be the corresponding points from Proposition 2.3, and let $s_A = s_A(C)$, $s_B = s_B(C)$ denote the corresponding length parameter values as functions of C. Set*

$$L_A := |CA(C)|, \ L_B := |CB(C)|.$$

For every $Q = A, B$ let $w_Q \in T_Q\gamma$ be the unit tangent vector of the geodesic CQ directed to C. Let $\zeta_Q \in T_C\Sigma$ denote the unit tangent vector of the same geodesic at C directed to Q. For every $v \in T_C\Sigma$ and $Q = A, B$, one has

$$\frac{\mathrm{d}s_Q}{dv} = \frac{v \times \zeta_Q}{\kappa(Q) L_Q \Psi(Q, w_Q, L_Q)}; \ v \times \zeta_Q := |v| \sin \angle(v, \zeta_Q), \tag{2.8}$$

where $\angle(v, \zeta_Q)$ is the oriented angle between the vectors v and ζ_Q: it is positive, if the latter vectors form an orienting basis of the space $T_C\Sigma$.

Proof Let us prove (2.8) for $Q = A$; the proof for B is analogous. As $A = A(C)$ moves by ε along the curve γ to the point A_ε with the natural parameter $s_A + \varepsilon$, the geodesic G_A tangent to γ at A is deformed to the geodesic G_{A_ε} intersecting G_A at a point converging to A, as $\varepsilon \to 0$. Let $\alpha(\varepsilon)$ denote their acute intersection angle at the latter point. One has

$$\alpha(\varepsilon) \simeq \kappa(A)\varepsilon. \tag{2.9}$$

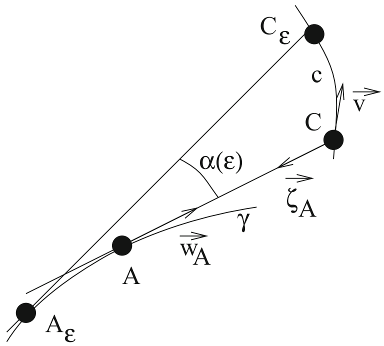


FIGURE 3. The tangent geodesics to γ at the points A and A_ε. The angle between them is $\alpha(\varepsilon) \simeq \kappa(A)\varepsilon$. One has $\mathrm{dist}(C, A_\varepsilon C_\varepsilon) \simeq \alpha(\varepsilon) L_A \Psi(A, w_A, L_A)$, $L_A = |AC|$, and $\lambda(C, C_\varepsilon) \simeq \mathrm{dist}(C, A_\varepsilon C_\varepsilon)/\sin\angle(v, \zeta_A)$

Both above statements follow from (2.2) and definition. One also has

$$\mathrm{dist}(C, G_{A_\varepsilon}) \simeq \alpha(\varepsilon) L_A \Psi(A, w_A, L_A) \simeq \varepsilon\kappa(A) L_A \Psi(A, w_A, L_A), \qquad (2.10)$$

by the definition of the function Ψ and (2.9).

Without loss of generality, we consider that v is a unit vector. Let us draw a curve c through C tangent to v and oriented by v. Let τ denote its natural parameter defined by this orientation. Let C_ε denote the point of intersection of the geodesic G_{A_ε} with c, see Fig. 3. Consider $\tau = \tau(C_\varepsilon)$ as a function of ε: $\tau = \tau(\varepsilon)$. One has

$$\frac{\mathrm{d}s_A}{\mathrm{d}v} = \left(\frac{\mathrm{d}\tau}{\mathrm{d}\varepsilon}(0)\right)^{-1}, \ \tau(C_\varepsilon) - \tau(C) \simeq \frac{\mathrm{dist}(C, G_{A_\varepsilon})}{\sin\angle(v, \zeta_A)} \simeq \varepsilon\frac{\mathrm{d}\tau}{\mathrm{d}\varepsilon}(0) = \varepsilon\left(\frac{\mathrm{d}s_A}{\mathrm{d}v}\right)^{-1},$$

as $\varepsilon \to 0$, by definition. Substituting (2.10) to this formula yields (2.8).

2.3. Geodesics passing through the same base point: azimuths of tangent vectors at equidistant points

Proposition 2.10 *Let the metric on Σ be C^3-smooth. Let $G_t(s), Z_t(s) \subset \Sigma$ be two families of geodesics parametrized by the natural length s and depending on a parameter $t \in [0, 1]$. Let they be issued from the same point $A_t = G_t(0) = Z_t(0)$. Let A_t lie in a given compact subset (the same for all t) in a local chart (x, y) (not necessarily a normal chart). Set*

$$\phi_t = \mathrm{az}(\dot{G}_t(0)) - \mathrm{az}(\dot{Z}_t(0)).$$

One has

$$\mathrm{az}(\dot{G}_t(s)) - \mathrm{az}(\dot{Z}_t(s)) \simeq \phi_t, \ \textit{as } s \to 0, \ \textit{uniformly in } t \in [0, 1]. \qquad (2.11)$$

Proof A geodesic, say, $G(s)$ is a solution of a second order vector differential equation with a given initial condition: a point $A \in \Sigma$ and the azimuth $\mathrm{az}(v(0))$ of a unit vector $v(0) \in T_A\Sigma$. Here, we set $v(s) = \dot{G}(s)$. It depends smoothly on the initial condition. The derivative of the vector function

$(G(s), \operatorname{az} v(s))$ in the initial conditions is a linear operator (3×3-matrix) function in s that is a solution of the corresponding linear equation in variations. The right-hand sides of the equation for geodesics and the corresponding equation in variations are, respectively, C^2- and C^1-smooth. Let us now fix the initial point A and consider the derivative of the azimuth $\operatorname{az}(v(s))$ in the initial azimuth $\operatorname{az}(v(0))$ for fixed s. If $s = 0$, then the latter derivative equals 1, since the initial condition in the equation in variations is the identity matrix. Therefore, in the general case the derivative of the azimuth $\operatorname{az}(v(s))$ in $\operatorname{az}(v(0))$ equals $1 + u_{A,v(0)}(s)$, where $u_{A,v(0)}(s)$ is a C^1-smooth function with $u_{A,v(0)}(0) = 0$. This together with the above discussion and Lagrange Increment Theorem for the derivative in $\operatorname{az}(v(0))$ implies (2.11).

2.4. Geodesic-curvilinear triangles in normal coordinates

Everywhere below in the present subsection Σ is a two-dimensional surface equipped with a C^4-smooth Riemannian metric g, and $O \in \Sigma$.

Proposition 2.11 *Let $A_uB_uC_u$ be a family of geodesic right triangles lying in a compact subset in Σ with right angle B_u. Set*

$$c = c_u = |A_uB_u|,\ b = b_u = |A_uC_u|,\ a = a_u = |B_uC_u|,\ \alpha = \alpha_u = \angle B_uA_uC_u.$$

Let $b_u, \alpha_u \to 0$, as $u \to u_0$. Then,

$$b \simeq c,\ b - c \simeq \frac{a^2}{2c} \simeq \frac{1}{2}c\alpha^2 \simeq \frac{1}{2}a\alpha,\ \angle B_uC_uA_u = \frac{\pi}{2} - \alpha + o(\alpha). \qquad (2.12)$$

Proof Consider normal coordinates (x_u, y_u) centered at A_u (depending smoothly on the base point A_u). The coordinates

$$(X_u, Y_u) := \left(\frac{x_u}{c_u}, \frac{y_u}{c_u}\right)$$

are normal coordinates centered at A_u for the Riemannian metric rescaled by division by c_u. For the rescaled metric, one has $|A_uB_u| = 1$. In the rescaled normal coordinates (X_u, Y_u), the rescaled metric has trivial 1-jet at 0 and tends to the Euclidean metric, as $u \to u_0$: its nonlinear part tends to zero, as $u \to u_0$, uniformly on the Euclidean disk of radius 2 in the coordinates (X_u, Y_u). One has obviously $|A_uB_u| \simeq |A_uC_u|$ in the rescaled metric, since $\alpha_u \to 0$. Rescaling back, we get the first asymptotic formula in (2.12).

Let S_u denote the circle of radius $|A_uB_u|$ centered at A_u, and let D_u denote its point lying on the geodesic A_uC_u: $|A_uB_u| = |A_uD_u|$; the arc B_uD_u of the circle S_u is its intersection with the geodesic angle $B_uA_uC_u$. In the rescaled coordinates (X_u, Y_u), the circle S_u tends to the Euclidean unit circle. Thus, its geodesic curvature in the rescaled metric tends to 1. The geodesic segment B_uC_u is tangent to S_u at the point B_u, and $\angle B_uC_uA_u \to \frac{\pi}{2}$. The two latter statements together with Proposition 2.2 (applied to $O = B$ and $\gamma = S_u$) imply that in the rescaled metric, one has $|B_uC_u| \simeq \alpha$,

$$|D_uC_u| = |A_uC_u| - |A_uB_u| \simeq \frac{|B_uC_u|^2}{2} \simeq \frac{1}{2}\alpha^2 \simeq \frac{1}{2}|B_uC_u|\alpha.$$

Rescaling back to the initial metric, we get the second, third and fourth formulas in (2.12). The fifth formula follows from Gauss–Bonnet Formula, which implies that the sum of angles in the triangle $A_uB_uC_u$ differs from π by a quantity $O(Area) = O(|A_uB_u||B_uC_u|) = O(\alpha|A_uB_u|^2) = o(\alpha)$.

Proposition 2.12 *Consider a family of C^3-smooth arcs $\gamma_u = A_uB_u$ of curves in Σ (lying in a compact set) with uniformly bounded geodesic curvature (from above) such that $|A_uB_u| \to 0$, as $u \to 0$. Let $\lambda(A_u, B_u)$ denote their lengths. Let α_u denote the angle at A_u between the arc γ_u and the geodesic segment A_uB_u. One has*

$$\lambda(A_u, B_u) = |A_uB_u| + O(|A_uB_u|^3),\ \alpha_u = O(|A_uB_u|). \tag{2.13}$$

Proof The proposition obviously holds in Euclidean metric. It remains valid in the normal coordinates centered at A_u. Indeed, the length of the arc γ_u in the Euclidean metric in the normal chart differs from its Riemannian length by a quantity $O(|A_uB_u|^3)$, since the difference of the metrics at a point $P \in \gamma_u$ is $O(|PA_u|^2) = O(|A_uB_u|^2)$ and the curvature of the arcs γ_u is bounded.

Proposition 2.13 *Consider a family of curvilinear triangles $T_u := A_uB_uC_u$ in Σ where the side A_uB_u is geodesic and the sides A_uC_u, B_uC_u are arcs of C^3-smooth curves with uniformly bounded geodesic curvature. Let the side A_uC_u be tangent to the side A_uB_u at A_u. Set*

$$\varepsilon := |A_uB_u|,\ \theta := \frac{\pi}{2} - \angle A_uB_uC_u.$$

Here, $\angle A_uB_uC_u$ is the angle at B_u between the (curvilinear) sides A_uB_u, B_uC_u of the triangle T_u. Let the triangles T_u lie in a compact subset in Σ, and $\varepsilon, \theta \to 0$, as $u \to 0$. Then,

$$\lambda(A_u, C_u) - |A_uB_u| = O(\varepsilon^3) + O(\varepsilon^2\theta). \tag{2.14}$$

Proof One has $|B_uC_u| = O(\varepsilon^2)$, by construction and since $\theta \to 0$. Hence, $|A_uC_u| \simeq \varepsilon$. Let D_u denote the point closest to C_u in the geodesic A_uB_u: the points A_u, C_u, D_u form a triangle Δ_u with right angle at D_u. One has

$$|C_uD_u| = O(\varepsilon^2),\ \lambda(A_u, C_u) - |A_uC_u| = O(\varepsilon^3), \tag{2.15}$$

by definition and (2.13),

$$|A_uC_u| - |A_uD_u| = O(\frac{|C_uD_u|^2}{|A_uC_u|}) = O(\varepsilon^3), \tag{2.16}$$

by (2.15) and (2.12) applied to Δ_u. Let us show that

$$|A_uD_u| - |A_uB_u| = \pm|B_uD_u| = O(\varepsilon^2\theta) + O(\varepsilon^3). \tag{2.17}$$

In the right triangle $\widehat{\Delta}_u$ with vertices B_u, C_u, and D_u, one has $\angle D_uB_uC_u = \frac{\pi}{2} \pm \theta + O(\varepsilon^2)$. Indeed, the latter angle is the sum (difference) of the two following angles at B_u: the angle $\frac{\pi}{2} - \theta$ of the triangle T_u if the sign in (2.17) is "$-$" (or its complementary angle $\pi+\theta$, if the sign is "$+$") ; the angle between the geodesic B_uC_u and the curved side B_uC_u in T_u, which is $O(|B_uC_u|) = O(\varepsilon^2)$. This implies the above formula for the angle $\angle D_uB_uC_u$, which in its turn implies that in the triangle $\widehat{\Delta}_u$ one has $\angle B_uC_uD_u = O(\theta) + O(\varepsilon^2)$ (the last formula in (2.12)). The latter formula together with (2.15) and (2.12) imply (2.17). Adding formulas (2.15), (2.16), and (2.17) yields (2.14).

3. The string foliation: proof of Theorem 1.3

3.1. Finite smoothness lemmas

Everywhere below in the present section, we are dealing with a function $f(x, y)$ of two variables (x, y): the variable y is scalar, and the variable x may be a vector variable. The function f is defined on the product

$$Z = \overline{U} \times V$$

of closure of a domain $U \subset \mathbb{R}^n_x$ and an interval $V = (-\varepsilon, \varepsilon) \subset \mathbb{R}_y$.

The following two lemmas will be used in the proof of Theorem 1.3.

Lemma 3.1 *Let a function f as above be C^k-smooth on Z, $k \geq 2$, and let*

$$f(x, y) = a(x)y^2(1 + o(1)), \text{ as } y \to 0, \quad \text{uniformly in } x \in \overline{U}; \quad a > 0. \tag{3.1}$$

Then, the function $g(x, y) := \operatorname{sign}(y)\sqrt{f(x, y)}$ is C^{k-1}-smooth on Z.

Lemma 3.2 *Let a function $f(x, y)$ as at the beginning of the section be C^k-smooth on Z and even in y: $f(x, y) = f(x, -y)$. Then, $g(x, z) := f(x, \sqrt{z})$ is $C^{[\frac{k}{2}]}$-smooth on $\widetilde{Z} := \overline{U} \times [0, \varepsilon^2)$, and its restriction to $\widetilde{Z} \setminus \{z = 0\}$ is C^k-smooth.*

In the proof of the lemmas for simplicity without loss of generality, we consider that the variable x is one-dimensional; in higher-dimensional case the proof is the same. We use the following definition and a more precise version of the asymptotic Taylor formula for finitely smooth functions.

Definition 3.3 Let $l, m \in \mathbb{Z}_{\geq 0}$. We say that

$$f(x, y) = o_l(y^m), \text{ as } y \to 0,$$

if for every $j, s \in \mathbb{Z}_{\geq 0}$, $j \leq l$, $s \leq m$ the derivative $\frac{\partial^{j+s} f}{\partial^j x \partial^s y}$ exists and is continuous on $Z = \overline{U} \times V$ and one has

$$\frac{\partial^{j+s} f}{\partial^j x \partial^s y}(x, y) = o(y^{m-s}), \text{ as } y \to 0, \text{ uniformly in } x \in \overline{U}. \tag{3.2}$$

Proposition 3.4 *Let $f(x, y)$ be as at the beginning of the section, and let f be C^k-smooth on Z. Then, for every $l, m \in \mathbb{Z}_{\geq 0}$ with $l + m \leq k$, one has*

$$f(x, y) = f(x, 0) + \sum_{j=1}^{m} a_j(x)y^j + R_m(x, y), \ a_j(x) = \frac{1}{j!}\frac{\partial^j f}{\partial y^j}(x, 0) \in C^l(\overline{U}),$$

$$R_m(x, y) = o_l(y^m), \text{ as } y \to 0, \text{ uniformly in } x \in \overline{U}. \tag{3.3}$$

Proof The first formula in (3.3) holds with

$$R_m(x, y) = \int_{0 \leq y_m \leq \dots \leq y_1 \leq y} \left(\frac{\partial^m}{\partial y^m} f(x, y_m) - \frac{\partial^m}{\partial y^m} f(x, 0)\right) \mathrm{d}y_m \mathrm{d}y_{m-1} \dots dy_1,$$

by the classical asymptotic Taylor formula with error term in integral form. The latter R_m is $o_l(y^m)$, whenever $f \in C^k$ and $k \geq l + m$.

Proposition 3.5 *One has*

$$y^{-s} o_l(y^m) = o_l(y^{m-s}) \text{ for every } m, s \in \mathbb{Z}_{\geq 0}, \ m \geq s. \tag{3.4}$$

The proposition follows from definition.

Proof of Lemma 3.1 The function $g(x,y) = \mathrm{sign}(y)\sqrt{f(x,y)}$ is well defined, by (3.1). It is obviously C^k-smooth outside the hyperplane $\{y = 0\}$. Fix arbitrary $l, m \in \mathbb{Z}_{\geq 0}$ such that $l + m \leq k - 1$. Let us prove continuity of the derivative $\frac{\partial^{l+m} g}{\partial x^l \partial y^m}$ on Z.

Case $m = 0$; then $k \geq l+1$. The above derivative is a linear combination of expressions

$$\mathrm{sign}(y) f^{\frac{1}{2}-s}(x,y) \prod_{j=1}^{s} \frac{\partial^{n_j} f(x,y)}{\partial x^{n_j}}, \ s \in \mathbb{N}, \ n_j \geq 1, \ \sum_{j=1}^{s} n_j = l. \tag{3.5}$$

The partial derivatives in (3.5) are C^1-smooth, since f is C^k-smooth and $n_j \leq l \leq k - 1$. One has

$$\mathrm{sign}(y) f^{\frac{1}{2}-s}(x,y) \simeq a^{\frac{1}{2}-s}(x) y^{1-2s}, \tag{3.6}$$

by definition. If $s = 1$, then $y^{1-2s} = y^{-1}$, and the expression (3.5) contains only one derivative $\phi(x,y) := \frac{\partial^{n_1} f(x,y)}{\partial x^{n_1}}$. One has $\phi \in C^1(Z)$, $\phi|_{y=0} = 0$, $\frac{\partial \phi}{\partial y}|_{y=0} = 0$, by smoothness and since $f(x,y) = O(y^2)$. Hence $\phi(x,y) = o(y)$ uniformly, as $y \to 0$. Therefore, the expression (3.5) is continuous on $Z \setminus \{y = 0\}$, and it extends continuously to Z as zero along the line $\{y = 0\}$. If $s \geq 2$, then $n_j \leq l - 1 \leq k - 2$. Hence, each derivative in (3.5) is C^2-smooth, has vanishing first derivative in y at $y = 0$ and is asymptotic to y^2 times a continuous function in x. Then, (3.5) is again continuous, by (3.6).

Case $m = 1$ is treated analogously with the following change: one of the derivatives in (3.5) will contain one differentiation in y and will be asymptotic to y times a continuous function in x.

Case $m \geq 2$. Then, $k \geq m + l + 1 \geq l + 3$. One has

$$g(x,y) = a^{\frac{1}{2}}(x) y \sqrt{w(x,y)}, \tag{3.7}$$

$$w(x,y) = 1 + \sum_{j=3}^{m+1} a^{-1}(x) a_j(x) y^{j-2} + \frac{o_l(y^{m+1})}{y^2}, \quad a, a_j \in C^l(\overline{U}), \tag{3.8}$$

by (3.3) applied to the function $f(x,y)$ and m replaced by $m+1$. The derivative $\frac{\partial^{l+m-1} g}{\partial x^l \partial y^{m-1}}$ exists and continuous for small y, by (3.8) and since $\frac{o_l(y^{m+1})}{y^2} = o_l(y^{m-1})$, see (3.4). Now, it remains to prove the same statement for the derivative $h := \frac{\partial^{l+m} g}{\partial x^l \partial y^m}$. Those terms in its expression that include the derivatives of the function $\frac{o_l(y^{m+1})}{y^2} = o_l(y^{m-1})$ with differentiation in y of orders less than m are well defined and continuous, as above. Each term in h that contains a derivative $\frac{\partial^{j+m}}{\partial x^j \partial y^m}(\frac{o_l(y^{m+1})}{y^2})$ contains only one such derivative, and it comes with the factor y from (3.7); here $j \leq l$. On the other hand, the latter derivative is $\frac{o_{l-j}(y)}{y^2} = o(\frac{1}{y})$, by (3.4). Thus, its product with the above factor y is a continuous function, as are the other factors in the term under question. Continuity of the derivative h is proved. Lemma 3.1 is proved.

Proof of Lemma 3.2 Fix $l, m \in \mathbb{Z}_{\geq 0}$ such that $l + m \leq [\frac{k}{2}]$. Then $l + 2m \leq k$, and one has

$$f(x,y) = \sum_{j=0}^{m} a_j(x) y^{2j} + o_l(y^{2m}),$$

where the functions $a_j(x)$ are C^l-smooth, by (3.3) and evenness. Set $z = y^2$. The derivative $\frac{\partial^{l+m}}{\partial x^l \partial z^m}$ of the above sum is obviously continuous, since the sum is a polynomial in z with coefficients being C^l-smooth functions in x. Let us prove continuity of the derivative of the remainder $o_l(y^{2m})$. One has

$$\frac{\partial}{\partial z} = \frac{1}{2y} \frac{\partial}{\partial y}.$$

Therefore, the above $(l+m)$-th partial derivative of the remainder $o_l(y^{2\,m})$ is $o(1)$, see (3.4). This proves continuity and Lemma 3.2.

3.2. Proof of Theorem 1.3

The fact that the exterior bisector line field Λ is tangent to the string construction curves is well known and proved as follows. Consider the value $L(A,B) = |AC| + |CB| - \lambda(A,B)$ as a function of C: here $A = A(C)$ and $B = B(C)$ are the same, as in Proposition 2.3. Its derivative along the string construction curve Γ_p through C should be zero. Let $v \in T_C\Sigma$ be a unit vector. Let α and β be, respectively, the oriented angles between the vector v and the vectors ζ_A and ζ_B in $T_C\Sigma$ directing the geodesics G_A, G_B from C to A and B, respectively, see Proposition 2.9. The derivative of the above function $L(A(C), B(C))$ along the vector v is equal to $-(\cos\alpha + \cos\beta)$. Therefore, it vanishes if and only if the line generated by v is the exterior bisector $\Lambda(C)$ of the angle $\angle ACB$. Therefore, the level sets of the function $L(A(C), B(C))$, i.e., the string construction curves are integral curves of the line field Λ.

It suffices to prove only statement (1) of Theorem 1.3: C^{k-1}-smoothness on $\mathcal{U}$ and $C^{r(k)}$-smoothness on $\overline{\mathcal{U}}$ of the line field Λ. Statement (2) on $C^{r(k)+1}$-regularity of its integral curves (the string construction curves) and continuity then follows from the next general fact: *for every C^r-smooth line field the $(r+1)$-jets of its integral curves at base points A are expressed analytically in terms of r-jets of the line field, and hence, depend continuously on A.*

Fix a C^k-smooth coordinate system (s,z) on Σ centered at the base point O of the curve γ such that γ is the s-axis, $s|_\gamma$ is the natural length parameter of the curve γ and $\mathcal{U} = \{z > 0\}$. For every $\sigma \in \mathbb{R}$ small enough, let $G(\sigma)$ denote the geodesic tangent to γ at the point with length parameter value σ. For every $\sigma, s \in \mathbb{R}$ small enough, let $A(\sigma, s)$ denote the point of intersection of the geodesic $G(\sigma)$ with the line parallel to the z-axis and having abscissa s. The mapping $(\sigma, s) \mapsto A(\sigma, s)$ is C^{k-1}-smooth, since so is the family of geodesics $G(\sigma)$ (by C^k-smoothness of the metric (in some initial C^{k+1}-smooth chart) and C^k-smoothness of the curve γ) and by transversality. Set

$$z(\sigma, s) := z(A(\sigma, s)).$$

Proposition 3.6 *The function*

$$y(\sigma, s) := \operatorname{sign}(\sigma - s)\sqrt{z(\sigma, s)}$$

is C^{k-2}-smooth on a neighborhood of zero in $\mathbb{R}^2$ and C^{k-1}-smooth outside the diagonal $\{\sigma = s\}$. The mapping

$$F : (\sigma, s) \mapsto (s, y(\sigma, s)) \tag{3.9}$$

is a C^{k-2}-smooth diffeomorphism of a neighborhood of the origin onto a neighborhood of the origin, and it is C^{k-1}-smooth outside the diagonal. It sends the diagonal to the axis $\{y = 0\}$.

Proof For every point $Q \in \Sigma$ lying in a smooth chart (s, z), let $u(Q)$ denote the orthogonal projection of the vector $\frac{\partial}{\partial z} \in T_Q\Sigma$ to the line $(\mathbb{R}\frac{\partial}{\partial s})^{\perp}$. Set $\mu(Q) := ||u(Q)||^{-1}$. Recall that $\kappa(s) > 0$. One has $\mu > 0$ near O,

$$z(\sigma, s) = \frac{1}{2}\mu(s, 0)\kappa(s)(s - \sigma)^2 + o((s - \sigma)^2), \text{ as } \sigma \to s, \tag{3.10}$$

uniformly in small s, by (2.3). This together with C^{k-1}-smoothness of the function z and Lemma 3.1 implies the statements of the proposition.

Let us now return to the proof of statement (1) of Theorem 1.3. Consider the mapping inverse to the mapping F from (3.9):

$$F^{-1} : (s, y) \mapsto (\sigma, s).$$

The function $\sigma = \sigma(s, y)$ is C^{k-2}-smooth, by Proposition 3.6, and it is C^{k-1}-smooth outside the axis $\{y = 0\}$. Recall that the geodesic $G(\sigma(s, y))$ passes through the point $A = (s, z) = (s, y^2) \in \mathcal{U}$. For every s and y, let $v = v(s, y) \in T_A\Sigma$ denote the unit tangent vector of the geodesic $G(\sigma(s, y))$ that orients it in the same way, as the orienting tangent vector of the curve γ at $\sigma(s, y)$. The vector function $v(s, y)$ is C^{k-2}-smooth in (s, y). For a given point $A = (s, z)$, set $y := \sqrt{z}$, the unit vectors $v(s, y), v(s, -y) \in T_A\Sigma$ direct the two geodesics through A that are tangent to γ, by construction. Their sum $w(s, y) = v(s, y) + v(s, -y)$ generates the line $\Lambda(A)$ of the line field Λ, by definition. The vector function $w(s, y)$ is even in y, C^{k-2}-smooth in both variables, and $|w| = 2|v| = 2$, whenever $y = 0$. Thus, w is $C^{[\frac{k}{2}]-1}$-smooth in (s, z) and C^{k-1}-smooth outside the curve $\gamma = \{z = 0\}$, by Proposition 3.6 and Lemma 3.2. Finally, w induces a vector field generating Λ that is $C^{[\frac{k}{2}]-1}$-smooth on $\overline{\mathcal{U}}$ and C^{k-1}-smooth on $\mathcal{U}$. Theorem 1.3 is proved.

4. Billiards on surfaces of constant curvature: proofs of Proposition 1.6 and Theorem 1.7

In Sect. 4.1, we prove Proposition 1.6. The proof of Theorem 1.7, which follows its proof given in [21, section 7] in the Euclidean case, takes the rest of the section. In Sect. 4.2, we prove the following coboundary property of a curve γ with the string Poritsky property: for every $A, B \in \gamma$, set $C = C_{AB}$, the ratio $|AC|/|BC|$ equals the ratio of values at A and B of some function on γ. In Sect. 4.3, we deduce Theorem 1.7 from the coboundary property by planimetric arguments using Ceva's Theorem.

4.1. Proof of Proposition 1.6

We re-state and prove Proposition 1.6 in a more general Riemannian context. To do this, let us recall the following definition.

Definition 4.1 [1, p. 345] (implicitly considered in [21]) Let Σ be a surface equipped with a Riemannian metric, $\gamma \subset \Sigma$ be a (germ of) curve with positive geodesic curvature. Let Γ_p denote the family of curves obtained from it by string construction. We say that γ has *evolution* (or *Graves*) *property,* if for every $p_1 < p_2$ the curve Γ_{p_1} is a caustic for the curve Γ_{p_2}.

Example 4.2 It is well known that each conic on a surface Σ of constant curvature has evolution property, and the corresponding curves Γ_p given by string construction are confocal conics. In the Euclidean case, this follows from the classical fact saying that the caustics of a billiard in a conic are confocal conics (Proclus–Poncelet Theorem). Analogous statements hold in non-zero constant curvature and in higher dimensions, see [29, theorem 3].

Proposition 4.3 *Let Σ be a surface equipped with a C^4-smooth Riemannian metric. Let $\gamma \subset \Sigma$ be a C^4-smooth germ of curve with positive geodesic curvature that has evolution property. Then, it has the string Poritsky property*[2]*. For every $p, q > 0$, the reflections from the string curves Γ_p and Γ_q commute as mappings acting on the space of those oriented geodesics that intersect both of them and lie on the concave side $\mathcal{U}$ from the curve γ.*

Remark 4.4 In the Euclidean case, the first part of Proposition 4.3 with a proof is contained in [1,21]. Commutativity then follows by arguments from [24, chapter 3]. The proof of the first part of Proposition 4.3 given below is analogous to arguments from [21], [24, ch.3]. The analogue of evolution property for outer billiards was introduced and studied by E. Amiran [2].

Proof of Proposition 4.3 Billiard reflections acting on the manifold of oriented geodesics preserve a canonical symplectic form ω. See [21, section 3], [24, chapter 3] in the planar case. In the general case the form ω is given by Melrose construction, see [23, section 1.5], [4,5,18,19] and Sect. 7.1 below. The string curves Γ_p form a foliation of $\overline{\mathcal{U}}$ by level curves of a function ϕ that is C^3-smooth on $\mathcal{U}$, C^1-smooth on $\overline{\mathcal{U}}$ and has no critical points. This follows from the fact that they are tangent to the line field Λ of the same regularity (Theorem 1.3). Consider the mapping R of the set $\overline{\mathcal{U}}$ to the space of oriented geodesics sending each point $Q \in \overline{\mathcal{U}}$ to the geodesic tangent to $\Lambda(Q)$. The orientations of the lines $\Lambda(Q)$ are chosen to converge to the orientation of the curve γ, as $Q \to \gamma$. This is a diffeomorphism onto $\overline{\mathcal{U}^*} := R(\overline{\mathcal{U}})$ of the same regularity, as Λ, by construction and since γ has positive geodesic curvature. The image $\Gamma_p^* := R(\Gamma_p)$ of each curve Γ_p is the family or geodesics tangent to

[2]Recently, after an arxiv draft of the present paper was written, it was shown in a joint paper of the author with Sergei Tabachnikov and Ivan Izmestiev [10] that for a C^∞-smooth curve γ the *evolution property is* **equivalent** *to the Poritsky property*. In addition, it is also equivalent to the statement that the foliation by the curves Γ_p and its orthogonal foliation form a *Liouville net* on the concave side $\mathcal{U}$ from the curve γ. See [10] for a survey of related results.

Γ_p and oriented as Γ_p. The curves Γ_p^* form a foliation by level curves of the function $\psi := \phi \circ R^{-1}$, which has the above regularity and no critical points. For every $q < p$, the curve Γ_q^* is invariant under the reflection T_p from the curve Γ_p (evolution property). Therefore, the restriction of the function ψ to the strip between the curves Γ_0^* and Γ_p^* is also T_p-invariant. Hence, its Hamiltonian vector field H_ψ is also invariant and tangent to the curves Γ_q^*. Thus, for every $q < p$, the reflection $T_p : \Gamma_q^* \to \Gamma_q^*$ acts by translation in the time coordinate t_q of the field H_ψ on Γ_q^*, and this also holds for $q = 0$. The time coordinate t_0 on Γ_0^* induces a parameter, also denoted by t_0, on the curve $\Gamma_0 = \gamma$. Therefore, γ has the Poritsky property with Poritsky–Lazutkin parameter t_0, by the above discussion. Any two reflections T_p and T_q commute while acting on the union of the curves Γ_r^* with $r \leq \min\{p, q\}$: the curves Γ_r^* are T_p- and T_q-invariant, and T_p, T_q act as translations there. Proposition 4.3 is proved.

Proposition 1.6 follows from Proposition 4.3 and Example 4.2.

4.2. Preparatory coboundary property of length ratio

Let Σ be an oriented surface of constant curvature $K \in \{0, \pm 1\}$: either Euclidean plane, or unit sphere in $\mathbb{R}^3$, or hyperbolic plane. Let $O \in \Sigma$, and let $\gamma \subset \Sigma$ be a regular germ of curve through O with positive geodesic curvature. We consider that γ is oriented counterclockwise with respect to the orientation of the surface Σ. For every point $X \in \gamma$ by G_X, we denote the geodesic tangent to γ at X. Let $A, B \in \gamma$ be two distinct points close to O such that the curve γ is oriented from B to A. Let $C = C_{AB}$ denote the unique intersection point of the geodesics G_A and G_B that is close to O. (Then, CA is the right geodesic tangent to γ through C.) Set

$$L_A := |CA|; \ L_B := |CB|;$$

here $|CX|$ is the length of the geodesic arc CX. Recall that we denote

$$\psi(x) = \begin{cases} x, & \text{if } \Sigma \text{ is Euclidean plane,} \\ \sin x, & \text{if } \Sigma \text{ is unit sphere,} \\ \sinh x, & \text{if } \Sigma \text{ is hyperbolic plane.} \end{cases} \tag{4.1}$$

Proposition 4.5 *Let Σ be as above, $\gamma \subset \Sigma$ be a germ of C^2-smooth curve at a point $O \in \Sigma$ with the string Poritsky property. There exists a positive continuous function $u(X)$, $X \in \gamma$, such that for every $A, B \in \gamma$ close enough to O one has*

$$\frac{\psi(L_A)}{\psi(L_B)} = \frac{u(A)}{u(B)}. \tag{4.2}$$

The above statement holds for

$$u = \frac{1}{\kappa}\frac{\mathrm{d}t}{\mathrm{d}s}; \quad t \text{ is the Poritsky parameter.}$$

Proof For every $p > 0$ small enough and every $C \in \Gamma_p$ close enough to O, there are two geodesics issued from the point C that are tangent to γ (Proposition 2.3). The corresponding tangency points $A = A(C)$ and $B = B(C)$ in γ depend smoothly on the point $C \in \Gamma_p$. Let s_p denote the natural length parameter of the curve Γ_p. We set $s = s_0$: the natural length parameter of the curve γ. We write $C = C(s_p)$, and consider the natural parameters $s_A(s_p)$, $s_B(s_p)$ of the points $A(C)$ and $B(C)$ as functions of s_p. Let $\alpha(C)$ denote the oriented angle between a vector $v \in T_C\Gamma_p$ orienting the curve Γ_p and a vector $\zeta_A \in T_C G_A$ directing the geodesic G_A from C to A. It is equal (but with opposite sign) to the oriented angle between the vector $-v$ and a vector $\zeta_B \in T_C G_B$ directing the geodesic G_B from C to B, since the tangent line to Γ_p at C is the exterior bisector of the angle between the geodesics G_A and G_B (Theorem 1.3). One has

$$\frac{\mathrm{d}s_A}{\mathrm{d}s_p} = \frac{\sin\alpha(C)}{\kappa(A(C))\psi(|AC|)}, \ \frac{\mathrm{d}s_B}{\mathrm{d}s_p} = \frac{\sin\alpha(C)}{\kappa(B(C))\psi(|BC|)}, \tag{4.3}$$

by (2.8), (2.7) and the above angle equality.

Let now t be the Poritsky parameter of the curve γ. Let $t_A(s_p)$ and $t_B(s_p)$ denote its values at the points $A(C)$ and $B(C)$, respectively, as functions of s_p. Their difference is constant, by the Poritsky property. Therefore,

$$\frac{\mathrm{d}t_A}{\mathrm{d}s_p} = \frac{\mathrm{d}t}{\mathrm{d}s}(A)\frac{\mathrm{d}s_A}{\mathrm{d}s_p} = \frac{\mathrm{d}t_B}{\mathrm{d}s_p} = \frac{\mathrm{d}t}{\mathrm{d}s}(B)\frac{\mathrm{d}s_B}{\mathrm{d}s_p}.$$

Substituting (4.3) to the latter formula and canceling out $\sin\alpha(C)$ yields (4.2) with $u = \frac{1}{\kappa}\frac{dt}{ds}$.

4.3. Conics and Ceva's Theorem on surfaces of constant curvature: proof of Theorem 1.7

Definition 4.6 Let Σ be a surface with Riemannian metric. We say that a germ of curve $\gamma \subset \Sigma$ at a point O with positive geodesic curvature has *tangent incidence property,* if the following statement holds. Let $A', B', C' \in \gamma$ be arbitrary three distinct points close enough to O. Let a, b, and c denote the geodesics tangent to γ at A', B', and C', respectively. Let A, B, and C denote the points of intersection $b \cap c$, $c \cap a$, and $a \cap b$, respectively. Then, the geodesics AA', BB', and CC' intersect at one point. See [21, p. 462, fig.5] and Fig. 4 as follows.

Proposition 4.7 *Every germ of C^2-smooth curve with the string Poritsky property on a surface of constant curvature has tangent incidence property.*

As it is shown below, Proposition 4.7 follows from Proposition 4.5 and the next theorem.

Theorem 4.8 [17, pp. 3201–3203] (Ceva's Theorem on surfaces of constant curvature.) *Let Σ be a simply connected complete surface of constant curvature. Let $\psi(x)$ be the corresponding function in (4.1): the length of circle of radius x divided by 2π. Let $A, B, C \in \Sigma$ be three distinct points. Let A', B', and C' be, respectively, some points on the* **sides** *BC, CA, and AB of the geodesic triangle ABC. Then, the geodesics AA', BB', and CC' intersect at one point, if and only if*

$$\frac{\psi(|AB'|)}{\psi(|B'C|)}\frac{\psi(|CA'|)}{\psi(|A'B|)}\frac{\psi(|BC'|)}{\psi(|C'A|)} = 1. \tag{4.4}$$

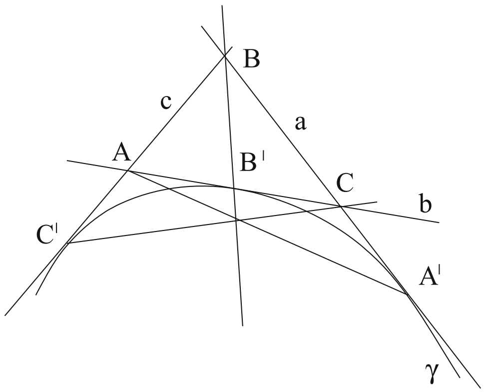


FIGURE 4. A curve γ with tangent incidence property

Addendum to Theorem 4.8. *Let now in the conditions of Theorem 4.8* A', B', and C' *be points on the* **geodesics** BC, CA, and AB, *respectively, so that some two of them, say* A' and C' *do not lie on the corresponding* **sides** *and the remaining third point* B' *lies on the corresponding side* AC, *see Fig. 4.*

(1) *In the Euclidean and spherical cases, the geodesics* AA', BB', and CC' *intersect at the same point, if and only if* (4.4) *holds.*
(2) *In the hyperbolic case (when* Σ *is of negative curvature) the geodesics* AA', BB', and CC' *intersect at the same point, if and only if some two of them intersect and* (4.4) *holds.*
(3) *Consider the standard model of the hyperbolic plane* Σ *in the Minkowski space* $\mathbb{R}^3$, *see Sect.* 1.1. *Consider the 2-subspaces defining the geodesics* AA', BB', and CC', *and let us denote the corresponding projective lines (i.e., their tautological projections to* $\mathbb{RP}^2$*) by* $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$, *respectively. The projective lines* $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ *intersect at one point (which may be not the projection of a point in* Σ*), if and only if* (4.4) *holds.*

Proof Statements (1) and (2) of the addendum follow from Theorem 4.8 by analytic extension, when some two points A' and C' go out of the corresponding sides BC and BA while remaining on the same (complexified) geodesics BC and BA. Statement (3) is proved analogously.

Proof of Proposition 4.7 Let O be the base point of the germ γ, and let A', B', and C' be its three subsequent points close enough to O. Let a, b, and c be, respectively, the geodesics tangent to γ at them. Then, each pair of the latter geodesics intersect at one point close to O. Let A, B, and C be the points of intersections $b\cap c$, $c\cap a$, and $a\cap b$, respectively. The point B' lies on the geodesic arc $AC \subset b$. This follows from the assumption that the point B' lies between A' and C' on the curve γ and the inequality $\kappa \neq 0$. In a similar way, we get that the points A' and C' lie on the corresponding geodesics a and c but outside the sides BC and AB of the geodesic triangle ABC so that A lies between C' and B, and C lies between A' and B. The geodesics BB'

and AA' intersect, by the two latter arrangement statements. Let $u : \gamma \to \mathbb{R}$ be the function from Proposition 4.5. One has $\frac{\psi(|BA'|)}{\psi(|BC'|)} = \frac{u(A')}{u(C')}$, by (4.2), and similar equalities hold with B replaced by A and C. Multiplying the three latter equalities, we get (4.4), since the right-hand side cancels out. Hence, the geodesics AA', BB' and CC' intersect at one point, by statements (1) and (2) of the addendum to Theorem 4.8. Proposition 4.7 is proved.

Theorem 4.9 *Each conic on a surface of constant curvature has tangent incidence property. Vice versa, each C^2-smooth curve on a surface of constant curvature that has tangent incidence property is a conic.*

Proof The first, easy statement of the theorem follows from Propositions 1.6 and 4.7. The proof of its second statement repeats the arguments from [21, p.462], which are given in the Euclidean case but remain valid in the other cases of constant curvature without change. Let us repeat them briefly in full generality for completeness of presentation. Let γ be a germ of curve with tangent incidence property on a surface Σ of constant curvature. Let A', B', and C' denote three distinct subsequent points of the curve γ, and let a, b, and c be, respectively, the geodesics tangent to γ at these points. Let A, B, and C denote, respectively, the points of intersections $b \cap c$, $c \cap a$, and $a \cap b$. Fix the points A' and C'. Consider the pencil $\mathcal{C}$ of conics through A' and C' that are tangent to $T_{A'}\gamma$ and $T_{C'}\gamma$. Then, each point of the surface Σ lies in a unique conic in $\mathcal{C}$ (including two degenerate conics: the double geodesic $A'C'$; the union of the geodesics $G_{A'}$ and $G_{C'}$). Let $\phi \in \mathcal{C}$ denote the conic passing through the point B'.

Claim. *The tangent line $l = T_{B'}\phi$ coincides with $T_{B'}\gamma$.*

Proof Let L denote the geodesic through B' tangent to l. Let C_1 and A_1 denote, respectively, the points of intersections $L \cap a$ and $L \cap c$. Both curves γ and ϕ have tangent incidence property. Therefore, the three geodesics AA', BB', CC' intersect at the same point denoted X, and the three geodesics $A'A_1$, BB', $C'C_1$ intersect at the same point Y; both X and Y lie on the geodesic BB'. We claim that this is impossible, if $l \neq T_{B'}\gamma$ (or equivalently, if $L \neq b$). Indeed, let to the contrary, $L \neq b$. Let us turn the geodesic b continuously towards L in the family of geodesics b_t through B', $t \in [0, 1]$: $b_0 = b$, $b_1 = L$, the azimuth of the line $T_{B'}b_t$ turns monotonously (clockwise or counterclockwise), as t increases. Let A_t and C_t denote, respectively, the points of the intersections $b_t \cap c$ and $b_t \cap a$: $A_0 = A$, $C_0 = C$. Let X_t denote the point of the intersection of the geodesics $A'A_t$ and $C'C_t$: $X_0 = X$, $X_1 = Y$. At the initial position, when $t = 0$, the point X_t lies on the fixed geodesic BB'. As t increases from 0 to 1, the points A and C remain fixed, while the points C_t and A_t move monotonously, so that as C_t moves towards (out from) B along the geodesic a, the point A_t moves out from (towards) B along the geodesic c, see Fig. 5. In the first case, when C_t moves towards B and A_t moves out from B, the point X_t moves out of the geodesic BB', to the half-plane bounded by BB' that contains A, and its distance to BB' increases. Hence, $Y = X_1$ does not lie on BB'. The second case is treated analogously. The contradiction thus obtained proves the claim.

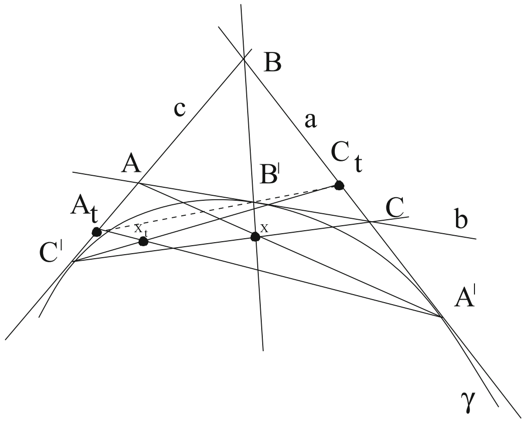


FIGURE 5. The intersection point X_t moves away from the geodesic BB'

For every point $Q \in \Sigma$ such that the conic $\phi_Q \in \mathcal{C}$ passing through Q is regular, set $l_Q := T_Q\phi_Q$. The lines l_Q form an analytic line field outside the union of three geodesics: $G_{A'}$, $G_{C'}$, $A'C'$. Its phase curves are the conics from the pencil $\mathcal{C}$. The curve γ is also tangent to the latter line field, by the above claim. Hence, γ is a conic. This proves Theorem 4.9.

Proof of Theorem 1.7 Let γ be a germ of C^2-smooth curve with the string Poritsky property on a surface of constant curvature. Then, it has tangent incidence property, by Proposition 4.7. Therefore, it is a conic, by Theorem 4.9. Theorem 1.7 is proved.

5. Case of outer billiards: proof of Theorem 1.12

Everywhere below in the present section Σ is a simply connected complete Riemannian surface of constant curvature, and $\gamma \subset \Sigma$ is a germ of C^2-smooth curve at a point $O \in \Sigma$ with positive geodesic curvature.

Proposition 5.1 *Let Σ, O, and γ be as above, and let γ have the area Poritsky property. Then, there exists a continuous function $u : \gamma \to \mathbb{R}_+$ such that for every $A, B \in \gamma$ close enough to O the following statement holds. Let α and β denote the angles between the chord AB and the curve γ at the points A and B, respectively. Then,*

$$\frac{\sin\alpha}{\sin\beta} = \frac{u(A)}{u(B)}. \tag{5.1}$$

Let t and s denote, respectively, the area Poritsky and length parameters of the curve γ. The above statement holds for the function

$$u := t'_s = \frac{\mathrm{d}t}{\mathrm{d}s}.$$

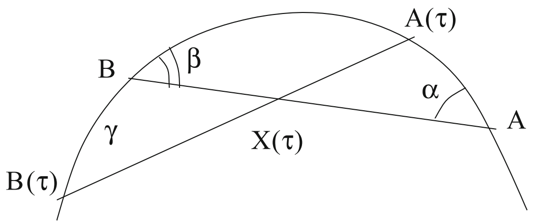


FIGURE 6. Curve γ with the area Poritsky property. The chords AB, $A(\tau)B(\tau)$

Proof Recall that for every $C, D \in \gamma$ by $\lambda(C, D)$, we denote the length of the arc CD of the curve γ. Fix A and B as above. Set $A(0) = A$, $B(0) = B$. For every small $\tau > 0$, let $A(\tau)$ denote the point of the curve γ such that $\lambda(A(\tau), A(0)) = \tau$ and the curve γ is oriented by the natural parameter from $A(0)$ to $A(\tau)$. Let $B(\tau) \in \gamma$ denote the family of points such that the area of the domain bounded by the chord $A(\tau)B(\tau)$ and the arc $A(\tau)B(\tau)$ of the curve γ remains constant, independent on τ. For every τ small enough, the chord $A(\tau)B(\tau)$ intersects the chord $A(0)B(0)$ at a point $X(\tau)$ tending to the middle of the chord $A(0)B(0)$, see Fig. 6. This follows from constance of area and homogeneity (constance of curvature) of the surface Σ. One has

$$t(A(\tau)) - t(A(0)) = t(B(\tau)) - t(B(0)) \text{ for every } \tau \text{ small enough,}$$

by the area Poritsky property. The above left- and right-hand sides are asymptotic to $u(A)\lambda(A(0), A(\tau))$ and $u(B)\lambda(B(0), B(\tau))$, respectively, as $\tau \to 0$, with $u = \frac{dt}{ds}$. Therefore,

$$\frac{\lambda(B(0), B(\tau))}{\lambda(A(0), A(\tau))} \to \frac{u(A)}{u(B)}, \text{ as } \tau \to 0. \tag{5.2}$$

The length $\lambda(A(0), A(\tau))$ is asymptotic to $\frac{1}{\sin\alpha}$ times $\mathrm{dist}(A(\tau), B(0)A(0))$: the distance of the point $A(\tau)$ to the geodesic $X(\tau)A(0) = B(0)A(0)$. Similarly, $\lambda(B(0), B(\tau)) \simeq \frac{1}{\sin\beta}\mathrm{dist}(B(\tau), B(0)A(0))$, as $\tau \to 0$. The above distances of the points $A(\tau)$ and $B(\tau)$ to the geodesic $A(0)B(0)$ are asymptotic to each other, since the intersection point $X(\tau)$ of the chords $A(\tau)B(\tau)$ and $A(0)B(0)$ tends to the middle of the chord $A(0)B(0)$ and by homogeneity. This implies that the left-hand side in (5.2) tends to the ratio $\frac{\sin\alpha}{\sin\beta}$, as $\tau \to 0$. This together with (5.2) proves (5.1).

Proposition 5.2 *Let Σ, O and γ be as at the beginning of the section. Let there exist a function u on γ that satisfies (5.1) for every $A, B \in \gamma$ close to O. Then, γ has tangent incidence property, see Definition 4.6.*

Proof Let A', B', and C' be three subsequent points of the curve γ. Let a, b, and c denote, respectively, the geodesics tangent to γ at these points. Let A, B, and C denote, respectively, the points of intersections $b \cap c$, $c \cap a$, and $a \cap b$ (all the points A', B', and C', and hence A, B, and C are close enough to the base point O), as at Fig. 4. Let ψ be the same, as in (4.1). One has

$$\frac{\sin\angle CA'B'}{\sin\angle CB'A'} = \frac{\psi(|CB'|)}{\psi(|CA'|)} = \frac{u(A')}{u(B')}, \tag{5.3}$$

by (5.1) and Sine Theorem on the Euclidean plane and its analogues for unit sphere and hyperbolic plane applied to the geodesic triangle $CA'B'$, see [14, p.215], [22, theorem 10.4.1]. Similar equalities hold for other pairs of points (B', C'), (C', A'). Multiplying all of them yields relation (4.4): the ratios of values of the function u at A', B', C' cancel out. This together with Theorem 4.8 and its addendum implies that γ has tangent incidence property and proves Proposition 5.2.

Proof of Theorem 1.12 A curve with the area Poritsky property on a surface of constant curvature has tangent incidence property, by Propositions 5.1 and 5.2. Hence, it is a conic, by Theorem 4.9. Theorem 1.12 is proved.

6. The function $L(A, B)$ and the Poritsky–Lazutkin parameter: proofs of Theorems 1.16 and 1.15, and Corollaries 1.17 and 1.18

Proof of Theorem 1.16 Let g denote the metric. Let $C = C_{AB}$ denote the point of intersection of the geodesics G_A and G_B tangent to γ at the points A and B, respectively. We will work in normal coordinates (x, y) centered at C and the corresponding polar coordinates (r, ϕ). Recall that the metric is C^4-smooth, and hence, it is C^2-smooth in normal coordinates. The next two claims concern asymptotics of different quantities, as $\text{dist}(A, B) \to 0$ so that A and B lie in a compact subarc in γ.

Claim 1. *The length $s_A - s_B$ of the arc AB of the curve γ differs from its Euclidean length in the coordinates (x, y) by a quantity $o((s_A - s_B)^3)$. The same statement also holds for the quantity $L(A, B)$. These asymptotics are uniform in the metric running through a closed bounded subset in the space of C^3-smooth Riemannian metrics.*

Proof It is known that the metric g is $O(r^2)$-close to the Euclidean metric, and the polar coordinates are g-orthogonal. In the polar coordinates g has the same radial part dr^2, as the Euclidean metric $dr^2 + r^2 d\phi^2$, and their angular parts differ by a quantity $\Delta = O(r^2) r^2 d\phi^2 = O(r^4 d\phi^2)$. The g-length of the arc AB is the integral of the g-norm of the Euclidean-unit tangent vector field to γ. The integration parameter is the Euclidean natural parameter. The contribution of the above difference Δ to the latter integral is bounded from above by the integral I of a quantity $O(r^2 \alpha)$, where α is the acute angle of a tangent vector $\dot\gamma(Q)$ with the radial line CQ. Set $\delta := |s_A - s_B|$. The arc AB lies in a $O(\delta)$-neighborhood of the point C. The distance of the arc AB to C is of order $O(\delta^2)$. Those points in the arc AB where α is bounded away from zero are on distance $O(\delta^2)$ from the origin C. Therefore, $\alpha = o(1)$, as $\delta \to 0$, uniformly on the complement of the arc AB to the disk $D_{\delta^{\frac{3}{2}}}$ of radius $\delta^{\frac{3}{2}}$ centered at C. Hence, the above integral of $O(r^2\alpha)$ over the complement to the disk $D_{\delta^{\frac{3}{2}}}$ is $o(\delta^3)$. The integral inside this disk is also $o(\delta^3)$, since its intersection with γ has length of order $O(\delta^{\frac{3}{2}})$, while the subintegral expression is $O(\delta^2)$. Finally, the upper bound I for the contribution of the non-Euclidean angular part Δ is $o(\delta^3)$. This implies the statement of the claim for the g-length $s_A - s_B$, and hence, for the expression $L(A,B)$: the g-lengths of the segments AC, BC coincide with their Euclidean lengths by the definition of normal coordinates. The asymptotics of Claim 1 are uniform in the metric, as are the intermediate asymptotics used in the proof.

Claim 2. *Let $\gamma \subset \mathbb{R}^2$ be a C^3-smooth curve with positive geodesic curvature. (Here, we deal with the standard Euclidean metric on $\mathbb{R}^2$.) For every point $A \in \gamma$ consider the osculating circle S_A at A of the curve γ. For every $B \in \gamma$ close to A let us consider the point $B' \in S_A$ closest to B ($BB' \perp S_A$) and the corresponding expressions $\lambda(A, B')$, $L(A, B') = L_{S_A}(A, B')$ written for the circle S_A. One has*

$$\lambda(A,B') - \lambda(A,B) = o((s_A - s_B)^3), \quad L(A,B') - L(A,B) = o((s_A - s_B)^3).$$

Proof Recall that we denote $\delta = |s_A - s_B|$. The lengths of the arcs $AB \subset \gamma$ and $AB' \subset S_A$ differ by a quantity $o(\delta^3)$. Indeed, the projection of the arc AB to the arc AB' along the radii of the circle S_A has norm of derivative of order $1 + o(\delta^2)$. This is implied by the two following statements: (1) the distance between the source and the image is of order $o(\delta^2)$ (the circle is osculating); (2) the slopes of the corresponding tangent lines differ by a quantity $o(\delta)$. The asymptotics $1 + o(\delta^2)$ for the norm of projection implies that $\lambda(A,B) - \lambda(A,B') = o(\delta^3)$. Let us now show that the straight-line parts of the expressions $L(A,B)$ and $L(A,B')$ also differ by a quantity $o(\delta^3)$. The tangent lines $T_B\gamma$ and $T_{B'}S_A$ pass through $o(\delta^2)$-close points B and B', and their slopes differ by a quantity $o(\delta)$, see the above statements (1) and (2). Note that $BB' \perp T_{B'}S_A$. This implies that the distance between their points C and C' of intersection with the line $T_A\gamma$ is $o(\delta)$. Consider the line through C orthogonal to the line $T_{B'}S_A$. Let H denote their intersection point. The difference of the straight-line parts of the expressions $L(A,B)$ and $L(A,B')$ is equal to $(|BC| - |B'H|) \pm (|CC'| - |C'H|)$. The second bracket is the difference of a leg and a hypotenuse, both of order $o(\delta)$, in a right triangle with angle $O(\delta)$ between them. Hence, the latter difference is $o(\delta^3)$, since the cosine of the angle is $1 + O(\delta^2)$. The first bracket is equal to the similar difference in another right triangle, with leg $B'H$ and hypotenuse being the segment BC shifted by the vector $\overrightarrow{BB'}$; both are of order $O(\delta)$, and the angle between them is $o(\delta)$. Hence, the first bracket is $o(\delta^3)$ (the cosine being now $1 + o(\delta^2)$). Finally, the difference of the straight-line parts of the expressions $L(A,B)$ and $L(A,B')$ is $o(\delta^3)$. The claim is proved.

Claims 1 and 2 reduce Theorem 1.16 to the case, when the metric is Euclidean and γ is a circle in $\mathbb{R}^2$. Let R denote its radius. Let AB be its arc cut by a sector of small angle ϕ. Then,

$$L(A,B) = R(2\tan(\frac{\phi}{2}) - \phi) \simeq \frac{R}{12}\phi^3 = \frac{\kappa^2}{12}|s_A - s_B|^3, \ \kappa = R^{-1}.$$

This proves Theorem 1.16.

Proof of Corollary 1.17 Let $C \in \Gamma_p$. Let $A = A(C)$, $B = B(C) \in \gamma$ denote the points such that the geodesics AC and BC are tangent to γ at A and B, respectively. We order them so that $A(C) = \mathcal{T}_p(B(C))$. One has $L(A(C), B(C)) = p$ for all $C \in \Gamma_p$, by definition. On the other hand, $L(A(C), B(C)) \simeq \frac{1}{12}\kappa^2(A(C))|s(A(C)) - s(B(C))|^3$, as C tends to a compact subarc $\widehat{\gamma} \Subset \gamma$, by Theorem 1.16. Therefore, all the quantities $\kappa^{\frac{2}{3}}(A(C))|s(A(C)) - s(B(C))|$ are uniformly asymptotically equivalent. Substituting $A(C) = \mathcal{T}_p(B(C))$, we get (1.6). Corollary 1.17 is proved.

Corollary 1.18 follows immediately from Corollary 1.17.

Proof of Theorem 1.15 Let the curve γ have the string Poritsky property. Let t denote its Poritsky parameter. Set $f := \frac{dt}{ds}$. For the proof of Theorem 1.15 it suffices to show that $f \equiv \kappa^{\frac{2}{3}}$ up to constant factor. Or equivalently,

$$\frac{f(Q)}{f(B)} = \frac{\kappa^{\frac{2}{3}}(Q)}{\kappa^{\frac{2}{3}}(B)} \quad \text{for every } B, Q \in \gamma. \tag{6.1}$$

Fix a small $p > 0$. Set $A := \mathcal{T}_p(B)$, $R = \mathcal{T}_p(Q)$. One has

$$t(A) - t(B) = t(R) - t(Q), \tag{6.2}$$

by the Poritsky property. On the other hand, the latter left- and right-hand sides are asymptotically equivalent, respectively, to $f(B)\lambda(A, B)$ and $f(Q)\lambda(R, Q)$. But

$$\kappa^{\frac{2}{3}}(B)\lambda(A, B) \simeq \kappa^{\frac{2}{3}}(Q)\lambda(R, Q), \text{ as } p \to 0,$$

by Corollary 1.17. Substituting the two latter asymptotics to (6.2) yields (6.1). Theorem 1.15 is proved.

7. Symplectic generalization of Theorem 1.15

In Sect. 7.1, we give a background material on symplectic properties of the billiard ball reflection map. In Sect. 7.2, we introduce weakly billiard-like maps. We consider the so-called string type families of weakly billiard-like maps, which generalize the family of billiard reflections from string construction curves defined by a curve with the string Poritsky property. We state Theorem 7.10, which is a symplectic generalization of Theorem 1.15 (C^6-smooth case) to the string type billiard-like map families. Theorem 7.10 will be proved in Sect. 7.4. For its proof, in Sect. 7.3, we introduce an analogue of Lazutkin coordinates, the so-called modified Lazutkin coordinates, for weakly billiard-like maps (Theorem 7.11) and prove Lemma 7.13 on asymptotics of orbits in these coordinates.

In Sect. 7.5, we show how to retrieve Theorem 1.15 for C^6-smooth curves from Theorem 7.10.

The idea to extend Theorem 1.15 to a more general symplectic context was suggested by Sergei Tabachnikov.

7.1. Symplectic properties of billiard ball map

The background material recalled here can be found in [4,5,16,18,19,23].

Let Σ be a surface with Riemannian metric. Let $\Pi : T\Sigma \to \Sigma$ denote the tautological projection. Let us recall that the *tautological 1-form* α on $T\Sigma$ (also called the *Liouville form*) is defined as follows: for every $(Q, P) \in T\Sigma$ with $Q \in \Sigma$ and $P \in T_Q\Sigma$ for every $v \in T_{(Q,P)}(T\Sigma)$ set

$$\alpha(v) := < P, \Pi_* v > . \tag{7.1}$$

The differential

$$\omega = \mathrm{d}\alpha$$

of the 1-form α is the *canonical symplectic form* on $T\Sigma$.

Let $O \in \Sigma$, and $\gamma \subset \Sigma$ be a germ of regular oriented curve at O. We parametrize it by its natural length parameter s. The corresponding function $s \circ \Pi$ on $T\gamma$ will be also denoted by s. For every $Q \in \gamma$ and $P \in T_Q\gamma$, set

$$\begin{aligned} &\dot\gamma(Q) = \frac{\mathrm{d}\gamma}{\mathrm{d}s}(Q) := \text{ the directing unit tangent vector to } \gamma \text{ at } Q, \\ &\sigma(Q, P) := < P, \dot\gamma(Q) >, \quad y(Q, P) := 1 - \sigma(Q, P). \end{aligned} \tag{7.2}$$

The restriction to $T\gamma$ of the form ω is a symplectic form, which will be denoted by the same symbol ω.

Proposition 7.1 (see [16, formula (3.1)] in the Euclidean case) *The coordinates (s, y) on $T\gamma$ are symplectic: $\omega = ds \wedge dy$ on $T\gamma$.*

Proof The proposition follows from the definition of the symplectic structure $\omega = \mathrm{d}\alpha$, α is the same, as in (7.1): in local coordinates (s, σ) one has $\alpha = \sigma \mathrm{d}s$, thus, $\omega = \mathrm{d}\sigma \wedge \mathrm{d}s = \mathrm{d}s \wedge dy$.

Let V denote the Hamiltonian vector field on $T\Sigma$ with the Hamiltonian $||P||^2$. It generates the geodesic flow. Consider the unit circle bundle:

$$S = \mathcal{T}_1\Sigma := \{||P||^2 = 1\} \subset T\Sigma.$$

It is known that for every point $x \in S$ the kernel of the restriction $\omega|_{T_xS}$ is the one-dimensional linear subspace spanned by the vector $V(x)$ of the field V. Each cross-section $W \subset S$ to the field V is identified with the (local) space of geodesics. The symplectic structure ω induces a well-defined symplectic structure on W called the *symplectic reduction.*

Remark 7.2 The symplectic reduction is invariant under holonomy along orbits of the geodesic flow. Namely, for every arc AB of its orbit and two germs of cross-sections W_1 and W_2 through A and B, respectively, the holonomy mapping $W_1 \to W_2$, $A \mapsto B$ along the arc AB is a symplectomorphism.

Consider the local hypersurface

$$\Gamma = \Pi^{-1}(\gamma) \cap S = (\mathcal{T}_1\Sigma)|_\gamma \subset S.$$

At those points $(Q, P) \in \Gamma$, for which the vector P is transverse to γ, the hypersurface Γ is locally a cross-section for the restriction to S of the geodesic flow. Thus, near the latter points, the hypersurface Γ carries a canonical symplectic structure given by the symplectic reduction. Set

$$\mathcal{O}_{\pm} := (O, \pm\dot{\gamma}(O)) \in \Gamma.$$

Let γ have positive geodesic curvature. For every $(Q, P) \in \Gamma$ close enough to $\mathcal{O}_{\pm}$, the oriented geodesic through Q issued in the direction P intersects γ at two points Q and Q' (which coincide if P is tangent to γ). Let P' denote the orienting unit tangent vector of the latter geodesic at Q'. This defines the germ at $\mathcal{O}_{\pm}$ of involution

$$\beta : (\Gamma, \mathcal{O}_{\pm}) \to (\Gamma, \mathcal{O}_{\pm}), \quad \beta(Q, P) = (Q', P'), \quad \beta^2 = Id, \tag{7.3}$$

which will be called the *billiard ball geodesic correspondence.*

Consider the following open subset in $T\gamma$: the unit ball bundle

$$\mathcal{T}_{\leq 1}\gamma := \{(Q, P) \in T\gamma \mid ||P||^2 \leq 1\}.$$

Let $\pi : (T\Sigma)|_{\gamma} \to T\gamma$ denote the mapping acting by orthogonal projections

$$\pi : T_Q\Sigma \to T_Q\gamma, \ Q \in \gamma.$$

It induces the following projection also denoted by π:

$$\pi : \Gamma \to \mathcal{T}_{\leq 1}\gamma. \tag{7.4}$$

Let $\mathcal{V}$ denote a convex domain with boundary containing γ. Every point $(Q, P) \in \mathcal{T}_{\leq 1}\gamma$ has two π-preimages $(Q, w_{\pm})$ in Γ: the vector w_+ (w_-) is directed inside (respectively, outside) the domain $\mathcal{V}$. The vectors $w_{\pm}$ coincide, if and only if $||P|| = 1$, and in this case they lie in $T_Q\gamma$. Thus, the mapping $\pi : \Gamma \to \mathcal{T}_{\leq 1}\gamma$ has two continuous inverse branches. Let $\mu_+ := \pi^{-1} : \mathcal{T}_{\leq 1}\gamma \to \Gamma$ denote the inverse branch sending P to w_+, cf. [16, section 2]. The above mappings define the germ of mapping

$$\delta_+ := \pi \circ \beta \circ \mu_+ : (\mathcal{T}_{\leq 1}\gamma, \mathcal{O}_{\pm}) \to (\mathcal{T}_{\leq 1}\gamma, \mathcal{O}_{\pm}). \tag{7.5}$$

Recall that Γ carries a canonical symplectic structure given by the above-mentioned symplectic reduction (as a cross-section), and $T\gamma$ carries the standard symplectic structure: the restriction to $T\gamma$ of the form $\omega = ds \wedge dy$.

Theorem 7.3 *[23, subsection 1.5], [4, 5, 18, 19] The mappings β, π, and hence, δ_+ given by (7.3), (7.4) and (7.5), respectively, are symplectic.*

Proof Symplecticity of the mapping β follows from the definition of symplectic reduction and its holonomy invariance (Remark 7.2). Symplecticity of the projection π follows from definition and the fact that the π-pullback of the tautological 1-form α on $T\gamma$ is the restriction to Γ of the tautological 1-form on $T\Sigma$. This implies symplecticity of $\mu_+ = \pi^{-1}$, and hence, δ_+.

Let $I : \Gamma \to \Gamma$ denote the reflection involution

$$I : (Q, P) \mapsto (Q, P^*),$$
$$Q \in \gamma, \ P^* := \text{ the vector symmetric to } P \text{ with respect to the line } T_Q\gamma.$$

Let $\Gamma_+ \subset \Gamma$ denote the subset of those points (Q, P) for which P either is directed inside the convex domain $\mathcal{V}$, or coincides with $\dot{\gamma}(Q)$.

Proposition 7.4 *The involution I preserves the tautological 1-form α, and hence, is symplectic. The involutions I and β are C^r-smooth germs of mappings $(\Gamma, \mathcal{O}_\pm) \to (\Gamma, \mathcal{O}_\pm)$, if the metric and the curve γ are C^{r+1}-smooth. The mapping δ_+ is conjugated to their product acting on Γ_+:*

$$\widetilde{\delta}_+ := I \circ \beta = \mu_+ \circ \delta_+ \circ \mu_+^{-1}. \tag{7.6}$$

The proposition follows immediately from definitions.

The billiard transformation T of reflection from the curve γ acts on the space of oriented geodesics that intersect γ and are close enough to the geodesic tangent to γ at O. Each of them intersects γ at two points (which coincide, if the geodesic is tangent to γ). To each oriented geodesic G we put into correspondence a point $(Q, P) \in \Gamma_+$, where Q is its first intersection point with γ (in the sense of the orientation of the geodesic G) and P is the orienting unit vector tangent to G at Q. This is a locally bijective correspondence.

Proposition 7.5 *Let the metric and the curve γ be C^3-smooth. The billiard mapping T written as a mapping $\Gamma_+ \to \Gamma_+$ via the above correspondence coincides with $\widetilde{\delta}_+$. Consider the coordinates (s, ϕ) on Γ: $s = s(Q)$ is the natural length parameter of a point $Q \in \gamma$; $\phi = \phi(Q, P)$ is the oriented angle of the vector $\dot{\gamma}(Q)$ with a vector $P \in T_Q\Sigma$. In the coordinates (s, ϕ) the mappings I, β and $T = \widetilde{\delta}_+$ are C^2-smooth and take the form*

$$I(s, \phi) = (s, -\phi), \ \ \beta(s, \phi) = (s + 2\kappa^{-1}(s)\phi + O(\phi^2), -\phi + O(\phi^2)), \tag{7.7}$$
$$\widetilde{\delta}_+(s, \phi) = (s + 2\kappa^{-1}(s)\phi + O(\phi^2), \phi + O(\phi^2)). \tag{7.8}$$

The asymptotics are uniform in s, as $\phi \to 0$. In the coordinates

$$(s, y), \quad y = 1 - \cos\phi, \tag{7.9}$$

see (7.2), the billiard mapping T coincides with δ_+ and takes the form

$$\delta_+(s, y) = (s + 2\sqrt{2}\kappa^{-1}(s)\sqrt{y} + O(y), y + O(y^{\frac{3}{2}})). \tag{7.10}$$

Proof All the statements of the proposition except for the formulas follow from definition. Formula (7.7) follows from the definitions of the mappings I and β: a geodesic issued from a point $Q \in \gamma$ at a small angle ϕ with the tangent vector $\dot{\gamma}(Q)$ intersects γ at a point Q' such that $\lambda(Q, Q') = 2\kappa^{-1}(Q)\phi + O(\phi^2)$. The latter formula follows from its Euclidean analogue (applied to the curve γ represented in normal coordinates centered at Q), Proposition 2.2 and smoothness. Formulas (7.7) and (7.6) imply (7.8), which in its turn implies (7.10), since $y = \frac{\phi^2}{2} + O(\phi^4)$.

7.2. Families of billiard-like maps with invariant curves: a symplectic version of Theorem 1.15

In this and the next subsections, we study the following class of area-preserving mappings generalizing the billiard mappings (7.10).

Definition 7.6 A *weakly billiard-like map* is a germ of mapping preserving the standard area form $dx \wedge \mathrm{d}y$,

$$F : (\mathbb{R} \times \mathbb{R}_{\geq 0}, (0, 0)) \to (\mathbb{R} \times \mathbb{R}_{\geq 0}, (0, 0)),$$

$$F = (f_1, f_2) : (x, y) \mapsto (x + w(x)\sqrt{y} + O(y), y + O(y^{\frac{3}{2}})),\ \ w(x) > 0, \tag{7.11}$$

for which the x-axis is a line of fixed points and such that the variable change

$$(x, y) \mapsto (x, \phi),\ \ y = \phi^2$$

conjugates F to a C^2-smooth germ $\widetilde{F}(x, \phi)$. The above asymptotics are uniform in x, as $y \to 0$. If, in addition to the above assumptions, the latter mapping $\widetilde{F}$ is a product of two involutions:

$$\begin{aligned}&\widetilde{F} = I \circ \beta,\ \ I(x, \phi) = (x, -\phi),\\ &\beta(x, \phi) = (x + w(x)\phi + O(\phi^2), -\phi + O(\phi^2)),\ \ \beta^2 = Id,\end{aligned} \tag{7.12}$$

then F will be called *a (strongly) billiard-like map.*

Example 7.7 The mapping δ_+ from (7.10) is strongly billiard-like in the coordinates (s, y) with $w(s) = 2\sqrt{2}\kappa^{-1}(s)$, see (7.6), (7.7), (7.8) and (7.10).

The next definition generalizes the notion of string curve family to weakly billiard-like maps.

Definition 7.8 A family $F_\varepsilon(x, y)$ of weakly billiard-like maps (7.11) depending on a parameter $\varepsilon \in [0, \varepsilon_0]$ is of *string type,* if the derivatives up to order 2 of the corresponding mappings $\widetilde{F}_\varepsilon(x, \phi)$ are continuous in (x, ϕ, ε) on a product $\{|x| \leq \delta_0\} \times [0, \phi_0] \times [0, \varepsilon_0]$ and there exist a $\delta \in (0, \delta_0]$ and a family γ_ε of F_ε-invariant graphs of continuous functions $h_\varepsilon : [-\delta, \delta] \to \mathbb{R}_{\geq 0}$,

$$\gamma_\varepsilon = \{y = h_\varepsilon(x)\}, \tag{7.13}$$

such that γ_ε converge to the x-axis: $h_\varepsilon(x) \to 0$ uniformly on $[-\delta, \delta]$.

Example 7.9 Let $\gamma \subset \Sigma$ be a germ of curve with positive geodesic curvature such that the corresponding string construction curves Γ_p are C^3-smooth and their 3-jets depend continuously on the base points. (For example, this holds automatically in the case, when the curve γ and the metric are C^6-smooth, see Theorem 1.3.) Then, the family of billiard reflection maps from the curves Γ_p is a string type family. The invariant curves γ_p from (7.13) are identified with one and the same curve in the space of oriented geodesics: the family of geodesics tangent to the curve γ and oriented by its tangent vectors $\dot{\gamma}$. See Sect. 7.5 for more details.

The next theorem deals with string type families of weakly billiard-like maps satisfying an analogue of the Poritsky property. It extends Theorem 1.15 on coincidence of Poritsky and Lazutkin parameters.

Theorem 7.10 *Let $F_\varepsilon(x, y)$ be a string type family of weakly billiard maps. Let for every ε small enough there exist a continuous strictly increasing parameter t_ε on γ_ε in which $F|_{\gamma_\varepsilon}$ is a translation by ε-dependent constant,*

$$t_\varepsilon \circ F|_{\gamma_\varepsilon} = t_\varepsilon + c(\varepsilon), \tag{7.14}$$

such that the parameter $t_\varepsilon = t_\varepsilon(x)$ *considered as a function of* x *converges to a strictly increasing function* $t_0(x)$ *uniformly on* $[-\delta, \delta]$*, as* $\varepsilon \to 0$*. Then,*

$$t_0 = aX + b, \ X := \int_0^x w^{-\frac{2}{3}}(z)\mathrm{d}z, \ a, b \equiv const. \tag{7.15}$$

Here, $w = w_0(x)$ *is the function from (7.11) corresponding to the mapping* F_ε *with* $\varepsilon = 0$*.*

Theorem 7.10 is proved in Sect. 7.4.

7.3. Modified Lazutkin coordinates and asymptotics

In the proof of Theorem 7.10, we use the following well-known theorem.

Theorem 7.11 *Let* F *be a weakly billiard-like map, and let* $w(x)$ *be the corresponding function in (7.11). The transformation*

$$\mathcal{L} : (x, y) \mapsto (X, Y), \ \begin{cases} X(x) = \int_0^x w^{-\frac{2}{3}}(z)dz \\ Y(x, y) := w^{\frac{2}{3}}(x)y \end{cases} \tag{7.16}$$

is symplectic. Its post-composition with the variable change $(X, Y) \mapsto (X, \Phi)$*,* $\Phi := \sqrt{Y}$*, conjugates* F *to a mapping with the asymptotics*

$$F : (X, \Phi) \mapsto (X + \Phi + o(\Phi), \Phi(1 + o(\Phi))), \ \text{as } \Phi \to 0, \tag{7.17}$$

uniform in X*. The coordinates* (X, Φ) *will be called the* **modified Lazutkin coordinates***.*

A version of Theorem 7.11 is implicitly contained in [15,16]. For completeness of presentation, we present its proof using the following proposition.

Proposition 7.12 *The* y*-component of a weakly billiard-like map (7.11) admits the following more precise formula:*

$$f_2(x, y) = y - \frac{2}{3}w'(x)y^{\frac{3}{2}} + o(y^{\frac{3}{2}}). \tag{7.18}$$

Proof Recall that $\widetilde{F} = (\widetilde{f}_1, \widetilde{f}_2)$ is a C^2-smooth mapping, $\phi = \sqrt{y}$, $\widetilde{f}_2(x, \phi) = \sqrt{f_2(x, \phi^2)}$. Consider the Taylor expansion of the function $\widetilde{f}_2$ in ϕ:

$$\begin{aligned} &\widetilde{f}_2(x, \phi) = \phi + c(x)\phi^2 + o(\phi^2), \\ &f_2(x, y) = \widetilde{f}_2^2(x, \phi) = y(1 + c(x)\sqrt{y} + o(\sqrt{y}))^2 = y + 2c(x)y^{\frac{3}{2}} + o(y^{\frac{3}{2}}), \\ &\frac{\partial f_2}{\partial y}(x, y) = \frac{1}{\sqrt{y}}\widetilde{f}_2\frac{\partial \widetilde{f}_2}{\partial \phi}(s, \phi) = 1 + 3c(x)\sqrt{y} + o(\sqrt{y}). \end{aligned} \tag{7.19}$$

This together with analogous calculations of the other partial derivatives,

$$\begin{aligned} &\frac{\partial f_1}{\partial x} = 1 + w'(x)\sqrt{y} + o(\sqrt{y}), \\ &\frac{\partial f_1}{\partial y} = O(y^{-\frac{1}{2}}), \frac{\partial f_2}{\partial x} = 2\widetilde{f}_2(x, \phi)\frac{\partial \widetilde{f}_2}{\partial x}(x, \phi) = o(\phi^2) = o(y), \end{aligned}$$

shows that the Jacobian of the mapping $F(x, y)$ equals $1 + (w'(x) + 3c(x))\sqrt{y} + o(\sqrt{y})$. But it should be equal to 1, by symplecticity. Therefore, $c(x) = -\frac{1}{3}w'(x)$. This together with (7.19) proves the proposition.

Proof of Theorem 7.11 Symplecticity of the transformation $\mathcal{L}$ follows from definition. Let us show that the coordinate change $(x,y)\mapsto(X,\Phi)$ conjugates F to a mapping with asymptotics (7.17). One has

$$X\circ F(x,y) = X + \int_x^{x+w(x)\sqrt{y}+O(y)} w^{-\frac{2}{3}}(z)dz$$
$$= X + w(x)w^{-\frac{2}{3}}(x)\sqrt{y} + O(y) = X + \Phi + O(\Phi^2), \qquad (7.20)$$
$$\Phi\circ F(x,y) = \sqrt{w^{\frac{2}{3}}(f_1(x,y))f_2(x,y)}$$
$$= w^{\frac{1}{3}}(x+w(x)\sqrt{y}+o(\sqrt{y}))\sqrt{y(1-\frac{2}{3}w'(x)y^{\frac{1}{2}}+o(y^{\frac{1}{2}}))}.$$

Substituting the expressions $\sqrt{y}=\Phi w^{-\frac{1}{3}}(x)$ and

$$w^{\frac{1}{3}}(x+w(x)\sqrt{y}+o(\sqrt{y})) = w^{\frac{1}{3}}(x) + \frac{1}{3}w^{-\frac{2}{3}}(x)w'(x)w(x)\sqrt{y}+o(\sqrt{y})$$
$$= w^{\frac{1}{3}}(x)(1+\frac{1}{3}w'(x)\sqrt{y}+o(\sqrt{y}))$$

to the above formula yields

$$\Phi\circ F(x,y) = \Phi(1+\frac{1}{3}w'(x)y^{\frac{1}{2}}+o(y^{\frac{1}{2}}))(1-\frac{1}{3}w'(x)y^{\frac{1}{2}}+o(y^{\frac{1}{2}})) = \Phi + o(\Phi^2).$$

This together with (7.20) proves (7.17).

We use the following lemma on asymptotics of orbits of a mapping (7.17).

Lemma 7.13 *Let $V_{\Delta,\sigma} := [-\Delta,\Delta]\times[0,\sigma]\subset\mathbb{R}^2_{(X,\Phi)}$, $F: V_{\Delta,\sigma}\to F(V_{\Delta,\sigma})$ be a homeomorphism with asymptotics (7.17) uniform in $X\in[-\Delta,\Delta]$. There exist functions $\alpha(z),\beta(z)>0$, $\alpha(z),\beta(z)\to 0$, as $z\to 0$ such that for every $\eta\in(0,\frac{\sigma}{4})$ small enough the following statements hold. Fix an arbitrary $\delta\in(0,\Delta)$. For every $q_0\in V_{\delta,\eta}$ its two-sided orbit in $V_{\delta,\sigma}$ is a finite sequence:*

$$\mathcal{O} := (q_{j_{\min}},\dots,q_{-1},q_0,q_1,\dots,q_{j_{\max}}),\ q_j=(X_j,\Phi_j),\ q_{j+1}=F(q_j), \qquad (7.21)$$
$$X_{j_{\min}-1} = X\circ F^{-1}(q_{j_{\min}}) < -\delta, \quad X_{j_{\max}+1} = X\circ F(q_{j_{\max}}) > \delta. \qquad (7.22)$$

The following inequalities hold for every $j=j_{\min}-1,\dots,j_{\max}+1$:

$$|\ln\frac{\Phi_j}{\Phi_0}| \le \alpha(\eta); \qquad (7.23)$$
$$e^{-\beta(\eta)}\Phi_0 \le X_{j+1}-X_j \le e^{\beta(\eta)}\Phi_0. \qquad (7.24)$$

Addendum to Lemma 7.13. *Let F_ε be a family of homeomorphisms defined on $V_{\Delta,\eta}$ and depending on a parameter $\varepsilon\in[0,\varepsilon_0]$ with asymptotics (7.17) being uniform in $(X,\varepsilon)\in[-\Delta,\Delta]\times[0,\varepsilon_0]$. Then, all the statements of the lemma hold with functions α and β independent on ε.*

Proof of Lemma 7.13 The second component of asymptotics (7.17) is equivalent to the uniform asymptotics $\ln\frac{\Phi\circ F(X,\Phi)}{\Phi}=o(\Phi)$: to the existence of a non-decreasing function $u(\Phi)>0$, $u(\Phi)\to 0$, as $\Phi\to 0$, such that

$$|\ln\frac{\Phi\circ F^{\pm 1}(X,\Phi)}{\Phi}|\leq \Phi u(\Phi). \tag{7.25}$$

The first component of asymptotics (7.17) is equivalent to the existence of a non-decreasing function $v(\Phi)>0$, $v(\Phi)\to 0$, as $\Phi\to 0$, for which

$$\Phi(1-v(\Phi))\leq X\circ F(X,\Phi)-X\leq \Phi(1+v(\Phi)). \tag{7.26}$$

Consider the maximal connected piece $\mathcal{O}_4$ of the orbit $\mathcal{O}$ containing q_0 whose points have Φ-coordinates satisfying the inequality $\frac{\Phi_0}{4}\leq\Phi_j\leq 4\Phi_0$:

$$\begin{aligned}&\mathcal{O}_4:=(q_{j_{\min,4}},\dots,q_0,\dots,q_{j_{\max,4}}),\ j_{\min}\leq j_{\min,4}\leq 0\leq j_{\max,4}\leq j_{\max},\\ &\frac{1}{4}\Phi_0\leq\Phi_j\leq 4\Phi_0 \text{ for every } j\in[j_{\min,4},j_{\max,4}].\end{aligned} \tag{7.27}$$

By definition, if $j_{\min,4}>j_{\min}$, then (7.27) does not hold for $j=j_{\min,4}-1$. Analogous statement holds for $j_{\max,4}$. Let us choose an $\eta>0$ small enough so that $u(4\eta),v(4\eta)<\frac{1}{4}$. Then, for every $j=j_{\min,4},\dots,j_{\max,4}$ one has

$$\frac{1}{8}\Phi_0\leq X_{j+1}-X_j\leq 8\Phi_0, \tag{7.28}$$

by (7.27) amd (7.26). Set $N:=j_{\max,4}-j_{\min,4}+1=|\mathcal{O}_4|$. One has

$$N\leq\frac{16\delta}{\Phi_0}+1<\frac{16\Delta}{\Phi_0}, \tag{7.29}$$

whenever $\Phi_0<\Delta-\delta$, by (7.28). For every $i\in[j_{\min,4}-1,j_{\max,4}]$ one has

$$|\ln\frac{\Phi_{i+1}}{\Phi_i}|\leq 4\Phi_0 u(4\eta), \tag{7.30}$$

by (7.25) and (7.27). Summing up the latter inequality and using (7.29), we get the following inequality for $j\in[j_{\min,4}-1,j_{\max,4}+1]$:

$$|\ln\frac{\Phi_j}{\Phi_0}|\leq Nu(4\eta)\Phi_0\leq\alpha(\eta):=16\Delta u(4\eta). \tag{7.31}$$

One has obviously $\alpha(\eta)\to 0$, as $\eta\to 0$. This proves (7.23) for $j\in[j_{\min,4}-1,j_{\max,4}+1]$. Inequality (7.24) for the same j with

$$\beta(\eta)=-\ln(1-v(4\eta))+\alpha(\eta)$$

follows from (7.23) and (7.26).

Claim. *For every $\eta>0$ small enough (such that $\alpha(\eta),4\eta u(4\eta)<\frac{1}{8}$) and every $q_0\in V_{\delta,\eta}$ one has $\mathcal{O}_4=\mathcal{O}$: that is, $j_{\min,4}=j_{\min}$, $j_{\max,4}=j_{\max}$.*

Proof Suppose the contrary, for some η as above and some $q_0=(X_0,\Phi_0)\in V_{\delta,\eta}$ one has, say, $j_{\max,4}<j_{\max}$. Set $j_0:=j_{\max,4}$. Then,

$$|\ln\frac{\Phi_{j_0+1}}{\Phi_{j_0}}|,|\ln\frac{\Phi_{j_0}}{\Phi_0}|<\frac{1}{8},$$

by (7.30) and (7.23) for $j = j_0$. Adding the latter inequalities we get $|\ln \frac{\Phi_{j_0+1}}{\Phi_0}| < \frac{1}{4}$, thus, $\frac{1}{4}\Phi_0 < \Phi_{j_0+1} < 4\Phi_0$. The contradiction thus obtained with the definition of the number $j_{\max,4}$ (maximality) proves the claim.

Let η be small, as in the claim. One has $q_{j_{\max}+1} = F(q_{j_{\max}}) \notin V_{\delta,\sigma}$, by definition. But $\Phi(q_{j_{\max}+1}) \leq e^{\alpha(\eta)}\Phi_0 < 4\eta < \sigma$, by (7.23). Therefore, $X_{j_{\max}+1} > \delta$, by definition and (7.24). This together with a similar argument for the point $q_{j_{\min}-1}$ implies (7.22). Lemma 7.13 is proved.

Proof of the Addendum to Lemma 7.13 The addendum follows from uniformity of asymptotics (7.17) in (X, ε) and from the above proof.

7.4. Proof of Theorem 7.10

Everywhere below we write the mappings F_ε in the coordinates $(X_\varepsilon, \Phi_\varepsilon)$ given by (7.17). We consider that F_ε are well defined on one and the same set $V_{\Delta,\eta} = [-\Delta, \Delta] \times [0, \eta] \subset \mathbb{R}^2_{(X_\varepsilon,\Phi_\varepsilon)}$ for all $\varepsilon \in [0, \varepsilon_0]$. Thus, we identify the above coordinates for all ε and denote them by (X, Φ). To show that the limit parameter t_0 is equal to the Lazutkin coordinate X up to multiplicative and additive constants, we have to show that for every four distinct points in the X-axis with X-coordinates $\mathcal{X}_j$,

$$-\Delta < \mathcal{X}_1 < \mathcal{X}_2 < \mathcal{X}_3 < \mathcal{X}_4 < \Delta,$$

the ratios of lengths of the segments

$$I_1 := [\mathcal{X}_1, \mathcal{X}_2], \quad I_3 := [\mathcal{X}_3, \mathcal{X}_4]$$

in the parameters X and t_0 are equal:

$$\frac{t_0(\mathcal{X}_2) - t_0(\mathcal{X}_1)}{t_0(\mathcal{X}_4) - t_0(\mathcal{X}_3)} = \frac{\mathcal{X}_2 - \mathcal{X}_1}{\mathcal{X}_4 - \mathcal{X}_3}. \tag{7.32}$$

Take a $\varepsilon > 0$ small enough, and consider the corresponding F_ε-invariant curve γ_ε. It can be represented as the graph $\{\Phi = H_\varepsilon(X)\}$ of a continuous function. The parameter t_ε on γ_ε in which F_ε is a translation induces a parameter on the X-axis via projection; the induced parameter will be also denoted by t_ε. Fix a $\delta \in (0, \Delta)$ such that $-\delta < \mathcal{X}_1 < \mathcal{X}_4 < \delta$. Consider the corresponding orbit $\mathcal{O}$ of the point $q_{0,\varepsilon} = (\mathcal{X}_1, H_\varepsilon(\mathcal{X}_1)) \in \gamma_\varepsilon$, see (7.21), and let us denote its points by $q_{j,\varepsilon} := F^j(q_{0,\varepsilon})$. Set

$$\nu(\varepsilon) := \Phi(q_{0,\varepsilon}) = H_\varepsilon(\mathcal{X}_1);\ \nu(\varepsilon) \to 0, \text{ as } \varepsilon \to 0.$$

The sequence $X_j := X(q_{j,\varepsilon})$ is strictly increasing with steps having uniform asymptotics $\nu(\varepsilon)(1+o(1))$, as $\varepsilon \to 0$, by Lemma 7.13 and its addendum. For every $i = 1, 2, 3, 4$ let $j_i = j_{i,\varepsilon}$ denote the maximal number j for which $X_j \leq \mathcal{X}_i$. By definition, $j_1 = 0$. For every $i = 2, 3, 4$ and every ε small enough, one has $\mathcal{X}_i - X_{j_i} < 2\nu(\varepsilon)$, by the above asymptotics. The sequence $t_\varepsilon(X_j)$ is an arithmetic progression, since $F_\varepsilon|_{\gamma_\varepsilon}$ acts as a translation in the parameter t_ε. Its step tends to zero, as $\varepsilon \to 0$, since t_ε limits to a strictly increasing continuous parameter t_0 and the X-lengths of steps tend to zero uniformly.

This implies that the ratio of the t_ε-lengths of the segments I_1 and I_3 has the same finite asymptotics, as the ratio

$$R_{1,3}(\varepsilon) := \frac{j_{2,\varepsilon} - j_{1,\varepsilon}}{j_{4,\varepsilon} - j_{3,\varepsilon}}.$$

But the ratio of their X-lengths has also the same asymptotics, as $R_{1,3}(\varepsilon)$, since all the steps of the sequence $X(q_{j,\varepsilon})$ are asymptotically equivalent to one and the same quantity $\nu(\varepsilon)$. This proves (7.32) and Theorem 7.10.

7.5. Deduction of Theorem 1.15 (case C^6) from Theorem 7.10

Let the metric on Σ be C^6-smooth. Let $\gamma \subset \Sigma$ be a germ of C^6-smooth curve with the Poritsky property. Let Γ_ε be the corresponding family of string curves. Let $\widetilde{F}_\varepsilon := \widetilde{\delta}_{+,\varepsilon}$ be the billiard ball maps (7.6) defined by reflections from the curves Γ_ε; see also (7.8). We write them in the coordinates $(s_\varepsilon, \phi_\varepsilon)$ associated to Γ_ε on the space of oriented geodesics, see Proposition 7.5. The curves Γ_ε form a foliation tangent to a C^2-smooth line field on the closure of the concave domain adjacent to γ. They are C^3-smooth, and their 3-jets depend continuously on points. Both statements follow from Theorem 1.3. This implies that the mappings $\widetilde{F}_\varepsilon = \widetilde{\delta}_{+,\varepsilon}(s_\varepsilon, \phi_\varepsilon)$ have derivatives of order up to 2 that are continuous in $(s_\varepsilon, \phi_\varepsilon, \varepsilon)$. Therefore, the corresponding maps $F_\varepsilon := \delta_{+,\varepsilon} = \delta_\varepsilon(s_\varepsilon, y_\varepsilon)$, $y_\varepsilon = 1 - \cos\phi_\varepsilon$, see (7.9), (7.5), (7.10), are strongly billiard-like.

The maps F_ε have invariant curves γ_ε, which are identified with the family of geodesics tangent to the curve γ and oriented as γ. In the coordinates $(s_\varepsilon, y_\varepsilon)$ the curves γ_ε are graphs of continuous functions converging to zero uniformly, as $\varepsilon \to 0$, by construction.

Let now γ have the string Poritsky property. Then, the Poritsky parameter t induces a parameter denoted by t_ε on each invariant curve γ_ε: by definition, the value of the parameter t_ε at a geodesic tangent to γ is the value of the Poritsky parameter t at the tangency point. The maps $\delta_\varepsilon : \gamma_\varepsilon \to \gamma_\varepsilon$ act by translations in the parameters t_ε. The parameters t_ε obviously converge uniformly to the Poritsky parameter $t = t_0$ of the curve $\gamma = \Gamma_0$, as $\varepsilon \to 0$. Therefore, the billiard ball maps F_ε satisfy the conditions of Theorem 7.10 with $w = 2\sqrt{2}\kappa^{-1}$, see Example 7.7. This together with Theorem 7.10 implies that $t_0 = at_L + b$, $a, b \equiv const$, and proves Theorem 1.15 in the case, when the metric and the curve γ are C^6-smooth.

8. Osculating curves with the string Poritsky property: proof of Theorem 1.19

Here, we prove Theorem 1.19, which states that a germ of curve with the string Poritsky property is uniquely determined by its 4-jet.

8.1. Cartan distribution, a generalized version of Theorem 1.19 and plan of the section

Everywhere below for a curve (function) γ by $j^r_p\gamma$, we denote its r-jet at the point p. Set

$$\mathcal{F}^r := \text{ the space of } r\text{-jets of functions of one variable } x \in \mathbb{R}.$$

Let Σ be a C^m-smooth two-dimensional manifold. For every $r \in \mathbb{Z}_{\geq 0}$, $r \leq m$, set

$$\mathcal{J}^r = \mathcal{J}^r(\Sigma) := \text{ the space of } r\text{-jets of regular curves in } \Sigma.$$

In more detail, a *germ of regular curve* is the graph of a germ of function $\{y = h(x)\}$ in appropriate local chart (x, y). Locally a neighborhood in $\mathcal{J}^r$ of the jet of a given C^r-germ of regular curve is thus identified with a neighborhood of a jet in $\mathcal{F}^r$. One has $\dim \mathcal{F}^r = \dim \mathcal{J}^r = r+2$. There are local coordinates $(x, b_0, \dots, b_r)$ on $\mathcal{F}^r$ defined by the condition that for every jet $j^r_p h \in \mathcal{F}^r$ one has

$$x(j^r_p h) = p, \;\; b_0(j^r_p h) = h(p), \;\; b_1(j^r_p h) = h'(p), \dots, b_r(j^r_p h) = h^{(r)}(p). \quad (8.1)$$

Recall that the *r-jet extension* of a function (curve) is the curve in the jet space $\mathcal{F}^r$ (respectively, $\mathcal{J}^r$) consisting of its r-jets at all points.

Definition 8.1 (see an equivalent definition in [20, pp. 122–123]). Consider the space $\mathcal{F}^r$ equipped with the above coordinates $(x, b_0, \dots, b_r)$. The *Cartan (or contact) distribution* $\mathcal{D}_r$ on $\mathcal{F}^r$ is the field of two-dimensional subspaces in its tangent spaces defined by the system of Pfaffian equations

$$db_0 = b_1 \mathrm{d}x, \;\; db_1 = b_2 \mathrm{d}x, \;\; \dots, \;\; db_{r-1} = b_r \mathrm{d}x. \quad (8.2)$$

For every C^m-smooth surface Σ and every $r \leq m$, the *Cartan (or contact) distribution (plane field) on* $\mathcal{J}^r$, which is also denoted by $\mathcal{D}_r$, is defined by (8.2) locally on its domains identified with open subsets in $\mathcal{F}^r$; the distributions (8.2) defined on intersecting domains V_i, V_j with respect to different charts (x_i, y_i) and (x_j, y_j) coincide and yield a global plane field on $\mathcal{J}^r$.

Remark 8.2 Recall that the r-jet extension of each function (curve) is tangent to the Cartan distribution.

Remark 8.3 The geodesic curvature of a germ of curve is a function of its 2-jet. We will call a 2-jet of curve *κ-nondegenerate,* if the corresponding geodesic curvature is positive. For every $r \geq 2$, an r-jet of curve with κ-nondegenerate 2-jet will be also called κ-nondegenerate. The property of being κ-nondegenerate depends on the Riemannian metric. We denote

$$\mathcal{J}^{r,0} = \mathcal{J}^{r,0}(\Sigma) = \{\text{the } \kappa-\text{nondegenerate } r-\text{jets of curves}\}.$$

The main result of the present section is the following theorem, which immediately implies Theorem 1.19. Proofs of both theorems will be given in Sect. 8.7.

Theorem 8.4 *Let Σ be a two-dimensional surface with a C^6-smooth Riemannian metric. There exists a C^1-smooth line field $\mathcal{P}$ on $\mathcal{J}^{4,0} = \mathcal{J}^{4,0}(\Sigma)$ lying in the Cartan plane field $\mathcal{D}_4$ such that the 4-jet extension of every C^5-smooth curve on Σ with positive geodesic curvature and the string Poritsky property (if any) is a phase curve of the field $\mathcal{P}$.*

Let γ be a germ of curve with the string Poritsky property at a point $O \in \Sigma$. The Poritsky–Lazutkin parameter t on γ is given by already known formula (1.4). We normalize it by additive and multiplicative constants so that $t(O) = 0$ and $\frac{dt}{ds}(O) = \kappa(O)$, see (8.4). We identify points of the curve γ with the corresponding values of the parameter t. Consider the function $L(A, B)$ defined in (1.2). Let $t(A) = a$, $t(B) = a + \tau$. The Poritsky property implies that the function $L(a, a + \tau) = L(0, \tau)$ is independent on a. In particular, the function

$$\Lambda(t) := L(0, t) - L(-t, 0) \tag{8.3}$$

vanishes. For the proof of Theorem 8.4, we show (in the Main Lemma stated in Sect. 8.2) that for every odd $n > 3$ the "differential equation" $\Lambda^{(n+1)}(0) = 0$ is equivalent to an equation saying that the coordinate $b_n = \frac{db_{n-1}}{dx}$ of the n-jet of the curve γ is equal to a function of the other coordinates $(x, b_0, \dots, b_{n-1})$. For $n = 5$, this yields an ordinary differential equation on $\mathcal{J}^{4,0}$ satisfied by the 4-jet extension of the curve γ. It will be represented by a line field contained in $\mathcal{D}_4$.

The proof of the Main Lemma takes the most of the section. For its proof, we study (in Sect. 8.3) two germs of curves γ and $\gamma_{n,b}$ at a point O having contact of order $n \geq 3$. More precisely, they are graphs of functions $y = h(x)$ and $y = h_{n,b}(x)$ such that $h_{n,b}(x) - h(x) = bx^n + o(x^n)$. We show that the corresponding functions $\Lambda(t)$ and $\Lambda_{n,b}(t)$ differ by $c_n bt^{n+1} + o(t^{n+1})$, with c_n being a known constant depending on the second jet of the curve γ; $c_n \neq 0$ for odd $n > 3$. To this end, we consider a local normal chart (x, y) centered at O with x-axis being tangent to γ at O. We compare different quantities related to both curves, all of them being considered as functions of x: the natural parameters, the curvature, etc. In Sect. 8.4, we show that the asymptotic Taylor coefficients of order $(n+1)$ of the functions $L(0, t)$ and $\Lambda(t)$ depend only on the n-jet of the metric at O. We show in Sect. 8.5 that the above Taylor coefficients are analytic functions of the n-jets of metric and the curve (using results of Sects. 8.3 and 8.4). In Sect. 8.6, we show that the degree $n + 1$ coefficient of the function $\Lambda(t)$ is a linear non-homogeneous function in $b_n = b_n(\gamma)$ with coefficients depending on b_j, $j < n$; the coefficient at b_n being expressed via c_n (using results of Sect. 8.3). This will prove the Main Lemma.

8.2. Differential equations in jet spaces and the Main Lemma

Let s denote the natural orienting length parameter of the curve γ, $s(O) = 0$. Let κ be its geodesic curvature considered as a function $\kappa(s)$, and let $\kappa > 0$. We already know that if the curve γ has the string Poritsky property, then its Poritsky–Lazutkin parameter t is expressed as a function of a point $Q \in \gamma$ in terms of the parameter s via formula (1.1), up to constant factor and additive constant, which can be chosen arbitrarily. We normalize it as follows:

$$t(Q) := \kappa^{\frac{1}{3}}(0) \int_0^{s(Q)} \kappa^{\frac{2}{3}}(s) ds \tag{8.4}$$

We can define the parameter t given by (8.4) on any curve γ, not necessarily having the Poritsky property. We identify the points of the curve γ with the corresponding values of the parameter t; thus, $t(O) = 0$.

Remark 8.5 The parameter t on a curve γ given by (8.4) is invariant under rescaling of the metric by constant factor. This follows from the fact that if the norm induced by the metric is multiplied by a constant factor C, then the Levi-Civita connection remains unchanged, the unit tangent vectors $\dot{\gamma}$ are divided by C, and the geodesic curvature $||\nabla_{\dot{\gamma}}\dot{\gamma}||$ of the curve γ considered as a function of a point in γ is divided by C.

Let $G = G(0)$ denote the geodesic tangent to γ at its base point O. We will work in normal coordinates (x, y) centered at O, in which G coincides with the x-axis. For every t let $G(t)$ denote the geodesic tangent to γ at the point t, and let $C(t)$ denote the point of the intersection $G \cap G(t)$.

Let $L(A, B)$ be the function of $A, B \in \gamma$ defined in (1.2). We consider $L(A, B)$ as a function of the corresponding parameters $t(A)$ and $t(B)$, thus,

$$L(0, t) = L(O, \gamma(t)) = |OC(t)| + |C(t)\gamma(t)| - \lambda(0, t), \tag{8.5}$$

where $\lambda(0, t) = \lambda(O, \gamma(t))$ is the length of the arc $O\gamma(t)$ of the curve γ.

The main part of the proof of Theorem 8.4 is the following lemma.

Lemma 8.6 (The Main Lemma). *Let $n \in \mathbb{N}$, $n \geq 5$. Let Σ be a surface equipped with a C^{n+1}-smooth Riemannian metric. Let $V \subset \Sigma$ be a domain equipped with a chart (x, y) where the metric is C^{n+1}-smooth. Let $\mathcal{J}_y^n(V)$ denote the space of those κ-nondegenerate n-jets of curves in V (see Remark 8.3) that are graphs of C^n-smooth functions $\{y = h(x)\}$; thus, it is naturally identified with an open subset $\mathcal{F}_y^n(V) \subset \mathcal{F}^n$. Let $(x, b_0, \ldots, b_n)$ denote the corresponding coordinates on $\mathcal{F}_y^n(V) \simeq \mathcal{J}_y^n(V)$ given by (8.1). Set*

$$J_2 := (x, b_0, b_1, b_2).$$

There exist C^1-smooth functions $\sigma_n(J_2)$ and $P_n(J_2; b_3, \ldots, b_{n-1})$,

$$\sigma_n \neq 0 \text{ for odd } n > 3; \ \sigma_n \equiv 0 \text{ for every even } n > 5, \tag{8.6}$$

such that every jet $J_n = (x, b_0, \ldots, b_n) \in \mathcal{J}_y^n(V)$ extending J_2 satisfies the following statement. Let γ be a C^n-smooth germ of curve representing the jet J_n, and let t be the parameter on γ defined by (8.4). Let $t > 0$, $L(0, t)$ be the same, as in (8.5). The corresponding function $\Lambda(t)$ from (8.3) admits an asymptotic Taylor formula of degree $n + 1$ at 0 of the following type:

$$\Lambda(t) = \sum_{k=3}^{n+1} \widehat{\Lambda}_k t^k + o(t^{n+1}), \tag{8.7}$$

$$\widehat{\Lambda}_{n+1} = \sigma_n(J_2) b_n - P_n(J_2; b_3, \ldots, b_{n-1}). \tag{8.8}$$

Definition 8.7 A *pure* n-jet of curve γ in $\mathbb{R}^2$ is a class of n-jets of curves modulo translations. If $\gamma = \{y = h(x)\}$, then it is identified with the collection of Taylor coefficients of the function $h(x)$ at monomials of degrees from 1 to n. A *pure* n-jet of metric on a planar domain is a class of n-jets of metrics modulo translations. It is identified with the collection of Taylor coefficients of the metric tensor at monomials of degrees from 0 to n.

Addendum to Lemma 8.6. *The function σ_n depends analytically on the pure 1-jet of the metric and the pure 2-jet of the curve. The function P_n depends analytically on the pure n-jet of the metric and the pure $(n-1)$-jet of the curve. The function σ_n is defined by the following formula. Set $u = u(J_2) := (1, b_1) \in T_{(x,b_0)}\Sigma$. Let $w \in T_{(x,b_0)}\Sigma$ denote the image of the vector $\frac{\partial}{\partial y} \in T_{(x,b_0)}\Sigma$ under the Riemannian-orthogonal projection to the line $\mathbb{R}u^{\perp}$. Let $\kappa = \kappa(J_2)$ denote the geodesic curvature of a curve γ representing the jet J_2 (it depends on the pure 2-jet of the curve and the pure 1-jet of the metric). Then, for every odd $n \geq 5$,*

$$\sigma_n(J_2) = \frac{(n-2)(n-3)}{6(n+1)!}||w||(||u||\kappa(J_2))^{-n}. \tag{8.9}$$

Lemma 8.6 and its addendum will be proved in Sect. 8.6.

8.3. Comparison of functions $L(0, t)$ and $\Lambda(t)$ for osculating curves

Let $n \geq 3$. Let Σ be a surface equipped with a Riemannian metric, $O \in \Sigma$. Let us consider normal coordinates (x, y) centered at O. We consider that the metric under question is C^4-smooth in the normal coordinates. This holds automatically for every C^6-smooth metric. Let $b \in \mathbb{R}$, and let $\gamma, \gamma_{n,b} \subset \Sigma$ be two germs of C^n-smooth curves at O with the same $(n-1)$-jet that are tangent to the x-axis at O,

$$\gamma = \{y = h(x)\},\ \gamma_{n,b} = \{y = h_{n,b}(x)\},\ h_{n,b}(x) = h(x) + bx^n + o(x^n),$$

$h, h_{n,b} \in C^n$. Here, $o(x^n)$ is a function tending to zero together with its derivatives up to order n, as $x \to 0$. Their geodesic curvatures at O are equal to the same number $\kappa(O) = |h''(0)| = |h''_{n,b}(0)|$, by (2.1). Without loss of generality, we consider that $\kappa(O) = h''(0) = 1$. One can achieve this by rescaling the norm of the metric by constant factor $\kappa(O)$, see Remark 8.5, and changing sign of the coordinate y.

The main result of the present subsection is the following lemma.

Lemma 8.8 *In the above conditions, let t be the parameter on γ given by (8.4). Let $L(0, t)$, $L_{n,b}(0, t)$ and $\Lambda(t)$, $\Lambda_{n,b}(t)$ be the functions from (8.3) defined for the curves γ and $\gamma_{n,b}$, respectively. For every $t > 0$, one has*

$$L_{n,b}(0,t) - L(0,t) = \frac{(n-2)(n-3)}{12(n+1)}bt^{n+1} + o(t^{n+1}),\ \text{as } t \to 0, \tag{8.10}$$

$$\Lambda_{n,b}(t) - \Lambda(t) = \begin{cases} \frac{(n-2)(n-3)}{6(n+1)}bt^{n+1} + o(t^{n+1}), & \text{if } n \text{ is odd}, \\ o(t^{n+1}), & \text{if } n \text{ is even}. \end{cases} \tag{8.11}$$

For the proof of Lemma 8.8, we first compare the natural parameters $s(x)$, $s_{n,b}(x)$ centered at O, the curvatures $\kappa(x)$, $\kappa_{n,b}(x)$ and the parameters $t(x)$, $t_{n,b}(x)$ given by (8.4) for the curves γ and $\gamma_{n,b}$ as functions of x. We also compare the corresponding inverse functions $x = x(t)$ and $x = x_{n,b}(t)$ as functions of t, see Proposition 8.9 below. Afterwards we prove formula (8.10) using the above-mentioned comparison results and the results of Sect. 2. Then, we deduce (8.11).

Proposition 8.9 *As $x \to 0$ (or equivalently, $t \to 0$), one has*

$$t(x) \simeq t_{n,b}(x) \simeq x, \quad x(t) \simeq t \simeq h'(x(t)), \tag{8.12}$$

$$s_{n,b}(x) - s(x) = \frac{n}{n+1} b x^{n+1} + o(x^{n+1}), \tag{8.13}$$

$$\kappa_{n,b}(x) - \kappa(x) = n(n-1) b x^{n-2} + o(x^{n-2}), \tag{8.14}$$

$$t_{n,b}(x) - t(x) = \frac{2n}{3} b x^{n-1} + o(x^{n-1}), \tag{8.15}$$

$$x_{n,b}(t) - x(t) = -\frac{2n}{3} b t^{n-1} + o(t^{n-1}). \tag{8.16}$$

Proof Formulas (8.12) follow from (8.4), since $\kappa(O) = h''(0) = 1$. In the parametrizations $\gamma = \gamma(x)$, $\gamma_{n,b} = \gamma_{n,b}(x)$, one has

$$s(x) = \int_0^x ||\dot{\gamma}(u)||\mathrm{d}u, \ s_{n,b}(x) = \int_0^x ||\dot{\gamma}_{n,b}(u)||\mathrm{d}u. \tag{8.17}$$

We claim that

$$||\dot{\gamma}_{n,b}(x)|| - ||\dot{\gamma}(x)|| = nbx^n + o(x^n). \tag{8.18}$$

Indeed, let us identify the tangent spaces $T_{(x,y)}\Sigma$ at different points (x, y) by translations. One has $\dot{\gamma}(x), \dot{\gamma}_{n,b}(x) = (1, x + o(x))$,

$$v(x) := \dot{\gamma}_{n,b}(x) - \dot{\gamma}(x) = (0, nbx^{n-1} + o(x^{n-1})) : \tag{8.19}$$

$h'(x) \simeq x$, since $h''(0) = \kappa(O) = 1$, by assumption. The metric has trivial 1-jet at the base point O. Therefore, the difference of metric tensors at the $O(x^n)$-close points $\gamma(x)$, $\gamma_{n,b}(x)$, which are $O(x)$-close to O, is $O(x^{n+1})$. Hence, it suffices to prove (8.18) for the vector $\dot{\gamma}_{n,b}(x)$ being translated to the point $\gamma(x)$. The Euclidean angle between the vectors $v(x)$ and $\dot{\gamma}(x)$ is $\frac{\pi}{2} - x + o(x)$, by (8.19). Therefore, the angle between them in the metric of the tangent plane $T_{\gamma(x)}\Sigma$ has the same asymptotics. Hence,

$$||\dot{\gamma}_{n,b}(x)||^2 = ||v(x) + \dot{\gamma}(x)||^2 = ||\dot{\gamma}(x)||^2 + 2nbx^n + o(x^n),$$

by Cosine Theorem and since $||v(x)||^2 = O(x^{2n-2}) = O(x^{n+1})$ $(n \geq 3)$. The latter formula together with the obvious formula $||\dot{\gamma}(x)|| = 1 + O(x)$ imply (8.18), which together with (8.17) implies (8.13).

Let us prove (8.14). The Christoffel symbols at the $O(x^n)$-close points $\gamma(x)$ and $\gamma_{n,b}(x)$ are $O(x^n)$-close, as in the above discussion. Therefore, the difference $\kappa_{n,b}(x) - \kappa(x)$ is equal up to $O(x^n)$ to the same difference, where each curvature is calculated in the metric (Christoffel symbols) of the point $\gamma(x)$. The difference of the Christoffel parts of the curvatures, which are quadratic in the vectors $\frac{1}{||\dot{\gamma}(x)||}\dot{\gamma}(x)$, $\frac{1}{||\dot{\gamma}_{n,b}(x)||}\dot{\gamma}_{n,b}(x)$, is $O(||v(x)||) = O(x^{n-1})$, by (8.18). The difference of their second derivative terms is equal

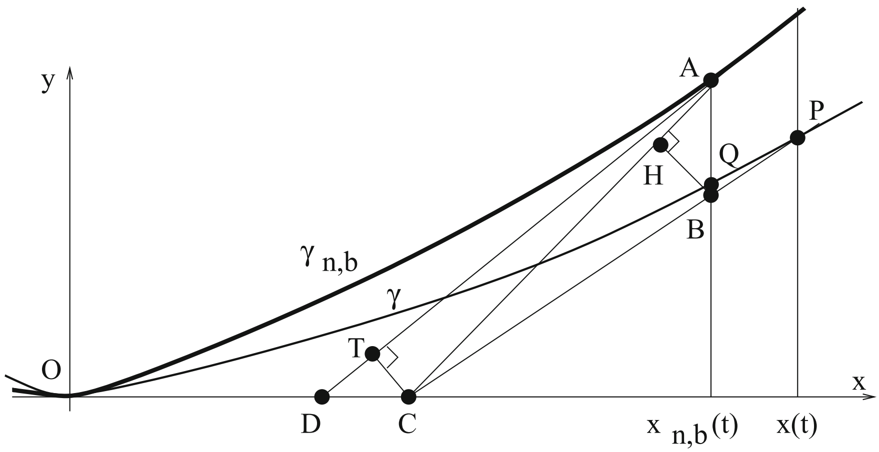


FIGURE 7. Auxiliary geodesics for calculation of the asymptotic of the difference $L_{n,b}(0,t) - L(0,t)$

to $h''_{n,b}(x) - h''(x) + O(x^n) = n(n-1)bx^{n-2} + o(x^{n-2})$, by definition and (8.18). This together with the above discussion implies (8.14).

Let us prove (8.15). One has

$$\begin{aligned} t_{n,b}(x) - t(x) &= \int_0^x (\kappa_{n,b}^{\frac{2}{3}}(u)||\dot\gamma_{n,b}(u)|| - \kappa^{\frac{2}{3}}(u)||\dot\gamma(u)||)du \\ &= \int_0^x (\kappa_{n,b}^{\frac{2}{3}}(u) - \kappa^{\frac{2}{3}}(u))||\dot\gamma(u)||du + O(x^n), \end{aligned}$$

by definition and (8.18). The latter right-hand side is asymptotic to $\frac{2}{3}\int_0^x n(n-1)bu^{n-2}\mathrm{d}u = \frac{2n}{3}bx^{n-1}$, by (8.14) and since $\kappa(0) = 1$. This proves (8.15).

Formula (8.16) follows from (8.15). Proposition 8.9 is proved.

In the proof of formula (8.10), we use the following notations:

$$\begin{aligned} &P = P(t) := \gamma(t),\ Q = Q(t) := (x_{n,b}(t), h(x_{n,b}(t))) \in \gamma,\ A = A(t) := \gamma_{n,b}(t), \\ &G(t) := \text{ the geodesic tangent to } \gamma \text{ at } P,\ G(0) = \text{ the } x-\text{axis}, \\ &C = C(t) := G(t) \cap G(0),\ V = V(t) := \{x = x_{n,b}(t)\},\ B = B(t) := G(t) \cap V, \\ &G_{n,b}(t) := \text{ the geodesic tangent to } \gamma_{n,b} \text{ at } A,\ D = D(t) := G_{n,b}(t) \cap G(0), \end{aligned}$$

see Fig. 7. By definition, $Q = Q(t) = \gamma \cap V$.

In what follows for any two points $E, F \in \Sigma$ close to O, the length of the geodesic segment connecting F to E will be denoted by $|EF|$. By definition,

$$L(0,t) = |OC| + |CP| - \lambda(O,P),\ L_{n,b}(0,t) = |OD| + |DA| - \lambda(O,A). \quad (8.20)$$

Recall that $\lambda(O,A)$, $\lambda(O,P)$ are lengths of arcs OA and OP of the curves $\gamma_{n,b}$ and γ, respectively. Set

$$\begin{aligned} &L_1 = L_1(t) := |OC| + |CB| - \lambda(O,Q),\ L_2 = L_2(t) := |OC| + |CA| - \lambda(O,A), \\ &\Delta_1 = \Delta_1(t) := L_1(t) - L(0,t) = \lambda(Q,P) - |BP|, \end{aligned}$$

$$\Delta_2 = \Delta_2(t) := L_2(t) - L_1(t),$$
$$\Delta_3 = \Delta_3(t) := L_{n,b}(0,t) - L_2(t):$$
$$L_{n,b}(0,t) - L(0,t) = \Delta_1(t) + \Delta_2(t) + \Delta_3(t). \tag{8.21}$$

In what follows, we find asymptotics of each Δ_j.

Proposition 8.10 *One has*

$$\Delta_1(t) = O(t^{2n-1}) = O(t^{n+2}) \text{ whenever } n \geq 3. \tag{8.22}$$

Proof In the curvilinear triangle QPB with $QP \subset \gamma$, PB being geodesic and QB vertical segment, one has $|PB| = O(x_{n,b}(t) - x(t)) = O(t^{n-1})$, by (8.16). Its angle at B is $\frac{\pi}{2} + O(t)$. Therefore, by (2.14),

$$\Delta_1 = \lambda(Q,P) - |PB| = O(|PB|^3) + O(t|PB|^2) = O(t^{3n-3}) + O(t^{2n-1}).$$

The latter right-hand side is $O(t^{2n-1}) = O(t^{n+2})$, since $n \geq 3$.

Proposition 8.11 *One has*

$$\Delta_2(t) = \frac{b}{n+1} t^{n+1} + o(t^{n+1}). \tag{8.23}$$

Proof By definition,

$$\begin{aligned} \Delta_2(t) &= |OC| + |CA| - \lambda(O,A) - (|OC| + |CB| - \lambda(O,Q)) \\ &= (|CA| - |CB|) - (\lambda(O,A) - \lambda(O,Q)), \end{aligned} \tag{8.24}$$

$$\lambda(O,A) - \lambda(O,Q) = s_{n,b}(x_{n,b}(t)) - s(x_{n,b}(t)) = \frac{nbt^{n+1}}{n+1} + o(t^{n+1}), \tag{8.25}$$

by (8.13) and (8.12). To find the asymptotics of the difference $|CA| - |CB|$, let us consider the height denoted by BH of the geodesic triangle ABC, which splits it into two right triangles ABH and CBH, see Fig. 7. We use the following asymptotic formula for lengths of their sides:

$$|AB| \simeq bt^n + o(t^n) \simeq |BH|, \tag{8.26}$$

$$|CB| \simeq |CP| \simeq |CA| \simeq \frac{t}{2} \tag{8.27}$$

$$|AH| \simeq bt^{n+1} + o(t^{n+1}) \simeq |AC| - |BC|. \tag{8.28}$$

Proof of (8.26) The Euclidean distance in the coordinates (x,y) between the points A and Q is $bx^n_{n,b}(t) + o(x^n_{n,b}(t)) = bt^n + o(t^n)$, by construction. Therefore, the distance between them in the metric g is asymptotic to the same quantity, since g is Euclidean on $T_O\Sigma$. The Euclidean distance between the points Q and B is of order $O((x(P) - x(B))^2) \simeq O(t^{2(n-1)}) = O(t^{n+1})$, by (8.16) and since $n \geq 3$: $2(n-1) \geq n+1$ for $n \geq 3$. The two latter statements together imply that $|AB| = bt^n + o(t^n)$; this is the first asymptotics in (8.26).

In the proof of the second asymptotics in (8.26) and in what follows, we use the two next claims.

Claim 1. *The azimuths of the tangent vectors of the geodesic arcs CA, CP, DA at all their points are uniformly asymptotically equivalent to $t = t(P)$, as $t \to 0$.*

Proof Let us prove the above statement for the geodesic arc CP; the proof for the arcs CA and DA is analogous. The slope of the tangent vector to the curve γ at the point P is asymptotic to $x(P) = x(t) \simeq t$, and it is equal to the slope of the tangent vector of the geodesic CP at P. On the other hand, let us apply formula (2.5) to the geodesic arc $\alpha = CP$: its right-hand side is $O(t)$. The length of the arc CP is $O(t)$. Hence, the difference between the azimuths of tangent vectors at any two points of the geodesic arc CP is $O(t^2)$. This proves the claim.

Claim 2. *The angle A of the geodesic triangle ABH is asymptotic to $\frac{\pi}{2} - t + o(t)$. Its angle B is asymptotic to $t + o(t)$, and $|AH| \simeq t|AB|$.*

Proof The first statement of the claim follows from Claim 1 applied to CA and the fact that the slopes of the tangent vectors to the geodesic arc BA are uniformly $O(|BA|) = O(t^n)$-close to $\frac{\pi}{2}$. This follows from the second formula in (2.13) and formula (2.5) applied to the geodesic arc BA. The second statement of the claim follows from the first one and (2.12).

One has $|AB| \simeq |HB|$, by Claim 2 and (2.12). This yields the second asymptotics in (8.26). Formula (8.26) is proved.

Proof of (8.27) The asymptotics $|CP| \simeq \frac{x(P)}{2} \simeq \frac{t}{2}$ follows from Claim 1 and the fact that the height of the point P over the x-axis is asymptotic to $\frac{x^2(P)}{2} \simeq \frac{t^2}{2}$. The other asymptotics in (8.27) follow from the above one, formula (8.26) and the fact that $|BP| = O(t^{n-1})$ (follows from (8.16)).

Proof of (8.28) The first asymptotic formula in (8.28) follows from (8.26) and the last statement of Claim 2. In the proof of the second formula in (8.28) we use the following claim.

Claim 3. *The angle $\phi := \angle BCH$ equals $2bt^{n-1} + o(t^{n-1})$.*

Proof The triangle BCH has right angle at H, $|BH| = bt^n + o(t^n)$, $|BC| \simeq \frac{t}{2}$, by (8.26) and (8.27). Hence, $\phi \simeq |BH|/\frac{t}{2} = 2bt^{n-1} + o(t^{n-1})$.

Now let us prove the second asymptotic formula in (8.28). One has

$$|BC| - |HC| \simeq \frac{1}{2}|BC|\phi^2,$$

by formula (2.12) applied to the family of triangles BCH. The right-hand side in the latter formula is $b^2t^{2n-1} + o(t^{2n-1}) = O(t^{n+2})$, by (8.27) and Claim 3 and since $2n - 1 \geq n + 2$ for $n \geq 3$. Thus,

$$\begin{aligned} &|BC| - |HC| = O(t^{n+2}), \\ &|AC| - |BC| = (|HC| - |BC|) + |AH| = |AH| + O(t^{n+2}) = bt^{n+1} + o(t^{n+1}), \end{aligned} \tag{8.29}$$

by the first formula in (8.28) proved above. Formula (8.28) is proved.

Substituting formulas (8.25) and (8.28) to (8.24) yields

$$\Delta_2(t) = bt^{n+1} - \frac{n}{n+1}bt^{n+1} + o(t^{n+1}) = \frac{b}{n+1}t^{n+1} + o(t^{n+1}).$$

Proposition 8.11 is proved.

Proposition 8.12 *One has*

$$\Delta_3(t) = \frac{n-6}{12} b t^{n+1} + o(t^{n+1}). \tag{8.30}$$

Proof Recall that

$$\begin{aligned}\Delta_3(t) &= L_{n,b}(0,t) - L_2(t) = |OD| + |DA| - \lambda(O,A) - (|OC| + |CA| - \lambda(O,A)) \\ &= |DA| - (DC + |CA|).\end{aligned} \tag{8.31}$$

Here, DC is the "oriented length" $DC := |OC| - |OD|$.

Let CT denote the height of the geodesic triangle DCA. To find an asymptotic formula for the right-hand side in (8.31), we first find asymptotics of the length of the height CT and the angle $\angle DAC$.

Claim 4. *Let $\alpha := \angle DAC$ denote the oriented angle between the geodesics AD and AC: it is said to be positive, if D lies between O and C, as at* Fig. 7. *One has $\alpha = \frac{6-n}{3} b t^{n-1} + o(t^{n-1})$.*

Proof Consider the following tangent lines of the geodesic arcs AD, AC, BC, CP and the curve γ:

$$\begin{aligned}&\ell_1 := T_A AD = T_A \gamma_{n,b}, \ \ell_2 := T_A AC, \ \ell_3 := T_B BC, \\ &\ell_4 := T_Q \gamma, \ \ell_5 := T_P CP = T_P \gamma.\end{aligned}$$

We orient all these lines "to the right". One has

$$\alpha \simeq \mathrm{az}(\ell_2) - \mathrm{az}(\ell_1), \tag{8.32}$$

by definition and since the Riemannian metric at the point A written in the normal coordinates (x, y) tends to the Euclidean one, as $t \to 0$. Let us find asymptotic formula for the above difference of azimuths by comparing azimuths of appropriate pairs of lines $\ell_1, \dots, \ell_5$. One has

$$\mathrm{az}(\ell_4) - \mathrm{az}(\ell_1) = -nbt^{n-1} + o(t^{n-1}),$$

since the above azimuth difference is asymptotically equivalent to the difference of the derivatives of the functions $h(x)$ and $h_{n,b}(x) = h(x) + bx^n + o(x^n)$ at the same point $x = x(B) \simeq t$: hence, to $-nbx^{n-1} + o(x^{n-1})$. One has

$$\mathrm{az}(\ell_5) - \mathrm{az}(\ell_4) \simeq h'(x(t)) - h'(x_{n,b}(t)) \simeq x(t) - x_{n,b}(t) = \frac{2n}{3} b t^{n-1} + o(t^{n-1}),$$

by (8.16) and since the function $h'(x) \simeq x$ has unit derivative at 0,

$$\mathrm{az}(\ell_3) - \mathrm{az}(\ell_5) = O(t(x(B) - x(P))) = O(t(x_{n,b}(t) - x(t))) = O(t^n),$$

by (2.5) and (8.16),

$$\mathrm{az}(\ell_2) - \mathrm{az}(\ell_3) \simeq \angle BCA = 2bt^{n-1} + o(t^{n-1}),$$

by (2.11), (2.5), (8.28) and Claim 3. The right-hand sides of the above asymptotic formulas for azimuth differences are all of order t^{n-1}, except for one, which is $O(t^n)$. Summing all of them yields the statement of Claim 4:

$$\alpha \simeq \mathrm{az}(\ell_2) - \mathrm{az}(\ell_1) = \frac{6-n}{3} b t^{n-1} + o(t^{n-1}).$$

Claim 5. *In the right triangle*[3] CDT $\angle TDC \simeq t$, $CT = \frac{6-n}{6}bt^n + o(t^n)$,

$$CD \simeq DT = \frac{6-n}{6}bt^{n-1} + o(t^{n-1}), \quad CD - DT = \frac{6-n}{12}bt^{n+1} + o(t^{n+1}). \tag{8.33}$$

Proof The angle asymptotics follows from Claim 1. The length asymptotics for the side CT is found via the adjacent right triangle ACT, from the formula $CT \simeq AC\angle CAT$ after substituting $\alpha = \angle CAT = \frac{6-n}{3}bt^{n-1} + o(t^{n-1})$ (Claim 4) and $AC \simeq \frac{t}{2}$, see (8.27). This together with formula (2.12) applied to the right triangle CDT and the asymptotics $\angle CDT \simeq t$ implies (8.33).

Now let us prove formula (8.30). Recall that

$$\Delta_3(t) = |DA| - (DC + |CA|) = (DT - DC) + (|AT| - |AC|), \tag{8.34}$$

see (8.31). One has $DT - DC = \frac{n-6}{12}bt^{n+1} + o(t^{n+1})$, by (8.33); $|AT| - |AC| = O(t^{n+2})$, by (2.12) and Claim 4, analogously to the proof of formula (8.29). Substituting the two latter formulas to (8.34) yields to (8.30). Proposition 8.12 is proved.

Proof of Lemma 8.8 Let us prove formula (8.10). Summing up formulas (8.22), (8.23), (8.30) and substituting their sum to (8.21) yields to (8.10):

$$L_{b,n}(0,t) - L(0,t) = \Delta_1(t) + \Delta_2(t) + \Delta_3(t) = \frac{b}{n+1}t^{n+1} + \frac{n-6}{12}bt^{n+1} + o(t^{n+1})$$
$$= (\frac{1}{n+1} + \frac{n-6}{12})bt^{n+1} + o(t^{n+1}) = \frac{(n-2)(n-3)}{12(n+1)}bt^{n+1} + o(t^{n+1}).$$

Let us prove formula (8.11). Consider the points of the curves γ and $\gamma_{n,b}$ with $x < 0$. Taking them in the coordinates $(\hat{x}, y)$, $\hat{x} := -x$ results in multiplying the coefficient b by $(-1)^n$. This implies that for every $t > 0$

$$L_{n,b}(-t,0) - L(-t,0) = (-1)^n \frac{(n-2)(n-3)}{12(n+1)}bt^{n+1} + o(t^{n+1}). \tag{8.35}$$

Thus, for odd (even) n, the main asymptotic terms in (8.35) and (8.10) are opposite (respectively, coincide). Hence, in the expression

$$\Lambda_{n,b}(t) - \Lambda(t) = (L_{n,b}(0,t) - L(0,t)) - (L_{n,b}(-t,0) - L(-t,0))$$

they are added (cancel out), and we get (8.11). Lemma 8.8 is proved.

8.4. Dependence of functions $L(0,t)$ and $\Lambda(t)$ on the metric

Here, we prove the following lemma, which shows that the $(n+1)$-jets of the quantities $L(0,t)$ and $\Lambda(t)$ depend only on the n-jet of the metric.

Lemma 8.13 *Let $n \geq 5$, Σ be a two-dimensional surface. Let $O \in \Sigma$, $\gamma \subset \Sigma$ be a germ of C^n-smooth curve at O with positive geodesic curvature. Let g and $\widetilde{g}$ be two C^{n+1}-smooth Riemannian metrics on Σ having the same n-jet at O: $\widetilde{g}(q) - g(q) = o(\mathrm{dist}^n(q,O))$, as $q \to O$. Then, the differences $L_{\widetilde{g}}(0,t) - L_g(0,t)$, $\Lambda_{\widetilde{g}}(t) - \Lambda_g(t)$ of quantities $L(0,t)$ and $\Lambda(t)$ defined by the metrics $\widetilde{g}$ and g are $o(t^{n+1})$.*

[3] *We treat the lengths of sides of the triangle CDT as oriented lengths (without module sign): we take them with the sign equal to $\mathrm{sign}(\alpha)$, where α is the same, as in Claim 4.*

Proof Let s, $\widetilde{s}$, t, $\widetilde{t}$, κ, and $\widetilde{\kappa}$ denote the natural and Lazutkin parameters centered at O, see (8.4), and the geodesic curvature of the curve γ defined by the metrics g and $\widetilde{g}$, respectively. One has $\kappa(O) = \widetilde{\kappa}(O)$, since $n \geq 3$. Let us rescale the metrics by the same constant factor so that $\kappa(O) = 1$. Fix C^n-smooth coordinates (x, y) centered at O so that the x-axis is tangent to the curve γ and $||\frac{\partial}{\partial x}|| = 1$ at O. Consider x as a local parameter on γ. We consider the above quantities as functions of x; $s(0) = \widetilde{s}(0) = t(0) = \widetilde{t}(0) = 0$.

Let $x(t)$, $\widetilde{x}(t)$ denote the functions inverse to $t(x)$ and $\widetilde{t}(x)$, respectively. Let $\gamma(t)$ and $\widetilde{\gamma}(t)$ denote the points of the curve γ with x-coordinates $x(t)$ and $\widetilde{x}(t)$, respectively. Let now $s(t)$ and $\widetilde{s}(t)$ denote the natural length parameters of the metrics g and $\widetilde{g}$, now considered as functions of the parameter t defined by the metric under question (g or $\widetilde{g}$).

Proposition 8.14 *One has* $t \simeq x \simeq \widetilde{t} \simeq s \simeq \widetilde{s}$,

$$\widetilde{s}(x) - s(x) = o(x^{n+1}),\ \widetilde{\kappa}(x) - \kappa(x) = o(x^{n-1}),\ \widetilde{t}(x) - t(x) = o(x^n), \tag{8.36}$$

$$\widetilde{x}(t) - x(t) = o(t^n),\ \operatorname{dist}(\gamma(t), \widetilde{\gamma}(t)) = o(t^n), \tag{8.37}$$

$$\widetilde{s}(t) - s(t) = o(t^n),\ \widetilde{s}'(t) - s'(t) = o(t^{n-1}). \tag{8.38}$$

Proof The asymptotic equivalences follow from (8.4). The first formula in (8.36) is obvious. The second one holds by definition and since the Christoffel symbols of the two metrics differ by a quantity $o(x^{n-1})$. The third formula follows from the second one. Formula (8.37) follows from the third formula in (8.36). Formula (8.38) follows from (8.36) and (8.37).

Fix a small value $t \in \mathbb{R}$, say, $t > 0$. Set

$$P = \gamma(t),\ A = \widetilde{\gamma}(t).$$

Let C ($\widetilde{C}$) be the point of intersection of the g-(respectively, $\widetilde{g}$-) geodesics $G(P)$, $G(O)$ tangent to γ at P and O. Let D ($\widetilde{D}$) be the analogous points of intersection of the geodesics tangent to γ at A and O. See Fig. 8a). The distance (arc length) between points E and F in a metric h will be denoted by $|EF|_h$ (respectively, $\lambda_h(E, F)$). One has

$$L_g(0, t) = |OC|_g + |CP|_g - \lambda_g(O, P),\ L_{\widetilde{g}}(0, t) = |O\widetilde{D}|_{\widetilde{g}} + |\widetilde{D}A|_{\widetilde{g}} - \lambda_{\widetilde{g}}(O, A),$$

by definition. Set

$$\Delta_1(t) := |OC|_g + |CP|_g - |OD|_g - |DA|_g - (\lambda_g(O, P) - \lambda_g(O, A)); \tag{8.39}$$

$$\Delta_2(t) := (|OD|_g - |OD|_{\widetilde{g}}) + (|DA|_g - |DA|_{\widetilde{g}}) - (\lambda_g(O, A) - \lambda_{\widetilde{g}}(O, A)); \tag{8.40}$$

$$\Delta_3(t) := (|OD|_{\widetilde{g}} - |O\widetilde{D}|_{\widetilde{g}}) + (|DA|_{\widetilde{g}} - |\widetilde{D}A|_{\widetilde{g}}). \tag{8.41}$$

One has

$$L_g(0, t) - L_{\widetilde{g}}(0, t) = \Delta_1 + \Delta_2 + \Delta_3. \tag{8.42}$$

Claim 1. *One has* $\Delta_1(t) = o(t^{n+1})$.

Proof Let us introduce the point B of intersection of the g-geodesic PC with the vertical line through A, see Fig. 8a): $x(B) = \widetilde{x}(t)$. One has

$$\Delta_1 = (|OC|_g + |CB|_g - |OD|_g - |DA|_g) + (BP_g - \widehat{\lambda}_g(A, P)). \qquad (8.43)$$

Here, BP_g and $\widehat{\lambda}_g(A, P)$ are the corresponding oriented lengths, which are positive if and only if A lies between O and P on the curve γ. Consider the curvilinear triangle APB formed by the arc AP of the curve γ, the g-geodesic PB and the vertical segment BA. Its sides AP and BA have g-length $o(t^n)$, by definition and (8.37). Its angle B is $\frac{\pi}{2} + O(\widetilde{x}(t)) = \frac{\pi}{2} + O(t)$, as in Claim 2 in Sect. 8.3 (Here and below, in order to use arguments from proofs of Claims 2 and 4 from Sect. 8.3, we use C^4-smoothness of the metric g in g-normal coordinates centered at O. This follows from its C^{n+1}-smoothness, $n \geq 5$.). This together with (2.14) implies that the second bracket in (8.43) is $o(t^{n+1})$. Let us prove the same statement for the first bracket. It is equal to

$$DC_g + |CA|_g - |DA|_g + (|CB|_g - |CA|_g) = DC_g + |CA|_g - |DA|_g + o(t^{n+1}), \qquad (8.44)$$

since $||CB|_g - |CA|_g| \leq |BA| = O((x(P) - x(B))^2) = o(t^{n+1})$. Here, DC_g is the oriented length $|OC|_g - |OD|_g$. One has

$$DC_g + |CA|_g - |DA|_g = o(t^{n+1}). \qquad (8.45)$$

Indeed, consider the height CT of the triangle ADC, which splits it into two right triangles. One has $\angle CAD = O(x(A) - x(P)) = o(t^n)$, as in the proof of Claim 4 in the previous subsection. This together with right triangle arguments using (2.12) analogous to those from the proof of Claim 5 (Sect. 8.3) implies (8.45). Substituting (8.45) to (8.44) and then substituting everything to (8.43) yields $\Delta_1(t) = o(t^{n+1})$. Claim 1 is proved.

Claim 2. *One has* $\Delta_2(t) = o(t^{n+1})$.

Proof All the points in (8.40) are $O(t)$-close to O. The g- and $\widetilde{g}$-distances between any two points (which will be denoted by E and F) differ by a quantity $o(t^{n+1})$. Indeed, the $\widetilde{g}$-length of the g-geodesic segment EF differs from its g-length by $o(t^{n+1})$, since the metrics differ by $o(t^n)$. The distance $|EF|_{\widetilde{g}}$ is no greater than the latter $\widetilde{g}$-length, and hence, no greater than $|EF|_g + o(t^{n+1})$. Applying the same arguments to interchanged metrics yields that the above distances differ by $o(t^{n+1})$. Similarly, $\lambda_g(O, A) - \lambda_{\widetilde{g}}(O, A) = o(t^{n+1})$. This proves the claim.

Let H and M denote the points in the $\widetilde{g}$-geodesics OD and DA, respectively, that are $\widetilde{g}$-closest to $\widetilde{D}$: $\widetilde{D}H \perp_{\widetilde{g}} OD$; $\widetilde{D}M \perp_{\widetilde{g}} DA$; see Fig. 8b).

Claim 3. *One has* $|\widetilde{D}H|_{\widetilde{g}} = o(t^{n+1})$, $|\widetilde{D}M|_{\widetilde{g}} = o(t^{n+1})$.

Proof Let EF_h denote the geodesic EF in the metric h. The $\widetilde{g}$-geodesic $O\widetilde{D}_{\widetilde{g}}$ is tangent to the g-geodesic OD_g at O. The metrics g and $\widetilde{g}$ have the same n-jet at O. Therefore, their Christoffel symbols have the same $(n-1)$-jet. Consider the geodesics OD_g, $O\widetilde{D}_{\widetilde{g}}$ as solutions of the differential equations for geodesics of the metrics g, $\widetilde{g}$ with the same initial condition $\dot{\gamma}(0)$ at O. Differentiating the above equations successively $(n-1)$ times we get that

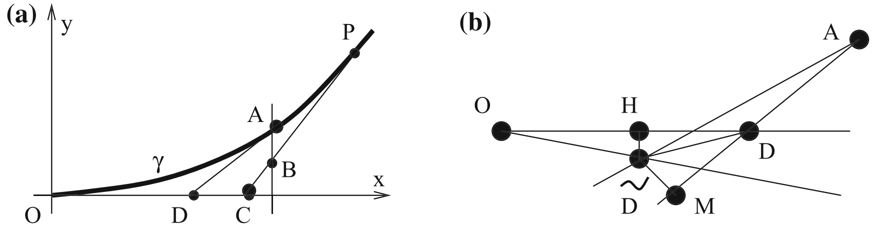


FIGURE 8. The curve γ, points P, A, C, D, B (**a**). The points $\widetilde{D}$, H, M; case (2) (**b**))

these solutions have the same $(n+1)$-jet, by the above statement on the Christoffel symbols. This implies that $\mathrm{dist}(D, O\widetilde{D}_{\widetilde{g}}) = o(t^{n+1})$. Therefore, the $\widetilde{g}$-geodesic $O\widetilde{D}$ should be turned at O by an angle of order $o(t^n)$ in order to hit the point D, by the above statement and since $|OD|_{\widetilde{g}} \simeq \frac{t}{2}$, as in (8.27). This implies that the points in $O\widetilde{D}_{\widetilde{g}}$ lying on a distance of order $O(t)$ from O are $o(t^{n+1})$-close to the geodesic $OD_{\widetilde{g}}$. This proves the statement of the claim for the distance $|\widetilde{D}H|_{\widetilde{g}}$. The proof for $|\widetilde{D}M|_{\widetilde{g}}$ is analogous. Namely, first we rescale the metric $\widetilde{g}$ by constant factor $1 + o(t^n)$ in order to achieve that a vector tangent to γ at A has equal g- and $\widetilde{g}$-norms. Then the g- and $\widetilde{g}$-geodesics AD, $A\widetilde{D}$ tangent to γ at A, being lifted to the tangent bundle, are phase curves of the vector fields defining g- and $\widetilde{g}$-geodesic flows, with the same initial condition at A. The latter vector fields differ by $o(t^{n-1})$ on the $\frac{1}{2}$-neighborhood (in the norm of the metric g) of the g-unit tangent bundle over the $6|t|$-neighborhood of O. This implies that the solutions of the corresponding second order equations on geodesics differ by $o(t^{n+1})$ on the time segment $[-2t, 2t]$ and $|\widetilde{D}M|_{\widetilde{g}} = o(t^{n+1})$, by a modified version of the above argument for $|\widetilde{D}H|_{\widetilde{g}}$.

Claim 4. *One has* $\Delta_3(t) = o(t^{n+1})$.

Proof All the distances below are measured in the metric $\widetilde{g}$. One has

$$|O\widetilde{D}| - |OH| = O(\frac{|\widetilde{D}H|^2}{|O\widetilde{D}|}) = o(t^{2n+1}) = o(t^{n+1}), \tag{8.46}$$

$$|A\widetilde{D}| - |AM| = O(\frac{|\widetilde{D}M|^2}{|A\widetilde{D}|}) = o(t^{2n+1}) = o(t^{n+1}), \tag{8.47}$$

by (2.12) (applied to the right $\widetilde{g}$-triangles $O\widetilde{D}H$ and $A\widetilde{D}M$) and Claim 3,

$$|OD| - |OH| = \pm|DH|, \ |AD| - |AM| = \pm|DM|, \tag{8.48}$$

see the cases of signs (which do not necessarily coincide) below. Taking sum of equalities (8.48) and its difference with the sum of (8.46), (8.47) yields

$$\Delta_3(t) = (\pm)|DH| \pm |DM| + o(t^{n+1}). \tag{8.49}$$

Case (1). In the right triangle $D\widetilde{D}H$, the angle D is bounded from below (along some sequence of parameter values t converging to 0). Then, the same statement holds in the right triangle $\widetilde{D}MD$, since the angle between the geodesics DA and OD tends to 0 as $O(t)$. This implies that $|DH| = O(|\widetilde{D}H|) = o(t^{n+1})$, and $|DM| = O(|\widetilde{D}M|) = o(t^{n+1})$, by Claim 3. This together with (8.49) implies Claim 4 (along the above sequence)

Case (2). In the right tringle $D\widetilde{D}H$, the angle D tends to zero along some sequence of parameter values t converging to 0, see Fig. 8b). Then, the same holds in $\widetilde{D}MD$. In this case, the signs in (8.49) are different. For example, if H lies between O and D, then the angle $\angle\widetilde{D}DA$ is obtuse and D lies between M and A. The opposite case is treated analogously. Let us denote

$$\alpha(t) := \angle\widetilde{D}DH,\ \beta(t) := \angle\widetilde{D}DM;\ \alpha(t), \beta(t) \to 0 \text{ as } t \to 0.$$

Applying (2.12) to the above right triangles together with Claim 3 yields

$$|\widetilde{D}D| - |DH| = O(\alpha(t)|\widetilde{D}H|) = o(t^{n+1}),$$
$$|\widetilde{D}D| - |DM| = O(\beta(t)|\widetilde{D}M|) = o(t^{n+1}).$$

Hence, $|DH| - |DM| = o(t^{n+1})$. This together with (8.49) implies the asymptotics of Claim 4 (along the above sequence). Claim 4 is proved.

Claims 1, 2 and 4 together with (8.42) imply the statement of Lemma 8.13 on the function L. In its turn, it implies the same statement on Λ.

8.5. Taylor coefficients of $\Lambda(t)$: analytic dependence on jets

Lemma 8.15 *Let (x, y) be coordinates on a neighborhood of a point $O \in \Sigma$, $n \geq 5$. Let a metric on Σ be C^{n+1}-smooth in the coordinates (x, y), and let γ be a germ of C^n-smooth curve on Σ at O. Then, the corresponding functions $L(0, t)$, $\Lambda(t)$ are $O(t^3)$. They admit asymptotic Taylor expansions up to t^{n+1}. Their coefficients at t^{n+1} are analytic functions of the pure n-jets of the metric and the curve γ.*

Proof The asymptotics $L(0, t), \Lambda(t) = O(t^3)$ follows from Theorem 1.16.

Case (1): the curve γ and the metric are analytic. Consider the metric and the curve with variable Taylor coefficients of orders up to n; the other, higher Taylor coefficients are fixed. Consider $L(0, t)$ and $\Lambda(t)$ as functions in t and in the latter variable Taylor coefficients. They are analytic on the product of a small complex disk centered at 0 with coordinate t and a domain in the space of collections of the above Taylor coefficients. In more detail, complexifying everything, we get that $L(0, t)$ has a well-defined holomorphic extension to complex domain. (The complexified lengths of segments in the definition of the function $L(0, t)$ become integrals of appropriate holomorphic forms along paths.) Well-definedness follows from the fact that through each point C in a complex neighborhood of the real curve γ there are two complex geodesics tangent to its complexification. This follows by quadraticity of tangencies (non-vanishing of geodesic curvature) and Implicit Function Theorem. Analytic extendability to the locus $\{t = 0\}$ follows from the Erasing Singularity Theorem on bounded functions holomorphic on complement to a hypersurface. Therefore, both functions admit a Taylor series in t with coefficients being analytic functions in the above Taylor coefficients.

Case (2) of general C^{n+1}-smooth metric g and C^n-smooth curve γ. Consider other, analytic metric $\widetilde{g}$ and curve $\widetilde{\gamma}$ representing their n-jets. The functions $\widetilde{L}(0,t) = L_{\widetilde{g},\widetilde{\gamma}}(0,t)$ and $\widetilde{\Lambda}(t) = \Lambda_{\widetilde{g},\widetilde{\gamma}}(t)$ defined by them are analytic and coincide with the functions $L(0,t) = L_{g,\gamma}(0,t)$ and $\Lambda(t) = \Lambda_{g,\gamma}(t)$ corresponding to g and γ up to $o(t^{n+1})$. Indeed, $L_{\widetilde{g},\gamma}(0,t) - L_{g,\gamma}(0,t) = o(t^{n+1})$, by Lemma 8.13, $L_{\widetilde{g},\widetilde{\gamma}}(0,t) - L_{\widetilde{g},\gamma}(0,t) = o(t^{n+1})$, by Lemma 8.8 applied to $b = 0$. Thus, $\widetilde{L}(0,t) - L(0,t) = o(t^{n+1})$, $\widetilde{\Lambda}(t) - \Lambda(t) = o(t^{n+1})$. This together with the discussion in Case (1) implies that $L(0,t)$ and $\Lambda(t)$ have asymptotic Taylor expansions of order up to t^{n+1} coinciding with those of $\widetilde{L}(0,t)$ and $\widetilde{\Lambda}(t)$, and hence, having coefficients being analytic functions of the n-jets of g and γ. They depend only on pure n-jets, since applying a translation of both the curve and the metric leaves $L(0,t)$ and $\Lambda(t)$ invariant. Lemma 8.15 is proved.

8.6. Proof of Lemma 8.6

Let Σ be a two-dimensional surface equipped with a C^{n+1}-smooth Riemannian metric g. Let $V \subset \Sigma$ be a domain equipped with a chart (x,y) (not necessarily normal) where the metric is C^{n+1}-smooth. Consider a C^n-smooth germ of curve γ at a point $O \in V$ with positive geodesic curvature that is a graph of C^n-smooth function $\{y = h(x)\}$; the tangent line $T_O\gamma$ is not necessarily horizontal. The corresponding function $\Lambda(t)$ admits an asymptotic Taylor expansion

$$\Lambda(t) = \sum_{k=3}^{n+1} \widehat{\Lambda}_k t^k + o(t^{n+1}).$$

Its coefficients are analytic functions of the pure n-jets of the metric and γ at O (Lemma 8.15). Therefore, without loss of generality we consider that O is the origin in the coordinates (x,y), applying a translation. Then,

$$\gamma = \{y = h(x)\}, \quad h(x) = b_1 x + \frac{b_2}{2}x^2 + \frac{1}{3!}b_3 x^3 + \cdots + \frac{1}{n!}b_n x^n + o(x^n).$$

By definition, the coordinates of the pure jet $j_O^n\gamma$ are $(b_1,\dots,b_n)$.

We already know that $\widehat{\Lambda}_{n+1}$ is an affine function in b_n, which follows from Lemma 8.8, see (8.11). To obtain a precise formula for its coefficient at b_n, we use the following proposition.

Proposition 8.16 *Let $n \geq 5$, Σ, O, (x,y), $h(x)$ be as above. Consider a family of tangent germs of C^n-smooth curves $\gamma_{n,b} = \{y = h_{n,b}(x)\}$ at O, $h_{n,b}(x) = h(x) + bx^n + o(x^n)$; $h_{n,0} := h$, $\gamma_{n,0} := \gamma$. Let $w \in T_O\Sigma$ denote the orthogonal projection of the vector $\frac{\partial}{\partial y}$ to $(T_O\gamma)^\perp$. Let $u = (1, b_1) \in T_O\gamma$: the tangent vector to γ with unit x-component. Let $\kappa(O)$ denote the geodesic curvature of the curve γ at O, which coincides with that of $\gamma_{n,b}$. Let $(\widetilde{x},\widetilde{y})$ be normal coordinates centered at O such that the $\widetilde{x}$-axis is tangent to γ. Set*

$$\hat{x} := \kappa(O)\widetilde{x}, \ \hat{y} := \kappa(O)\widetilde{y}.$$

In the coordinates $(\hat{x}, \hat{y})$, the family of curves $\gamma_{n,b}$ is the family of graphs of C^n-functions $\{\hat{y} = \hat{h}_{n,b}(\hat{x})\}$, set $\hat{h}_{n,0} := \hat{h}$, such that $\hat{h}(\hat{x}) = \frac{\hat{x}^2}{2} + O(\hat{x}^3)$,

$$\hat{h}_{n,b}(\hat{x}) = \hat{h}(\hat{x}) + \mu_n b \hat{x}^n + o(\hat{x}^n), \ \mu_n = ||w|| ||u||^{-n} \kappa^{1-n}(O). \tag{8.50}$$

Proof Note that the normal coordinates $(\widetilde{x}, \widetilde{y})$ are C^n-smooth, and the metric g is C^{n-1} smooth there, since $g \in C^{n+1}$. Hence, the curves under question are also C^n-smooth in these coordinates. Fix a point $A = (\widetilde{x}, 0)$ on the $\widetilde{x}$-axis. Let ℓ denote the geodesic through A orthogonal to the $\widetilde{x}$-axis. We have to calculate the gap (i.e., distance) $\widetilde{\Delta}(\widetilde{x})$ between the intersection points of the geodesic ℓ with the curves $\gamma_{n,b}$ and γ. Let $\Delta(\widetilde{x})$ denote the gap between the points of the intersection of the curves with the vertical line $\{x = x(A)\}$. Their ratio $\widetilde{\Delta}(\widetilde{x})/\Delta(\widetilde{x})$ tends to the cosine of the angle between the vector $\frac{\partial}{\partial y} \in T_O\Sigma$ and the line $(T_O\gamma)^{\perp}$, as $\widetilde{x} \to 0$. One has $\Delta(\widetilde{x}) = ||\frac{\partial}{\partial y}|| b x^n + o(x^n)$. Hence, by definition,

$$\widetilde{\Delta}(\widetilde{x}) = ||w|| b x^n + o(x^n). \tag{8.51}$$

One has $dx = \alpha d\widetilde{x} + \beta d\widetilde{y}$ on $T_O\Sigma$, $\alpha = dx(\frac{\partial}{\partial \widetilde{x}}) = ||u||^{-1}$, by definition; $x = \alpha\widetilde{x} + \beta\widetilde{y} + O(|\widetilde{x}|^2 + |\widetilde{y}|^2)$. One has $\widetilde{y} = \frac{\kappa(O)}{2}\widetilde{x}^2 = O(\widetilde{x}^2)$ along each curve $\gamma_{n,b}$, by (2.1). This together with (8.51) implies that

$$\widetilde{\Delta}(\widetilde{x}) = ||w|| ||u||^{-n} b \widetilde{x}^n + o(\widetilde{x}^n). \tag{8.52}$$

Hence, in the coordinates $(\widetilde{x}, \widetilde{y})$,

$$\gamma_{n,b} = \{\widetilde{y} = \widetilde{h}_{n,b}(\widetilde{x})\}, \quad \widetilde{h}_{n,b}(\widetilde{x}) = \widetilde{h}_{n,0}(\widetilde{x}) + ||w|| ||u||^{-n} b \widetilde{x}^n + o(\widetilde{x}^n).$$

Now, rescaling to the coordinates $(\hat{x}, \hat{y})$ yields that $\gamma_{n,b}$ is a family of graphs of functions $\hat{h}_{n,b}(\hat{x})$ satisfying (8.50). The proposition is proved.

Proposition 8.17 *Consider the above family of curves $\gamma_{n,b}$ and the corresponding functions $\Lambda^{n,b}(t)$ from (8.3), set $\Lambda^{n,0} := \Lambda$. For every $n \geq 5$, one has*

$$\widehat{\Lambda}^{n,b}_{n+1} = \widehat{\Lambda}_{n+1} + \nu_n b, \ \nu_n := \begin{cases} \frac{(n-2)(n-3)}{6(n+1)} ||w|| (||u|| \kappa(O))^{-n}, \ \textit{for odd } n, \\ 0, \ \textit{for even } n. \end{cases} \tag{8.53}$$

Proof The coordinates $(\hat{x}, \hat{y})$ are normal coordinates for the rescaled metric $\hat{g} := \kappa(O) g$. The common geodesic curvature at O of the curves $\gamma_{n,b}$ in the metric $\hat{g}$ is equal to 1, by Remark 8.5. Therefore, for the metric $\hat{g}$, one has $\widehat{\Lambda}^{n,b}_{n+1} - \widehat{\Lambda}_{n+1} = \frac{(n-2)(n-3)}{6(n+1)} \mu_n b$ for odd n, and the latter difference vanishes for even n, by Lemma 8.8 and (8.50). Rescaling the metric back to g by the factor $\kappa^{-1}(O)$ rescales the functions $\Lambda^{n,b}$ and their Taylor coefficients by the same factor (Remark 8.5). This implies (8.53).

Proposition 8.18 *Let $n \geq 5$, and let the metric on Σ be C^{n+1}-smooth. Let γ be a germ of C^n-smooth curve at a point $O \in \Sigma$ lying in a chart with coordinates (x, y). Let γ be a graph $\{y = h(x)\}$. Let $b_1, \dots, b_n$ denote the coordinates of the pure n-jet $j^n_O h$. Let $w, u \in T_O\Sigma$ be the vectors from Proposition 8.16.*

Then, the Taylor coefficient $\widehat{\Lambda}_{n+1}$ *of the corresponding function* $\Lambda(t)$ *is equal to*

$$\widehat{\Lambda}_{n+1} = \sigma_n b_n - P_n, \tag{8.54}$$

$$\sigma_n = \frac{(n-2)(n-3)}{6(n+1)!}||w||\,(||u||\kappa(O))^{-n} \ \textit{for odd } n, \tag{8.55}$$

$\sigma_n = 0$ *for even* n*, where* P_n *is an analytic function in* $b_1, \dots, b_{n-1}$ *and in the pure* n*-jet of the metric at* O.

Proof The fact that $\widehat{\Lambda}_{n+1}$ depends on b_n as an affine function with factor σ_n at b_n follows from definition and Proposition 8.17; the b from Proposition 8.17 is $\frac{1}{n!}$ times the difference of the b_n-coordinates of jets of functions $h_{n,b}(x)$ and $h(x)$. The function P_n is thus independent on b_n and hence has the required type, by Lemma 8.15.

Proof of Lemma 8.6 and its addendum. All the statements of Lemma 8.6 and its addendum follow from the above proposition, except for the following points discussed below. Note that σ_n depends only on the pure 2-jet of the curve γ and the pure 1-jet of the metric, by definition. The function P_n is an analytic function of the pure n-jet of the metric and the pure $(n-1)$-jet of the curve γ. Let us treat it as a function of a point and a pure $(n-1)$-jet of curve. We have to prove its smoothness. To this end, we use the assumption that the metric is C^{n+1}-smooth. (This is the main place in the proof where we use this assumption.) Then, its pure n-jet is a C^1-smooth function of a point. Similarly, σ_n is smooth, by (8.55). This together with the above analyticity statement proves C^1-smoothness and finishes the proof of Lemma 8.6.

8.7. Proof of Theorems 8.4 and 1.19

Proof of Theorem 8.4 Let $O \in \Sigma$. Let (x, y) be local coordinates on a neighborhood $V = V(O) \subset \Sigma$. Let $\mathcal{J}^4_y(V)$ denote the space of κ-nondegenerate 4-jets of curves, as in Lemma 8.6, which are graphs of functions $\{y = h(x)\}$. Let $J_2 = (x, b_0, b_1, b_2)$, $\sigma_5 = \sigma_5(J_2)$ and $h_5 := P_5(J_2; b_3, b_4)$ be the same, as in (8.8). Consider the field of kernels K_4 of the following 1-form ν_4 on $\mathcal{J}^4_y(V)$:

$$\nu_4 := db_4 - \sigma_5^{-1} h_5(x, b_0, b_1, b_2, b_3, b_4)dx;\ K_4 := \mathrm{Ker}(\nu_4).$$

Let $\mathcal{D}_4$ denote the contact distribution on $\mathcal{J}^4_y(V)$, see (8.2):

$$\mathcal{D}_4 = \mathrm{Ker}(\mathrm{d}b_0 - b_1\mathrm{d}x, \mathrm{d}b_1 - b_2\mathrm{d}x, \mathrm{d}b_2 - b_3\mathrm{d}x, \mathrm{d}b_3 - b_4\mathrm{d}x).$$

Set

$$\mathcal{P} := K_4 \cap \mathcal{D}_4. \tag{8.56}$$

This is a line field, since the above intersections are obviously transverse and $\dim(\mathcal{D}_4) = 2$. It is C^1-smooth, since so are σ_2 and h_5 (Lemma 8.6). Let γ be an arbitrary C^5-smooth germ of curve γ based at a point $A \in V$ with positive geodesic curvature such that the line $T_A\gamma$ is not parallel to the y-axis. Let γ have the string Poritsky property. Then, $\Lambda(t) \equiv 0$, hence, $\widehat{\Lambda}_6 = 0$, thus,

$$\sigma_5(J_2)b_5 - h_5(J_2; b_3, b_4) = 0, \tag{8.57}$$

by (8.8). On the other hand, the 5-jet extension of the curve γ is tangent to the contact distribution $\mathcal{D}_5$, and hence, to the hyperplane field $\{db_4 = b_5 dx\}$. This together with (8.57) implies that its 4-jet extension is tangent to the hyperplane field $\{db_4 = \frac{h_5}{\sigma_5} dx\}$. Thus, it is tangent to the kernel field K_4, and hence, to $\mathcal{P} = K_4 \cap \mathcal{D}_4$. This proves Theorem 8.4.

Proof of Theorem 1.19 Two germs of curves with the string Poritsky property and the same 4-jet correspond to the same point in $\mathcal{J}^4$. Therefore, their 4-jet extensions coincide with one and the same phase curve of the line field $\mathcal{P}$, by Theorem 8.4 and the Uniqueness Theorem for ordinary differential equations. Thus, the germs coincide. This proves Theorem 1.19.

Acknowledgements

I wish to thank Sergei Tabachnikov, to whom this paper owes much, for introducing me into the curves with the Poritsky property and to related areas. I wish to thank him and Misha Bialy and Maxim Arnold for helpful discussions. I wish to thank Felix Schlenk for helpful remarks. I wish to thank the referee for his work of reading the paper and helpful remarks. Most results of the paper were obtained while I was visiting the Mathematical Sciences Research Institute (MSRI) in Berkeley, California. I wish to thank MSRI for its hospitality and support.

References

[1] Amiran, E.: Caustics and evolutes for convex planar domains. J. Differ. Geometry **28**, 345–357 (1988)

[2] Amiran, E.: Lazutkin coordinates and invariant curves for outer billiards. J. Math. Phys. **36**(3), 1232–1241 (1995)

[3] Arnold, M., Tabachnikov, S.: Remarks on Joachimsthal integral and Poritsky property. Arnold Math. J. **7**, 483–491 (2021)

[4] Arnold, V.: Mathematical Methods of Classical Mechanics. Springer, New York (1978)

[5] Arnold, V.: Contact geometry and wave propagation. Monogr. **34** de l'Enseign. Math. (1989) Univ. of Geneva

[6] Bialy, M., Mironov, A.E.: Algebraic Birkhoff conjecture for billiards on Sphere and Hyperbolic plane. J. Geom. Phys. **115**, 150–156 (2017)

[7] Bennett, A.G.: Hyperbolic geometry - circles. https://www.maa.org/press/periodicals/loci/joma/hyperbolic-geometry-circles

[8] Bolsinov, A.V., Matveev, V. S., Miranda, E., Tabachnikov, S.: Open problems, questions and challenges in finite-dimensional integrable systems. Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. **376 (2131)** (2018)

[9] Glutsyuk, A.: On polynomially integrable Birkhoff billiards on surfaces of constant curvature. J. Eur. Math. Soc. **23**, 995–1049 (2021)

[10] Glutsyuk, A., Izmestiev, I., Tabachnikov, S.: Four equivalent properties of integrable billiards. Israel J. Math. **241**, 693–719 (2021)

[11] Glutsyuk, A., Shustin, E.: On polynomially integrable planar outer billiards and curves with symmetry property. Math. Ann. **372**, 1481–1501 (2018)

[12] Izmestiev, I.: Spherical and hyperbolic conics. In: Eighteen Essays in Non-Euclidean Geometry (editors: V.Alberge and A.Papadopoulos), IRMA Lectures in Mathematics and Theoretical Physics **29**, 262–320. Eur. Math. Soc. Publishing House (2019)

[13] Kaloshin, V., Sorrentino, A.: On local Birkhoff Conjecture for convex billiards. Ann. Math. **188**(1), 315–380 (2018)

[14] Klein, F.: Non-Euclidean geometry. Russian translation. NKTP SSSR, Moscow, Leningrad (1936)

[15] Lazutkin, V.F.: The existence of caustics for a billiard problem in a convex domain. Math. USSR Izvestija **7**, 185–214 (1973)

[16] Marvizi, S., Melrose, R.: Spectral invariants of convex planar regions. J. Differ. Geom. **17**, 475–502 (1982)

[17] Masaltsev, L.A.: Incidence theorems in spaces of constant curvature. J. Math. Sci. **72**(4), 3201–3206 (1994)

[18] Melrose, R.: Equivalence of glancing hypersurfaces. Invent. Math. **37**, 165–192 (1976)

[19] Melrose, R.: Equivalence of glancing hypersurfaces 2. Math. Ann. **255**, 159–198 (1981)

[20] Olver, P.: Equivalence, Invariants and Symmetry. Cambridge University Press, Cambridge (1995)

[21] Poritsky, H.: The billiard ball problem on a table with a convex boundary—an illustrative dynamical problem. Ann. Math. **51**(2), 446–470 (1950)

[22] Sossinsky, A.B.: Geometries. Student Mathematical Library, v. 64, AMS (2012)

[23] Tabachnikov, S.: Billiards. Panor. Synth. **1**, SMF, vi+142 (1995)

[24] Tabachnikov, S.: Geometry and billiards. Student Mathematical Library, **30**, American Mathematical Society, Providence, RI; Mathematics Advanced Study Semesters, University Park, PA (2005)

[25] Tabachnikov, S.: On algebraically integrable outer billiards. Pac. J. Math. **235**(1), 101–104 (2008)

[26] Tabachnikov, S.: Poncelet's theorem and dual billiards. L'Enseign. Math. **39**, 189–194 (1993)

[27] Tabachnikov, S.: Commuting dual billiard maps. Geom. Dedicat. **53**, 57–68 (1994)

[28] Tabachnikov, S., Dogru, F.: Dual billiards. Math. Intell. **27**(4), 18–25 (2005)

[29] Veselov, A.P.: Confocal surfaces and integrable billiards on the sphere and in the Lobachevsky space. J. Geom. Phys. **7**(1), 81–107 (1990)

Alexey Glutsyuk
CNRS, France (UMR 5669 (UMPA, ENS de Lyon) and UMI 2615 (Interdisciplinary Scientific Center J.-V.Poncelet));
HSE University, Moscow, Russian Federation

e-mail: aglutsyu@ens-lyon.fr

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:46
https://doi.org/10.1007/s11784-022-00966-5
Published online June 3, 2022

Journal of Fixed Point Theory and Applications

Publisher Correction

Correction to: On curves with the Poritsky property

Alexey Glutsyuk

Correction to: J. Fixed Point Theory Appl. (2022) 24:35 https://doi.org/10.1007/s11784-022-00948-7

There were typographical errors in the original publication of this article. The article has been corrected. We apologize for these mistakes.

Note that the results and conclusions are not affected by the present erratum.

The dedication "To Professor Claude Viterbo on the occasion of his 60th birthday. CNRS, France (UMR 5669 (UMPA, ENS de Lyon) and UMI 2615 (Interdisciplinary Scientific Center J.-V.Poncelet))" shall read "To Professor Claude Viterbo on the occasion of his 60th birthday."
On Page 16, line 2 before Subsection 2.4: the piece "with $u_{A,v(0)}$". The correct piece is "with $u_{A,v(0)}(0) = 0$".
On Page 23, Subsection 4.2, line 4: "counter clockwise". The correct piece is "counterclockwise".
On Page 25, Theorem 4.8, Statement (1), line 2: "(4.4)holds". The correct piece is "(4.4) holds".
On Page 43, line 2 before formula (8.5): "Let $L(A, B)$ the". The correct piece is "Let $L(A, B)$ be the".
On Page 43, formula (8.6): "$\sigma_n \neq 0$ for odd $n > 3$; $\sigma_n \equiv 0$ for $n = 3$ and for every even $n > 3$". The correct version of formula (8.6) is

$$\sigma_n \neq 0 \text{ for odd } n > 3;\ \sigma_n \equiv 0 \text{ for every even } n > 5.$$

On Page 44, the line before formula (8.9): "$n \geq 3$". The correct piece is "$n \geq 5$".

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.
The original article can be found online at https://doi.org/10.1007/s11784-022-00948-7.

Various proof headings "Proof of Lemma 8.8." were erroneous after following statements: proof headings "Proof of Lemma 8.8" after the 10 following statements:
Page 51: Lemma 8.13 and Proposition 8.14. Page 52: Claims 1 and 2. Page 53: Claims 3 and 4. Page 54: Lemma 8.15. Pages 56-57: Propositions 8.16, 8.17 and 8.18.
The correct version of each of the above-mentioned proof headings is "Proof".
On Page 56, Proposition 8.17, line 2, "$n \geq 3$". The correct version of this piece is "$n \geq 5$".
On Page 60, Affiliations
"CNRS, France (UMR 5669 (UMPA, ENS de Lyon) and UMI 2615 (Interdisciplinary Scientific Center J.-V.Poncelet)) and HSE University Moscow Russian Federation"
The correct version of this piece is
"- CNRS, France (UMR 5669 (UMPA, ENS de Lyon) and UMI 2615 (Interdisciplinary Scientific Center J.-V.Poncelet));
- HSE University, Moscow, Russian Federation."

Alexey Glutsyuk
CNRS, France (UMR 5669 (UMPA, ENS de Lyon) and UMI 2615 (Interdisciplinary Scientific Center J.-V.Poncelet));
HSE University, Moscow, Russian Federation
e-mail: aglutsyu@ens-lyon.fr

J. Fixed Point Theory Appl. (2022) 24:41
https://doi.org/10.1007/s11784-022-00949-6
Published online April 20, 2022

Journal of Fixed Point Theory and Applications

Examples around the strong Viterbo conjecture

Jean Gutt, Michael Hutchings and Vinicius G. B. Ramos

Abstract. A strong version of a conjecture of Viterbo asserts that all normalized symplectic capacities agree on convex domains. We review known results showing that certain specific normalized symplectic capacities agree on convex domains. We also review why all normalized symplectic capacities agree on S^1-invariant convex domains. We introduce a new class of examples called "monotone toric domains", which are not necessarily convex, and which include all dynamically convex toric domains in four dimensions. We prove that for monotone toric domains in four dimensions, all normalized symplectic capacities agree. For monotone toric domains in arbitrary dimension, we prove that the Gromov width agrees with the first equivariant capacity. We also study a family of examples of non-monotone toric domains and determine when the conclusion of the strong Viterbo conjecture holds for these examples. Along the way, we compute the cylindrical capacity of a large class of "weakly convex toric domains" in four dimensions.

Mathematics Subject Classification. 53D35, 53D42.

Keywords. Symplectic capacities, toric domains, Viterbo's conjecture.

1. Introduction

If X and X' are domains[1] in $\mathbb{R}^{2n} = \mathbb{C}^n$, a *symplectic embedding* from X to X' is a smooth embedding $\varphi : X \hookrightarrow X'$ such that $\varphi^\star\omega = \omega$, where ω denotes the standard symplectic form on $\mathbb{R}^{2n}$. If there exists a symplectic embedding from X to X', we write $X \underset{s}{\hookrightarrow} X'$.

M. Hutchings, partially supported by NSF grant DMS-2005437
V. G. B. Ramos partially supported by grants from the Serrapilheira Institute, FAPERJ and CNPq
This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

[1]In this paper, a "domain" is the closure of an open set. One can of course also consider domains in other symplectic manifolds, but we will not do so here.

An important problem in symplectic geometry is to determine when symplectic embeddings exist, and more generally to classify the symplectic embeddings between two given domains. Modern work on this topic began with the Gromov non-squeezing theorem [11], which asserts that the ball

$$B^{2n}(r) = \left\{ z \in \mathbb{C}^n \;\middle|\; \pi|z|^2 \le r \right\}$$

symplectically embeds into the cylinder

$$Z^{2n}(R) = \left\{ z \in \mathbb{C}^n \;\middle|\; \pi|z_1|^2 \le R \right\}$$

if and only if $r \le R$. Many questions about symplectic embeddings remain open, even for simple examples such as ellipsoids and polydisks.

If there exists a symplectic embedding $X \underset{s}{\hookrightarrow} X'$, then we have the volume constraint $\mathrm{Vol}(X) \le \mathrm{Vol}(X')$. To obtain more nontrivial obstructions to the existence of symplectic embeddings, one often uses various symplectic capacities. Definitions of the latter term vary; here we define a *symplectic capacity* to be a function c which assigns to each domain in $\mathbb{R}^{2n}$, possibly in some restricted class, a number $c(X) \in [0, \infty]$, satisfying the following axioms:

(Monotonicity) If X and X' are domains in $\mathbb{R}^{2n}$, and if there exists a symplectic embedding $X \underset{s}{\hookrightarrow} X'$, then $c(X) \le c(X')$.

(Conformality) If r is a positive real number then $c(rX) = r^2 c(X)$.

We say that a symplectic capacity c is *normalized* if it is defined at least for convex domains and satisfies

$$c\big(B^{2n}(1)\big) = c\big(Z^{2n}(1)\big) = 1.$$

The first example of a normalized symplectic capacity is the *Gromov width* defined by

$$c_{\mathrm{Gr}}(X) = \sup \left\{ r \;\middle|\; B^{2n}(r) \underset{s}{\hookrightarrow} X \right\}.$$

This trivially satisfies all of the axioms except for the normalization requirement $c_{\mathrm{Gr}}(Z^{2n}(1))$, which holds by Gromov non-squeezing. A similar example is the *cylindrical capacity* defined by

$$c_Z(X) = \inf \left\{ R \;\middle|\; X \underset{s}{\hookrightarrow} Z^{2n}(R) \right\}.$$

Additional examples of normalized symplectic capacities are the Hofer–Zehnder capacity c_{HZ} defined in [16] and the Viterbo capacity c_{SH} defined in [31]. There are also useful families of symplectic capacities parametrized by a positive integer k, including the Ekeland–Hofer capacities c_k^{EH} defined in [8,9] using calculus of variations; the "equivariant capacities" c_k^{CH} defined in [12] using positive equivariant symplectic homology; and in the four-dimensional case, the ECH capacities c_k^{ECH} defined in [17] using embedded contact homology. For each of these families, the $k = 1$ capacities c_1^{EH}, c_1^{CH}, and c_1^{ECH} are normalized. Some additional symplectic capacities defined using rational symplectic field theory were recently introduced in [27,28]. For more about symplectic capacities in general, we refer to [6,25] and the references therein.

The goal of this paper is to discuss some results and examples related to the following conjecture, which apparently has been folkore since the 1990s.

Conjecture 1.1. (strong Viterbo conjecture) *If X is a convex domain in $\mathbb{R}^{2n}$, then all normalized symplectic capacities of X are equal.*

Viterbo conjectured the following statement[2] in [32]:

Conjecture 1.2. (Viterbo conjecture) *If X is a convex domain in $\mathbb{R}^{2n}$ and if c is a normalized symplectic capacity, then*

$$c(X) \leq (n! \operatorname{Vol}(X))^{1/n}. \tag{1.1}$$

The inequality (1.1) is true when c is the Gromov width c_{Gr}, by the volume constraint, because $\operatorname{Vol}(B^{2n}(r)) = r^n/n!$. Thus, Conjecture 1.1 implies Conjecture 1.2. The Viterbo conjecture recently gained more attention as it was shown in [4] that it implies the Mahler conjecture[3] in convex geometry.

Lemma 1.3. *If X is a domain in $\mathbb{R}^{2n}$, then $c_{\mathrm{Gr}}(X) \leq c_Z(X)$, with equality if and only if all normalized symplectic capacities of X agree (when they are defined for X).*

Proof. It follows from the definitions that if c is a normalized symplectic capacity defined for X, then $c_{\mathrm{Gr}}(X) \leq c(X) \leq c_Z(X)$. □

Thus, the strong Viterbo conjecture is equivalent to the statement that every convex domain X satisfies $c_{\mathrm{Gr}}(X) = c_Z(X)$. We now discuss some examples where it is known that $c_{\mathrm{Gr}} = c_Z$. Hermann [13] showed that all T^n-invariant convex domains have to satisfy $c_{\mathrm{Gr}} = c_Z$. This generalizes to S^1-invariant convex domains by the following elementary argument:

Proposition 1.4. (Y. Ostrover, private communication) *Let X be a compact convex domain in $\mathbb{C}^n$ which is invariant under the S^1 action by $e^{i\theta} \cdot z = (e^{i\theta} z_1, \ldots, e^{i\theta} z_n)$. Then $c_{\mathrm{Gr}}(X) = c_Z(X)$.*

Proof. By compactness, there exists $z_0 \in \partial X$ minimizing the distance to the origin. Let $r > 0$ denote this minimal distance. Then the ball $(|z| \leq r)$ is contained in X, so by definition $c_{\mathrm{Gr}}(X) \geq \pi r^2$.

By applying an element of $U(n)$, we may assume without loss of generality that $z_0 = (r, 0, \ldots, 0)$. By a continuity argument, we can assume without loss of generality that ∂X is a smooth hypersurface in $\mathbb{R}^{2n}$. By the distance minimizing property, the tangent plane to ∂X at z_0 is given by $(z \cdot (1, 0, \ldots, 0) = r)$ where $\cdot$ denotes the real inner product. By convexity, X is contained in the half-space $(z \cdot (1, 0, \ldots, 0) \leq r)$. By the S^1 symmetry, X is also contained in the half-space $(z \cdot (e^{i\theta}, 0, \ldots, 0) \leq r)$ for each $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. Thus, X is contained in the intersection of all these half-spaces, which is the cylinder $|z_1| \leq r$. Then $c_Z(X) \leq \pi r^2$ by definition. □

[2] Viterbo also conjectured that equality holds in (1.1) only if $\operatorname{int}(X)$ is symplectomorphic to an open ball.

[3] The Mahler conjecture [22] states that for any n-dimensional normed space V, we have

$$\operatorname{Vol}(B_V) \operatorname{Vol}(B_{V^*}) \geq \frac{4^n}{n!},$$

where B_V denotes the unit ball of V, and B_{V^*} denotes the unit ball of the dual space V^*. For some examples of Conjectures 1.1 and 1.2 related to the Mahler conjecture, see [26].

Remark 1.5. A similar argument shows that if $k \geq 3$ is an integer and if $X \subset \mathbb{C}^n$ is a convex domain invariant under the $\mathbb{Z}/k$ action by $j \cdot z = (e^{2\pi ij/k}z_1, \ldots, e^{2\pi ij/k}z_n)$, then

$$\frac{c_Z(X)}{c_{\mathrm{Gr}}(X)} \leq \frac{k}{\pi}\tan(\pi/k).$$

The role of the convexity hypothesis in Conjecture 1.1 is somewhat mysterious. We now explore to what extent non-convex domains can satisfy $c_{\mathrm{Gr}} = c_Z$.

To describe some examples, if Ω is a domain in $\mathbb{R}^n_{\geq 0}$, define the *toric domain*

$$X_\Omega = \left\{ z \in \mathbb{C}^n \;\middle|\; \pi(|z_1|^2, \ldots, |z_n|^2) \in \Omega \right\}.$$

The factors of π ensure that

$$\mathrm{Vol}(X_\Omega) = \mathrm{Vol}(\Omega). \tag{1.2}$$

Let $\partial_+\Omega$ denote the set of $\mu \in \partial\Omega$ such that $\mu_j > 0$ for all $j = 1, \ldots, n$.

Definition 1.6. A *monotone toric domain* is a compact toric domain X_Ω with smooth boundary such that if $\mu \in \partial_+\Omega$ and if v an outward normal vector at μ, then $v_j \geq 0$ for all $j = 1, \ldots, n$. See Figure 1c.

A *strictly monotone toric domain* is a compact toric domain X_Ω with smooth boundary such that if $\mu \in \overline{\partial_+\Omega}$ and if v is a nonzero outward normal vector at μ, then $v_j > 0$ for all $j = 1, \ldots, n$.

One of our main results is the following:

Theorem 1.7. (proved in Sect. 4) *If X_Ω is a monotone toric domain in $\mathbb{R}^4$, then $c_{\mathrm{Gr}(X_\Omega)} = c_Z(X_\Omega)$.*

Note that monotone toric domains do not have to be convex; see Sect. 2 for details on when toric domains are convex. (Toric domains that are convex are already covered by Proposition 1.4.)

To clarify the hypothesis in Theorem 1.7, let X be a compact domain in $\mathbb{R}^{2n}$ with smooth boundary, and suppose that X is "star-shaped", meaning that the radial vector field on $\mathbb{R}^{2n}$ is transverse to ∂X. Then there is a well-defined Reeb vector field R on ∂X. We say that X is *dynamically convex* if, in addition to the above hypotheses, every Reeb orbit γ has Conley–Zehnder index $\mathrm{CZ}(\gamma) \geq n + 1$ if nondegenerate, or in general has minimal Conley-Zehnder index[4] at least $n + 1$. It was shown by Hofer–Wysocki–Zehnder [14] that if X is strictly convex, then X is dynamically convex. However, the Viterbo conjecture implies that not every dynamically convex domain is symplectomorphic to a convex domain; see Remark 1.9 below.

[4]If γ is degenerate then there is an interval of possible Conley–Zehnder indices of nondegenerate Reeb orbits near γ after a perturbation, and for dynamical convexity we require the minimum number in this interval to be at least $n+1$. In the 4-dimensional case ($n = 2$), this means that the dynamical rotation number of the linearized Reeb flow around γ, which we denote by $\rho(\gamma) \in \mathbb{R}$, is greater than 1.

Proposition 1.8. (proved in Sect. 2) *Let X_Ω be a compact star-shaped toric domain in $\mathbb{R}^4$ with smooth boundary. Then X_Ω is dynamically convex if and only if X_Ω is a strictly monotone toric domain.*

Thus, Theorem 1.7 implies that all dynamically convex toric domains in $\mathbb{R}^4$ have $c_{\mathrm{Gr}} = c_Z$.

If X is a star-shaped domain with smooth boundary, let $A_{\min}(X)$ denote the minimal period of a Reeb orbit on ∂X.

Remark 1.9. Without the toric hypothesis, not all dynamically convex domains in $\mathbb{R}^4$ have $c_{\mathrm{Gr}} = c_Z$. In particular, it is shown in [1] that for $\varepsilon > 0$ small, there exists a dynamically convex domain X in $\mathbb{R}^4$ such that

$$A_{\min}(X)^2/(2\,\mathrm{vol}(X)) \geq 2 - \varepsilon.$$

One has $c_1^{\mathrm{CH}}(X) \geq A_{\min}(X)$ by [12, Thm. 1.1], and $c_{\mathrm{Gr}}(X)^2 \leq 2\,\mathrm{vol}(X)$ by the volume constraint. Thus,

$$\frac{c_Z(X)}{c_{\mathrm{Gr}}(X)} \geq \sqrt{2-\varepsilon}.$$

Remark 1.10. It is also not true that all star-shaped toric domains have $c_{\mathrm{Gr}} = c_Z$. Counterexamples have been known for a long time, see, e.g., [13], and in Sect. 5 we discuss a new family of counterexamples.

For monotone toric domains in higher dimensions, we do not know how to prove that all normalized symplectic capacities agree, but we can at least prove the following:

Theorem 1.11. (proved in Sect. 3) *If X_Ω is a monotone toric domain in $\mathbb{R}^{2n}$, then*

$$c_{\mathrm{Gr}}(X_\Omega) = c_1^{\mathrm{CH}}(X_\Omega). \tag{1.3}$$

Returning to convex domains, some normalized symplectic capacities are known to agree (not the Gromov width or cylindrical capacity, however), as we review in the following theorem:

Theorem 1.12. (Ekeland, Hofer, Zehnder, Abbondandolo–Kang, Irie) *If X is a convex domain in $\mathbb{R}^{2n}$, then:*

(a) $c_1^{\mathrm{EH}}(X) = c_{\mathrm{HZ}}(X) = c_{\mathrm{SH}}(X) = c_1^{\mathrm{CH}}(X)$.
(b) If in addition ∂X is smooth[5], then all of the capacities in (a) agree with $A_{\min}(X)$.

Proof. Part (b) implies part (a) because every convex domain can be C^0 approximated by one with smooth boundary; and the capacities in (a) are C^0 continuous functions of the convex domain X, by monotonicity and conformality.

[5] Without the smoothness assumption, it is shown in [3, Prop. 2.7] that $c_{\mathrm{HZ}}(X)$ agrees with the minimum action of a "generalized closed characteristic" on ∂X.

Part (b) was shown for $c_{\mathrm{HZ}}(X)$ by Hofer–Zehnder in [16], and for $c_{\mathrm{SH}}(X)$ by Irie [20] and Abbondandolo–Kang [2]. The agreement of these two capacities with $c_1^{\mathrm{CH}}(X)$ for convex domains now follows from the combination of [12, Theorem 1.24] and [10, Lemma 3.2], as explained by Irie in [20, Remark 2.15]. Finally, part (b) for $c_1^{\mathrm{EH}}(X)$ has been claimed and understood for a long time, but since we could not find a complete proof in the literature we give one here in Sect. 6. □

Organization of the paper

In Sect. 2, we discuss different kinds of toric domains and when they are convex or dynamically convex. In Sect. 3, we consider the first equivariant capacity and prove Theorem 1.11. In Sect. 4, we use ECH capacities to prove Theorem 1.7. In Sect. 5, we consider a family of examples of non-monotone toric domains and determine when they do or do not satisfy the conclusions of Conjectures 1.1 and 1.2. Along the way, we compute the cylindrical capacity of a large class of "weakly convex toric domains" in four dimensions (Theorem 5.6). In Sect. 6, we review the definition of the first Ekeland–Hofer capacity and complete the (re)proof of Theorem 1.12.

2. Toric domains

In this section, we review some important classes of toric domains and discuss when they are convex or dynamically convex.

If Ω is a domain in $\mathbb{R}^n$, define

$$\widehat{\Omega} = \left\{ \mu \in \mathbb{R}^n \;\middle|\; (|\mu_1|, \ldots, |\mu_n|) \in \Omega \right\}.$$

Definition 2.1. [12] A *convex toric domain* is a toric domain X_Ω such that $\widehat{\Omega}$ is compact and convex. See Figure 1a.

This terminology may be misleading because a "convex toric domain" is not the same thing as a compact toric domain that is convex in $\mathbb{R}^{2n}$; see Proposition 2.3 below.

Definition 2.2. [12] A *concave toric domain* is a toric domain X_Ω such that Ω is compact and $\mathbb{R}^n_{\geq 0} \setminus \Omega$ is convex. See Figure 1b.

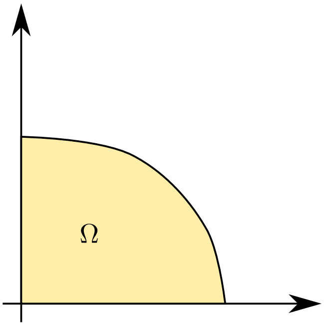
We remark that if X_Ω is a convex toric domain or concave toric domain and if X_Ω has smooth boundary, then it is a monotone toric domain.

Proposition 2.3. *A toric domain X_Ω is a convex subset of $\mathbb{R}^{2n}$ if and only if the set*

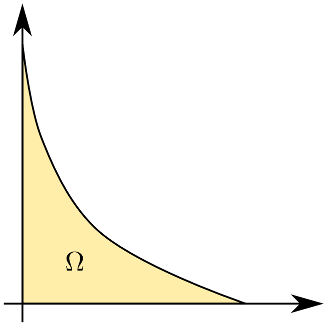
$$\widetilde{\Omega} = \left\{ \mu \in \mathbb{R}^n \;\middle|\; \pi\left(|\mu_1|^2, \ldots, |\mu_n|^2\right) \in \Omega \right\} \tag{2.1}$$

is convex in $\mathbb{R}^n$.

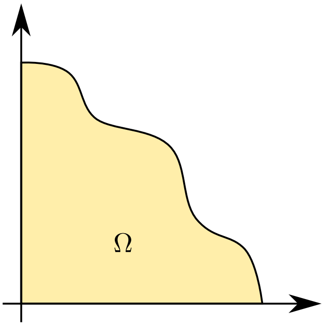
Proof. ($\Rightarrow$) The set $\widetilde{\Omega}$ is just the intersection of the toric domain X_Ω with the subspace $\mathbb{R}^n \subset \mathbb{C}^n$. If X_Ω is convex, then its intersection with any linear subspace is also convex.



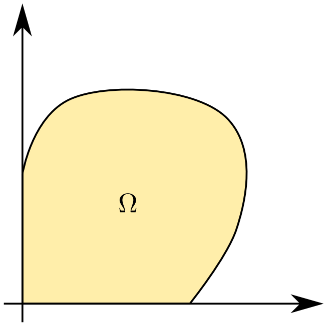
(a) A convex toric domain



(b) A concave toric domain



(c) A monotone toric domain



(d) A weakly convex toric domain

FIGURE 1. Examples of toric domains X_Ω in $\mathbb{R}^4$

($\Leftarrow$) Suppose that the set $\widetilde{\Omega}$ is convex. Let $z, z' \in X_\Omega$ and let $t \in [0,1]$. We need to show that

$$(1-t)z + tz' \in X_\Omega.$$

That is, we need to show that

$$(|(1-t)z_1 + tz_1'|, \ldots, |(1-t)z_n + z_n'|) \in \widetilde{\Omega}. \tag{2.2}$$

We know that the 2^n points $(\pm|z_1|, \ldots, \pm|z_n|)$ are all in $\widetilde{\Omega}$, as are the 2^n points $(\pm|z_1'|, \ldots, \pm|z_n'|)$. By the triangle inequality, we have

$$|(1-t)z_j + tz_j'| \leq (1-t)|z_j| + t|z_j'|$$

for each $j = 1, \ldots, n$. It follows that the point in (2.2) can be expressed as $(1-t)$ times a convex combination of the points $(\pm|z_1|, \ldots, \pm|z_n|)$, plus t times a convex combination of the points $(\pm|z_1'|, \ldots, \pm|z_n'|)$. Since $\widetilde{\Omega}$ is convex, it follows that (2.2) holds. $\square$

Example 2.4. If X_Ω is a convex toric domain, then X_Ω is a convex subset of $\mathbb{R}^{2n}$.

Proof. Similarly to the above argument, this boils down to showing that if $w, w' \in \mathbb{C}$ and $0 \le t \le 1$ then

$$|(1-t)w + tw'|^2 \le (1-t)|w|^2 + t|w'|^2.$$

The above inequality holds because the right hand side minus the left hand side equals $(t-t^2)|w-w'|^2$. □

However, the converse is not true:

Example 2.5. Let $p > 0$, and let Ω be the positive quadrant of the L^p unit ball,

$$\Omega = \left\{ \mu \in \mathbb{R}^n_{\ge 0} \;\middle|\; \sum_{j=1}^n \mu_j^p \le 1 \right\}.$$

Then X_Ω is a concave toric domain if and only if $p \le 1$, and a convex toric domain if and only if $p \ge 1$. By Proposition 2.3, the domain X_Ω is convex in $\mathbb{R}^{2n}$ if and only if $p \ge 1/2$.

We now work out when four-dimensional toric domains are dynamically convex.

Proof of Proposition 1.8. As a preliminary remark, note that if a Reeb orbit has rotation number $\rho > 1$, then so does every iterate of the Reeb orbit. Thus, X_Ω is dynamically convex if and only if every *simple* Reeb orbit has rotation number $\rho > 1$.

Since X_Ω is star-shaped, Ω itself is also star-shaped. Since X_Ω is compact with smooth boundary, $\overline{\partial_+\Omega}$ is a smooth arc from some point $(0,b)$ with $b>0$ to some point $(a,0)$ with $a>0$.

We can find the simple Reeb orbits and their rotation numbers by the calculations in [5, §3.2] and [12, §2.2]. The conclusion is the following. There are three types of simple Reeb orbits on ∂X_Ω:

(i) There is a simple Reeb orbit corresponding to $(a,0)$, whose image is the circle in ∂X_Ω with $\pi|z_1|^2 = a$ and $z_2 = 0$.

(ii) Likewise, there is a simple Reeb orbit corresponding to $(0,b)$, whose image is the circle in ∂X_Ω with $z_1 = 0$ and $\pi|z_2|^2 = b$.

(iii) For each point $\mu \in \partial_+\Omega$ where $\partial_+\Omega$ has rational slope, there is an S^1 family of simple Reeb orbits whose images sweep out the torus in ∂X_Ω where $\pi(|z_1|^2, |z_2|^2) = \mu$.

Let s_1 denote the slope of $\overline{\partial_+\Omega}$ at $(a,0)$, and let s_2 denote the slope of $\overline{\partial_+\Omega}$ at $(0,b)$. Then the Reeb orbit in (i) has rotation number $\rho = 1 - s_1^{-1}$, and the Reeb orbit in (ii) has rotation number $\rho = 1 - s_2$. For a Reeb orbit in (iii), let $\nu = (\nu_1, \nu_2)$ be the outward normal vector to $\partial_+\Omega$ at μ, scaled so that ν_1, ν_2 are relatively prime integers. Then each Reeb orbit in this family has rotation number $\rho = \nu_1 + \nu_2$.

If X_Ω is strictly monotone, then $s_1, s_2 < 0$, and for each Reeb orbit of type (iii) we have $\nu_1, \nu_2 \ge 1$. It follows that every simple Reeb orbit has rotation number $\rho > 1$.

Conversely, suppose that every simple Reeb orbit has rotation number $\rho > 1$. Applying this to the Reeb orbits (i) and (ii), we obtain that $s_1, s_2 < 0$. Thus, $\partial_+\Omega$ has negative slope near its endpoints. The arc $\overline{\partial_+\Omega}$ can never go horizontal or vertical in its interior, because otherwise there would be a Reeb orbit of type (iii) with $\nu = (1, 0)$ or $\nu = (0, 1)$, so that $\rho = 1$. Thus, X_Ω is strictly monotone. □

3. The first equivariant capacity

We now prove Theorem 1.11. (Some related arguments appeared in [12, Lem. 1.19].) If $a_1, \ldots, a_n > 0$, define the "L-shaped domain"

$$L(a_1, \ldots, a_n) = \left\{\mu \in \mathbb{R}^n_{\geq 0} \,\middle|\, \mu_j \leq a_j \text{ for some } j\right\}.$$

Lemma 3.1. *If $a_1, \ldots, a_n > 0$, then*

$$c_1^{\mathrm{CH}}\left(X_{L(a_1,\ldots,a_n)}\right) = \sum_{j=1}^n a_j.$$

Proof. Observe that

$$\mathbb{R}^n_{\geq 0} \setminus L(a_1, \ldots, a_n) = (a_1, \infty) \times \cdots \times (a_n, \infty).$$

is convex. Thus, $X_{L(a_1,\ldots,a_n)}$ satisfies all the conditions in the definition of "concave toric domain", except that it is not compact.

A formula for c_k^{CH} of a concave toric domain is given in [12, Thm. 1.14]. The $k = 1$ case of this formula asserts that if X_Ω is a concave toric domain in $\mathbb{R}^{2n}$, then

$$c_1^{\mathrm{CH}}(X_\Omega) = \min\left\{\sum_{i=1}^n \mu_i \,\middle|\, \mu \in \overline{\partial_+\Omega}\right\}. \tag{3.1}$$

By an exhaustion argument (see [12, Rmk. 1.3]), this result also applies to $X_{L(a_1,\ldots,a_n)}$. For $\Omega = L(a_1, \ldots, a_n)$, the minimum in (3.1) is realized by $\mu = (a_1, \ldots, a_n)$. □

Lemma 3.2. *If X_Ω is a monotone toric domain in $\mathbb{R}^{2n}$ and if $\mu \in \partial_+\Omega$, then $\Omega \subset L(\mu_1, \ldots, \mu_n)$.*

Proof. By an approximation argument, we can assume without loss of generality that X_Ω is strictly monotone. Then $\partial_+\Omega$ is the graph of a positive function f over an open set $U \subset \mathbb{R}^{n-1}_{\geq 0}$ with $\partial_j f < 0$ for $j = 1, \ldots, n-1$. It follows that if $(\mu'_1, \ldots, \mu'_{n-1}) \in U$ and $\mu'_j > \mu_j$ for all $j = 1, \ldots, n-1$, then $f(\mu'_1, \ldots, \mu'_{n-1}) < f(\mu_1, \ldots, \mu_{n-1})$. Consequently Ω does not contain any point μ' with $\mu'_j > \mu_j$ for all $j = 1, \ldots, n$. This means that $\Omega \subset L(\mu_1, \ldots, \mu_n)$. Figure 2 illustrates this inclusion for $n = 2$. □

Proof of Theorem 1.11.. For $a > 0$, consider the simplex

$$\Delta^n(a) = \left\{\mu \in \mathbb{R}^n_{\geq 0} \,\middle|\, \sum_{j=1}^n \mu_i \leq a\right\}.$$

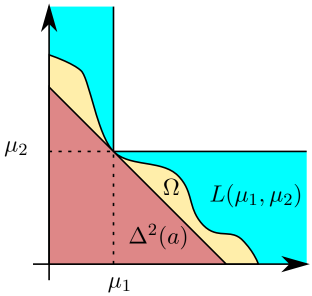


FIGURE 2. The inclusions $\Delta^n(a) \subset \Omega \subset L(\mu_1, \dots, \mu_n)$ for $n = 2$

Observe that the toric domain $X_{\Delta^n(a)}$ is the ball $B^{2n}(a)$. Now let $a > 0$ be the largest real number such that $\Delta^n(a) \subset \Omega$; see Fig. 2.

We have $B^{2n}(a) \subset X_\Omega$, so by definition $a \leq c_{\mathrm{Gr}}(X_\Omega)$. Since c_1^{CH} is a normalized symplectic capacity, $c_{\mathrm{Gr}}(X_\Omega) \leq c_1^{\mathrm{CH}}(X_\Omega)$. By the maximality property of a, there exists a point $\mu \in \overline{\partial_+\Omega}$ with $\sum_{j=1}^n \mu_j = a$. By an approximation argument we can assume that $\mu \in \partial_+\Omega$. By Lemma 3.2, $X_\Omega \subset X_{L(\mu_1,\dots,\mu_n)}$. By the monotonicity of c_1^{CH} and Lemma 3.1, we then have

$$c_1^{\mathrm{CH}}(X_\Omega) \leq c_1^{\mathrm{CH}}\left(X_{L(\mu_1,\dots,\mu_n)}\right) = \sum_{j=1}^n \mu_j = a.$$

Combining the above inequalities gives $c_{\mathrm{Gr}}(X_\Omega) = c_1^{\mathrm{CH}}(X_\Omega) = a$. □

4. ECH capacities

We now recall some facts about ECH capacities which we will use to prove Theorem 1.7.

Definition 4.1. A *weakly convex toric domain* in $\mathbb{R}^4$ is a compact toric domain $X_\Omega \subset \mathbb{R}^4$ such that Ω is convex, and $\overline{\partial_+\Omega}$ is an arc with one endpoint on the positive μ_1 axis and one endpoint on the positive μ_2 axis. See Figure 1d.

Theorem 4.2. (Cristofaro-Gardiner [7]) *In $\mathbb{R}^4$, let X_Ω be a concave toric domain, and let $X_{\Omega'}$ be a weakly convex toric domain. Then there exists a symplectic embedding* $\mathrm{int}(X_\Omega) \underset{s}{\hookrightarrow} X_{\Omega'}$ *if and only if $c_k^{\mathrm{ECH}}(X_\Omega) \leq c_k^{\mathrm{ECH}}(X_{\Omega'})$ for all $k \geq 0$.*

To make use of this theorem, we need some formulas to compute the ECH capacities c_k^{ECH}. To start, consider a 4-dimensional concave toric domain X_Ω. Associated to X_Ω is a "weight sequence" $W(X_\Omega)$, which is a finite or countable multiset of positive real numbers defined in [5], see also [23], as follows. Let r be the largest positive real number such that the triangle $\Delta^2(r) \subset \Omega$. We can write $\Omega \setminus \Delta^2(r) = \widetilde{\Omega}_1 \sqcup \widetilde{\Omega}_2$, where $\widetilde{\Omega}_1$ does not intersect the μ_2-axis and $\widetilde{\Omega}_2$ does not intersect the μ_1-axis. It is possible that $\widetilde{\Omega}_1$

and/or $\widetilde{\Omega}_2$ is empty. After translating the closures of $\widetilde{\Omega}_1$ or $\widetilde{\Omega}_2$ by $(-r,0)$ and $(0,-r)$ and multiplying them by the matrices $\begin{bmatrix}1&1\\0&1\end{bmatrix}$ and $\begin{bmatrix}1&0\\1&1\end{bmatrix}$, respectively, we obtain two new domains Ω_1 and Ω_2 in $\mathbb{R}^2_{\geq 0}$ such that X_{Ω_1} and X_{Ω_2} are concave toric domains. We then inductively define

$$W(X_\Omega) = (r) \cup W(X_{\Omega_1}) \cup W(X_{\Omega_2}), \tag{4.1}$$

where '$\cup$' denotes the union of multisets, and the term $W(X_{\Omega_i})$ is omitted if Ω_i is empty.

Let us call two subsets of $\mathbb{R}^2$ "affine equivalent" if one can be obtained from the other by the composition of a translation and an element of $\mathrm{GL}(2,\mathbb{Z})$. If $W(X_\Omega) = (a_1, a_2, \ldots)$, then the domain Ω is canonically decomposed into triangles, which are affine equivalent to the triangles $\Delta^2(a_1), \Delta^2(a_2), \ldots$ and which meet only along their edges; the first of these triangles is $\Delta^2(r)$. See [19, §3.1] for more details. We now recall the "Traynor trick":

Proposition 4.3. *[29] If $T \subset \mathbb{R}^2_{\geq 0}$ is a triangle affine equivalent to $\Delta^2(a)$, then there is a symplectic embedding* $\mathrm{int}(B^4(a)) \underset{s}{\hookrightarrow} X_{\mathrm{int}(T)}$.

As a result, there is a symplectic embedding

$$\coprod_i \mathrm{int}(B^4(a_i)) \subset X_\Omega.$$

Consequently, by the monotonicity property of ECH capacities, we have

$$c_k^{\mathrm{ECH}}\left(\coprod_i \mathrm{int}(B^4(a_i))\right) \leq c_k^{\mathrm{ECH}}(X_\Omega). \tag{4.2}$$

Theorem 4.4. [5] *If X_Ω is a four-dimensional concave toric domain with weight expansion $W(X_\Omega) = (a_1, a_2, \ldots)$, then equality holds in (4.2).*

To make this more explicit, we know from [17] that[6]

$$c_k^{\mathrm{ECH}}\left(\coprod_i \mathrm{int}(B^4(a_i))\right) = \sup_{k_1+\cdots=k} \sum_i c_{k_i}^{\mathrm{ECH}}(\mathrm{int}(B^4(a_i))) \tag{4.3}$$

and

$$c_k^{\mathrm{ECH}}(\mathrm{int}(B^4(a))) = c_k^{\mathrm{ECH}}(B^4(a)) = da, \tag{4.4}$$

where d is the unique nonnegative integer such that

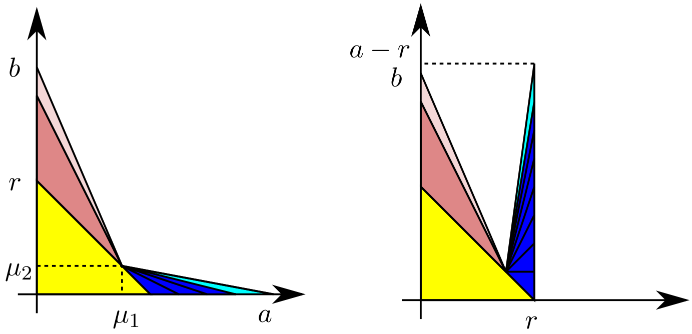
$$d^2 + d \leq 2k \leq d^2 + 3d.$$

To state the next lemma, given $a_1, a_2 > 0$, define the polydisk

$$P(a_1, a_2) = \left\{ z \in \mathbb{C}^2 \;\middle|\; \pi|z_1|^2 \leq a_1,\ \pi|z_2|^2 \leq a_2 \right\}.$$

This is a convex toric domain $X_{\Omega'}$ where Ω' is a rectangle of side lengths a_1 and a_2.

[6] For the sequence of numbers a_i coming from a weight expansion, or for any finite sequence, the supremum in (4.3) is achieved, so we can write 'max' instead of 'sup'.



(a) Weights of X_Ω **(b)** Ball packing into a polydisk

FIGURE 3. Embedding a concave toric domain into a polydisk

Lemma 4.5. *Let X_Ω be a four-dimensional concave toric domain. Let $(a, 0)$ and $(0, b)$ be the points where $\overline{\partial_+\Omega}$ intersects the axes. Let μ be a point on $\overline{\partial_+\Omega}$ minimizing $\mu_1 + \mu_2$, and write $r = \mu_1 + \mu_2$. Then there exists a symplectic embedding*

$$\operatorname{int}(X_\Omega) \underset{s}{\hookrightarrow} P(r, \max(b, a - r)).$$

Proof. One might hope for a direct construction using some version of "symplectic folding" [24], but we will instead use the above ECH machinery. By Theorem 4.2, it is enough to show that

$$c_k^{\mathrm{ECH}}(X_\Omega) \le c_k^{\mathrm{ECH}}(P(r, \max(b, a - r))) \tag{4.5}$$

for each nonnegative integer k.

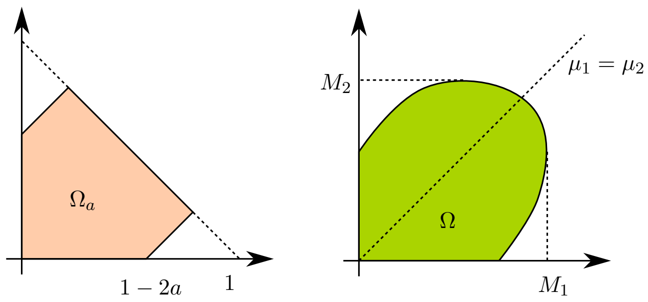
Consider the weight expansion $W(X_\Omega) = (a_1, a_2, \ldots)$ where $a_1 = r$. The decomposition of Ω into triangles corresponding to the weight expansion consists of the triangle $\Delta^2(r)$, plus some additional triangles in the triangle with corners $(0, r), (\mu_1, \mu_2), (0, b)$, plus some additional triangles in the triangle with corners $(\mu_1, \mu_2), (r, 0), (a, 0)$; see Fig. 3a. The latter triangle is affine equivalent to the triangle with corners $(\mu_1, \mu_2), (r, 0), (r, a-r)$; see Fig. 3b. This allows us to pack triangles affine equivalent to $\Delta^2(a_1), \Delta^2(a_2), \ldots$ into the rectangle with horizontal side length r and vertical side length $\max(b, a - r)$. Thus, by the Traynor trick, we have a symplectic embedding

$$\coprod_i \operatorname{int}(B(a_i)) \underset{s}{\hookrightarrow} P(r, \max(b, a - r)).$$

Then Theorem 4.4 and the monotonicity of ECH capacities imply (4.5). $\square$

Proof of Theorem 1.7. Let r be the largest positive real number such that $\Delta^2(r) \subset \Omega$. We have $B^4(r) \subset X_\Omega$, so $r \le c_{\mathrm{Gr}}(X_\Omega)$, and we just need to show that $c_Z(X_\Omega) \le r$.

Let μ be a point on $\partial_+\Omega$ such that $\mu_1 + \mu_2 = r$. By an approximation argument, we can assume that X_Ω is strictly monotone, so that the tangent



(a) The domain Ω_a **(b)** A domain to which Theorem 5.6 applies

FIGURE 4. Some domains

line to $\partial_+\Omega$ at μ is not horizontal or vertical. Then we can find $a, b > r$ such that Ω is contained in the quadrilateral with vertices $(0,0)$, $(a,0)$, (μ_1, μ_2), and $(0,b)$. It then follows from Lemma 4.5 that there exists a symplectic embedding $\operatorname{int}(X_\Omega) \underset{s}{\hookrightarrow} P(r,R)$ for some $R > 0$. Since $P(r,R) \subset Z^4(r)$, it follows that $c_Z(X_\Omega) \leq r$. $\square$

5. A family of non-monotone toric examples

We now study a family of examples of non-monotone toric domains, and we determine when they satisfy the conclusions of Conjecture 1.1 or Conjecture 1.2.

For $0 < a < 1/2$, let Ω_a be the convex polygon with corners $(0,0)$, $(1-2a, 0)$, $(1-a, a)$, $(a, 1-a)$ and $(0, 1-2a)$, and write $X_a = X_{\Omega_a}$; see Fig. 4a. Then X_a is a weakly convex (but not monotone) toric domain.

Proposition 5.1. *Let $0 < a < 1/2$. Then the Gromov width and cylindrical capacity of X_a are given by*

$$c_{\mathrm{Gr}}(X_a) = \min(1-a, 2-4a), \tag{5.1}$$

$$c_Z(X_a) = 1-a. \tag{5.2}$$

Corollary 5.2. *Let $0 < a < 1/2$ and let X_a be as above. Then:*

(a) The conclusion of Conjecture 1.1 holds for X_a, i.e., all normalized symplectic capacities defined for X_a agree, if and only if $a \leq 1/3$.

(b) The conclusion of Conjecture 1.2 holds for X_a, i.e., every normalized symplectic capacity c defined for X_a satisfies $c(X_a) \leq \sqrt{2\operatorname{Vol}(X_a)}$, if and only if $a \leq 2/5$.

Proof of Corollary 5.2. (a) By Lemma 1.3, we need to check that $c_{\mathrm{Gr}}(X_a) = c_Z(X_a)$ if and only if $a \leq 1/3$. This follows directly from (5.1) and (5.2).

(b) Since c_Z is the largest normalized symplectic capacity, the conclusion of Conjecture 1.2 holds for X_a if and only if

$$c_Z(X_a) \leq \sqrt{2\operatorname{Vol}(X_a)}. \tag{5.3}$$

By equation (1.2), we have

$$\mathrm{Vol}(X_{\Omega_a}) = \frac{1-4a^2}{2}.$$

It follows from this and (5.2) that (5.3) holds if and only if $a \le 2/5$. □

Remark 5.3. To recap, the conclusion of Conjecture 1.1 holds if and only if the ratio $c_Z/c_{\mathrm{Gr}} = 1$, and the conclusion of Conjecture 1.2 holds if and only if the ratio $c_Z^n/(n!\,\mathrm{Vol}) \le 1$. The above calculations show that both of these ratios for X_a go to infinity as $a \to 1/2$.

To prove Proposition 5.1, we will use the following formula for the ECH capacities of a weakly convex toric domain X_Ω. Let r be the smallest positive real number such that $\Omega \subset \Delta^2(r)$. Then $\Delta^2(r) \setminus \Omega = \widetilde{\Omega}_1 \sqcup \widetilde{\Omega}_2$ where $\widetilde{\Omega}_1$ does not intersect the μ_2-axis, and $\widetilde{\Omega}_2$ does not intersect the μ_1-axis. It is possible that $\widetilde{\Omega}_1$ and/or $\widetilde{\Omega}_2$ is empty. As in the discussion preceding (4.1), the closures of $\widetilde{\Omega}_1$ and $\widetilde{\Omega}_2$ are affine equivalent to domains Ω_1 and Ω_2 such that X_{Ω_1} and X_{Ω_2} are concave toric domains. Denote the union (as multisets) of their weight sequences by

$$W(X_{\Omega_1}) \cup W(X_{\Omega_2}) = (a_1, \ldots).$$

We then have:

Theorem 5.4. (Choi–Cristofaro-Gardiner [7]) *If X_Ω is a four-dimensional weakly convex toric domain as above, then*

$$c_k^{\mathrm{ECH}}(X_\Omega) = \inf_{l \ge 0} \left\{ c_{k+l}^{\mathrm{ECH}}\left(B^4(r)\right) - c_l^{\mathrm{ECH}}\left(\coprod_i B^4(a_i)\right)\right\}. \tag{5.4}$$

We need one more lemma, which follows from [21, Cor. 4.2]:

Lemma 5.5. *Let $\mu_1, \mu_2 \ge a > 0$. Let Ω be the "diamond" in $\mathbb{R}^2_{\ge 0}$ given by the convex hull of the points $(\mu_1 \pm a, \mu_2)$ and $(\mu_1, \mu_2 \pm a)$. Then there is a symplectic embedding*

$$\mathrm{int}(B^4(2a)) \underset{s}{\hookrightarrow} X_\Omega.$$

Proof of Proposition 5.1. To prove (5.1), we first describe the ECH capacities of X_a. In the formula (5.4) for X_a, we have $r = 1$, while the weight expansions of Ω_1 and Ω_2 are both (a, a); the corresponding triangles are shown in Figure 5(b). Thus, by Theorem 5.4 and equation (4.3), we have

$$c_k^{\mathrm{ECH}}(X_a) = \inf_{l_1,\ldots,l_4 \ge 0} \left\{ c_{k+l_1+l_2+l_3+l_4}^{\mathrm{ECH}}\left(B^4(1)\right) - \sum_{i=1}^4 c_{l_i}^{\mathrm{ECH}}\left(B^4(a)\right)\right\}. \tag{5.5}$$

We also note from (4.4) that

$$c_1^{\mathrm{ECH}}(B^4(r)) = c_2^{\mathrm{ECH}}(B^4(r)) = r, \qquad c_5^{\mathrm{ECH}}(B^4(r)) = 2r.$$

Taking $k = 1$ and $(l_1, \ldots, l_4) = (1, 0, 0, 0)$ in equation (5.5), we get

$$c_1^{\mathrm{ECH}}(X_{\Omega_a}) \le 1 - a. \tag{5.6}$$

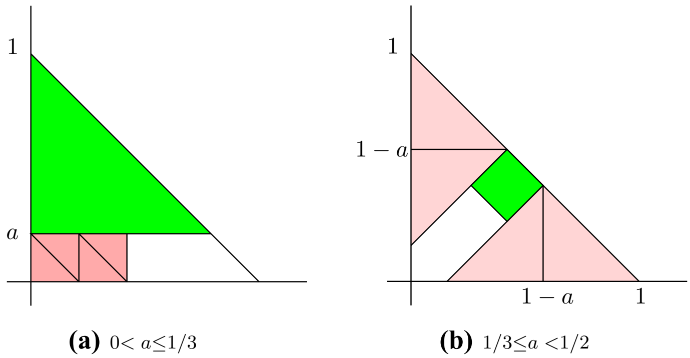


FIGURE 5. Ball packings

Taking $k = 1$ and $(l_1, \dots, l_4) = (1, 1, 1, 1)$ in equation (5.5), we get

$$c_1^{\mathrm{ECH}}(X_{\Omega_a}) \le 2 - 4a. \tag{5.7}$$

By (5.6) and (5.7) and the fact that c_1^{ECH} is a normalized symplectic capacity, we conclude that

$$c_{\mathrm{Gr}}(X_{\Omega_a}) \le \min(1 - a, 2 - 4a). \tag{5.8}$$

To prove the reverse inequality to (5.8), suppose first that $0 < a \le 1/3$. It is enough to prove that there exists a symplectic embedding $\operatorname{int}(B^4(1 - a)) \underset{s}{\hookrightarrow} X_{\Omega_a}$. By Theorem 4.2, it is enough to show that

$$c_k^{\mathrm{ECH}}(B^4(1 - a)) \le c_k^{\mathrm{ECH}}(X_{\Omega_a})$$

for all nonnegative integers k. By equation (5.5), the above inequality is equivalent to

$$c_k^{\mathrm{ECH}}(B^4(1 - a)) + \sum_{i=1}^{4} c_{l_i}^{\mathrm{ECH}}(B^4(a)) \le c_{k+l_1+l_2+l_3+l_4}^{\mathrm{ECH}}(B^4(1)) \tag{5.9}$$

for all nonnegative integers $k, l_1, \dots, l_4 \ge 0$. To prove (5.9), by the monotonicity of ECH capacities and the disjoint union formula (4.3), it suffices to find a symplectic embedding

$$\operatorname{int}\left(B^4(1 - a) \sqcup \coprod_4 B^4(a)\right) \underset{s}{\hookrightarrow} B^4(1).$$

This embedding exists by the Traynor trick (Proposition 4.3) using the triangles shown in Figure 5(a).

Finally, when $1/3 \le a < 1/2$, it is enough to show that there exists a symplectic embedding $\operatorname{int}(B^4(2 - 4a)) \underset{s}{\hookrightarrow} X_{\Omega_a}$. This exists by Lemma 5.5 using the diamond shown in Figure 5(b).

This completes the proof of (5.1). Equation (5.2) follows from Theorem 5.6 below. □

Theorem 5.6. *Let $X_\Omega \subset \mathbb{R}^4$ be a weakly convex toric domain, see Definition 4.1. For $j = 1, 2$, let*

$$M_j = \max\{\mu_j \mid \mu \in \Omega\}.$$

Assume that there exists $(M_1, \mu_2) \in \overline{\partial_+\Omega}$ with $\mu_2 \le M_1$, and that there exists $(\mu_1, M_2) \in \overline{\partial_+\Omega}$ with $\mu_1 \le M_2$. Then

$$c_Z(X_\Omega) = \min(M_1, M_2).$$

That is, under the hypotheses of the theorem, see Figure 4b, an optimal symplectic embedding of X_Ω into a cylinder is given by the inclusion of X_Ω into $(\pi|z_1|^2 \le M_1)$ or $(\pi|z_2|^2 \le M_2)$.

Proof. From the above inclusions we have $c_Z(X_\Omega) \le \min(M_1, M_2)$. To prove the reverse inequality, suppose that there exists a symplectic embedding

$$X_\Omega \underset{s}{\hookrightarrow} Z^4(R). \tag{5.10}$$

We need to show that $R \ge \min(M_1, M_2)$. To do so, we will use ideas[7] from [18].

Let $\varepsilon > 0$ be small. Let $(A, 0)$ and $(0, B)$ denote the endpoints of $\overline{\partial_+\Omega}$. By an approximation argument, we can assume that $\overline{\partial_+\Omega}$ is smooth, and that $\partial_+\Omega$ has positive slope less than ε near $(A, 0)$ and slope greater than ε^{-1} near $(0, B)$. As in the proof of Proposition 1.8, there are then three types of Reeb orbits on ∂X_Ω:

(i) There is a simple Reeb orbit whose image is the circle with $\pi|z_1|^2 = A$ and $z_2 = 0$. This Reeb orbit has symplectic action (period) equal to A, and rotation number $1 - \varepsilon^{-1}$.

(ii) There is a simple Reeb orbit whose image is the circle with $z_1 = 0$ and $\pi|z_2|^2 = B$. This Reeb orbit has symplectic action B and rotation number $1 - \varepsilon^{-1}$.

(iii) For each point $\mu \in \partial_+\Omega$ where $\partial_+\Omega$ has rational slope, there is an S^1 family of simple Reeb orbits in the torus where $\pi(|z_1|^2, |z_2|^2) = \mu$. If $\nu = (\nu_1, \nu_2)$ is the outward normal vector to $\partial_+\Omega$ at μ, scaled so that ν_1, ν_2 are relatively prime integers, then these Reeb orbits have rotation number $\nu_1 + \nu_2$ and symplectic action $\mu \cdot \nu$. See [12, §2.2].

We claim now that:

(*) Every Reeb orbit on ∂X_Ω with positive rotation number has symplectic action at least $\min(M_1, M_2)$.

To prove this claim, we only need to check the type (iii) simple Reeb orbits where $\nu_1 + \nu_2 \ge 1$. For such an orbit we must have $\nu_1 \ge 1$ or $\nu_2 \ge 1$. Suppose first that $\nu_1 \ge 1$. By the hypotheses of the theorem there exists μ_2' such that

[7] The main theorem in [18] gives a general obstruction to a symplectic embedding of one four-dimensional convex toric domain into another, which sometimes goes beyond the obstruction coming from ECH capacities. This theorem can be generalized to weakly convex toric domains; but rather than carry out the full generalization, we will just explain the simple case of this that we need.

$(M_1, \mu_2') \in \overline{\partial_+ \Omega}$ and $M_1 \geq \mu_2'$. Since Ω is convex and ν is an outward normal at μ, the symplectic action

$$\mu \cdot \nu \geq (M_1, \mu_2') \cdot \nu = M_1 + (\nu_1 - 1)(M_1 - \mu_2') + (\nu_1 + \nu_2 - 1)\mu_2' \geq M_1.$$

Likewise, if $\nu_2 \geq 1$, then the symplectic action $\mu \cdot \nu \geq M_2$.

As in [18, §5.3], starting from the symplectic embedding (5.10), by replacing X_Ω with an appropriate subset and replacing $Z^4(R)$ with an appropriate superset, we obtain a symplectic embedding $X' \underset{s}{\hookrightarrow} \mathrm{int}(Z')$, where:

- Z' is an ellipsoid whose boundary has one simple Reeb orbit γ_+ with symplectic action $\mathcal{A}(\gamma_+) = R + \varepsilon$ and Conley–Zehnder index $\mathrm{CZ}(\gamma_+) = 3$, another simple Reeb orbit with very large symplectic action, and no other simple Reeb orbits.
- X' is a (non-toric) star-shaped domain with smooth boundary, all of whose Reeb orbits are nondegenerate. Every Reeb orbit on $\partial X'$ with rotation number greater than or equal to 1 has action at least $\min(M_1, M_2) - \varepsilon$.

The symplectic embedding gives rise to a strong symplectic cobordism W whose positive boundary is $\partial Z'$ and whose negative boundary is $\partial X'$. The argument in [18, §6] shows that for a generic "cobordism-admissible" almost complex structure J on the "completion" of W, there exists an embedded J-holomorphic curve u with one positive end asymptotic to the Reeb orbit γ_+ in $\partial Z'$, negative ends asymptotic to some Reeb orbits $\gamma_1, \ldots, \gamma_m$ in $\partial X'$, and Fredholm index $\mathrm{ind}(u) = 0$. The Fredholm index is computed by the formula

$$\mathrm{ind}(u) = 2g + [\mathrm{CZ}(\gamma_+) - 1] - \sum_{i=1}^{m} [\mathrm{CZ}(\gamma_i) - 1] \tag{5.11}$$

where g denotes the genus of u. Furthermore, since J-holomorphic curves decrease symplectic action, we have

$$\mathcal{A}(\gamma_+) \geq \sum_{i=1}^{m} \mathcal{A}(\gamma_i). \tag{5.12}$$

We claim now that at least one of the Reeb orbits γ_i has action at least $\min(M_1, M_2) - \varepsilon$. Then the inequality (5.12) gives

$$R + \varepsilon \geq \min(M_1, M_2) - \varepsilon,$$

and since $\varepsilon > 0$ was arbitrarily small, we are done.

To prove the above claim, suppose to the contrary that all of the Reeb orbits γ_i have action less than $\min(M_1, M_2) - \varepsilon$. Then all the Reeb orbits γ_i have rotation number $\rho(\gamma_i) < 1$, which means that they all have Conley–Zehnder index $\mathrm{CZ}(\gamma_i) \leq 1$. It now follows from (5.11) that $\mathrm{ind}(u) \geq 2$, which is a contradiction[8]. □

[8] One way to think about the information that we are getting out of (5.11), as well as the general symplectic embedding obstruction in [18], is that we are making essential use of the fact that every holomorphic curve has nonnegative genus.

6. The first Ekeland–Hofer capacity

The goal of this section is to (re)prove the following theorem. This is well-known in the community and is attributed to Ekeland, Hofer and Zehnder [9,15]. It was first mentioned by Viterbo in [30, Proposition 3.10].

Theorem 6.1. (Ekeland–Hofer–Zehnder) *Let $W \subset \mathbb{R}^{2n}$ be a compact convex domain with smooth boundary. Then*

$$c_1^{\mathrm{EH}}(W) = A_{\min}(W).$$

We start by recalling the definition of the first Ekeland-Hofer capacity c_1^{EH}. Let $E = H^{1/2}(S^1, \mathbb{R}^{2n})$. That is, if $x \in L^2(S^1, \mathbb{R}^{2n})$ is written as a Fourier series $x = \sum_{k\in\mathbb{Z}} e^{2\pi ikt} x_k$ where $x_k \in \mathbb{R}^{2n}$, then

$$x \in E \iff \sum_{k\in\mathbb{Z}} |k||x_k|^2 < \infty.$$

Recall that there is an orthogonal splitting $E = E^+ \oplus E^0 \oplus E^-$ and orthogonal projections $P^\circ : E \to E^\circ$ where $\circ = +, 0, -$. The symplectic action of $x \in E$ is defined to be

$$A(x) = \frac{1}{2}\left(\|P^+x\|^2_{H^{1/2}} - \|P^-x\|^2_{H^{1/2}}\right).$$

It follows from a simple calculation that if x is smooth, then $A(x) = \int_x \lambda_0$, where λ_0 denotes the standard Liouville form on $\mathbb{R}^{2n}$.

Let $\mathcal{H}$ denote the set of $H \in C^\infty(\mathbb{R}^{2n})$ such that

- $H|_U \equiv 0$ for some $U \subset \mathbb{R}^{2n}$ open,
- $H(z) = c|z|^2$ for $z >> 0$ where $c \notin \{\pi, 2\pi, 3\pi, \dots\}$.

For $H \in \mathcal{H}$, the action functional $\mathcal{A}_H : H^{1/2}(S^1, \mathbb{R}^{2n}) \to \mathbb{R}$ is defined by

$$\mathcal{A}_H(x) = A(x) - \int_0^1 H(x(t))dt. \tag{6.1}$$

Note that the natural action of S^1 on itself induces an S^1-action on E. Let Γ be the set of homeomorphisms $h : E \to E$ such that h can be written as

$$h(x) = e^{\gamma_+(x)}P^+x + P^0x + e^{\gamma_-(x)}P^-x + K(x),$$

where $\gamma_+, \gamma_- : E \to \mathbb{R}$ are continuous, S^1-invariant and map bounded sets to bounded sets, and $K : E \to E$ is continuous, S^1-equivariant and maps bounded sets to precompact sets. Let S^+ denote the unit sphere in E^+ with respect to the $H^{1/2}$ norm. The first Ekeland-Hofer capacity is defined in [9] by

$$c_1^{\mathrm{EH}}(W) = \inf\{c_{H,1} \mid H \in \mathcal{H}, W \subset \operatorname{supp} H\},$$

where

$$c_{H,1} = \inf\{\sup \mathcal{A}_H(\xi) \mid \xi \subset E \text{ is } S^1- \text{ invariant, and } \forall h \in \Gamma : h(\xi) \cap S^+ \neq \emptyset\}.$$

Proof of Theorem 6.1. Since W is star-shaped, there is a unique differentiable function $r : \mathbb{R}^{2n} \to \mathbb{R}$ which is C^∞ in $\mathbb{R}^{2n} \setminus \{0\}$ satisfying $r(cz) = c^2 r(z)$ for $c \geq 0$ such that

$$\begin{aligned} W &= \{z \in \mathbb{R}^{2n} \mid r(z) \leq 1\}, \\ \partial W &= \{z \in \mathbb{R}^{2n} \mid r(z) = 1\}. \end{aligned}$$

Let $\alpha = A_{\min}(W)$ and fix $\varepsilon > 0$. Let $f \in C^\infty_{\geq 0}(\mathbb{R})$ be a convex function such that $f(r) = 0$ for $r \leq 1$ and $f(r) = Cr - (\alpha + \varepsilon)$ for $r \geq 2$ for some constant $C > \alpha$. In particular,

$$f(r) \geq Cr - (\alpha + \varepsilon), \quad \text{for all } r. \tag{6.2}$$

We now choose a convex function $H \in C^\infty(\mathbb{R}^{2n})$ such that

$$\begin{array}{ll} H(z) = f(r(z)), & \text{if } r(z) \leq 2, \\ H(z) \geq f(r(z)), & \text{for all } z \in \mathbb{R}^{2n}, \\ H(z) = c\,|z|^2, & \text{if } z >> 0 \text{ for some } c \in \mathbb{R}_{>0} \setminus \pi\mathbb{Z}. \end{array} \tag{6.3}$$

Let $x_0 \in E$ be an action-minimizing Reeb orbit on ∂W, reparametrized as a map $x_0 : \mathbb{R}/\mathbb{Z} = S^1 \to \mathbb{R}^{2n}$ of speed α, so that $A(x_0) = \alpha$ and $r(x_0) \equiv 1$ and $\dot{x}_0 = \alpha J \nabla r(x_0)$. From a simple calculation we deduce that x_0 is a critical point of the functional $\Psi : E \to \mathbb{R}$ defined by

$$\Psi(x) = A(x) - \alpha \int_0^1 r(x(t))\, dt. \tag{6.4}$$

Observe that $\Psi(cx) = c^2 \Psi(x)$ for $c \geq 0$. So sx_0 is a critical point of Ψ for all $s \geq 0$. Let $\xi = [0, \infty) \cdot P^+ x_0 \oplus E^0 \oplus E^-$.

We now claim that $\Psi(x) \leq 0$ for all $x \in \xi$. To prove this, let $\xi_s = sP^+ x_0 \oplus E^0 \oplus E^-$. Observe that $\Psi|_{\xi_s}$ is a concave function. Since sx_0 is a critical point of $\Psi|_{\xi_s}$ it follows that $\max \Psi(\xi_s) = \Psi(sx_0) = s^2 \Psi(x_0) = 0$.

From (6.1), (6.2), (6.3) and (6.4), we obtain

$$\mathcal{A}_H(x) \leq \Psi(x) + \alpha + \varepsilon + (C - \alpha) \int_0^1 r(x(t))\, dt \leq \alpha + \varepsilon.$$

Note that ξ is S^1-invariant. Moreover, it is proven in [8] that $h(\xi) \cap S^+ \neq \emptyset$ for all $h \in \Gamma$. So $c_{H,1} \leq \alpha + \varepsilon$. Hence $c_1^{\mathrm{EH}}(W) \leq \alpha + \varepsilon$ for all $\varepsilon > 0$. Therefore,

$$c_1^{\mathrm{EH}}(W) \leq \alpha.$$

To prove the reverse inequality, recall from [9, Prop. 2] that $c_1^{\mathrm{EH}}(W)$ is the symplectic action of some Reeb orbit on ∂W. Thus,

$$c_1^{\mathrm{EH}}(W) \geq \alpha.$$

$\square$

Acknowledgements

We thank A. Oancea, Y. Ostrover and M. Usher for useful discussions, J. Kang and E. Shelukhin for bringing some parts of the literature to our attention and F. Schlenk for detailed comments on an earlier version of this paper.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Abbondandolo, A., Bramham, B., Hryniewicz, U., Salamão, P.: Systolic ratio, index of closed orbits and convexity for tight contact forms on the three-sphere. Compos. Math. **154**, 2643–2680 (2018)

[2] Abbondandolo, A., Kang, J.: Symplectic homology of convex domains and Clarke's duality. arXiv:1907.07779

[3] Artstein-Avidan, S., Ostrover, Y.: Bounds for Minkowski billiard trajectories in convex bodies. IMRN **2014**, 165–193

[4] Artstein-Avidan, S., Karasev, R., Ostrover, Y.: From symplectic measurements to the Mahler conjecture. Duke Math. J. **163**, 2003–2022 (2014)

[5] Choi, K., Cristofaro-Gardiner, D., Frenkel, D., Hutchings, M., Ramos, V.G.B.: Symplectic embeddings into four-dimensional concave toric domains. J. Topol. **7**, 1054–1076 (2014)

[6] Cieliebak, K., Hofer, H., Latschev, J., Schlenk, F.: Quantitative symplectic geometry, in dynamics, ergodic theory, and geometry. Math. Sci. Res. Inst. Publ. bf **54**, 1–44 (2007)

[7] Cristofaro-Gardiner, D.: Symplectic embeddings from concave toric domains into convex ones. J. Differ. Geom. **112**, 199–232 (2019)

[8] Ekeland, I., Hofer, H.: Symplectic topology and Hamiltonian dynamics. Math. Z. **200**, 355–378 (1989)

[9] Ekeland, I., Hofer, H.: Symplectic topology and Hamiltonian dynamics II. Math. Z. **203**, 553–567 (1990)

[10] Ginzburg, V., Shon, J.: On the filtered symplectic homology of prequantization bundles. Int. J. Math. **29**, 1850071, 35pp (2018)

[11] Gromov, M.: Pseudoholomorphic curves in symplectic manifolds. Invent. Math. **82**, 307–347 (1985)

[12] Gutt, J., Hutchings, M.: Symplectic capacities from positive S^1-equivariant symplectic homology, Algebr. Geom. Topol. **18**, 3537–3600 (2018)

[13] Hermann, D.: Non-equivalence of symplectic capacities for open sets with restricted contact type boundary. www.math.u-psud.fr/~biblio/pub/1998/abs/ppo1998_32.html

[14] Hofer, H., Wysocki, K., Zehnder, E.: The dynamics on three-dimensional strictly convex energy surfaces. Ann. Math. **148**, 197–289 (1998)

[15] Hofer, H., Zehnder, E.: Periodic solutions on hypersurfaces and a result by C. Viterbo. Invent. Math. **90**, 1–9 (1987)

[16] Hofer, H., Zehnder, E.: A new capacity for symplectic manifolds, Analysis, et cetera, 405–427. Academic Press, Boston MA (1990)

[17] Hutchings, M.: Quantitative embedded contact homology. J. Differ. Geom. **88**, 321–366 (2011)

[18] Hutchings, M.: Beyond ECH capacities. Geom. Topol. **20**, 1085–1126 (2016)

[19] Hutchings, M.: ECH capacities and the Ruelle invariant (to appear in this same Festschrift volume of Journal of Fixed Point Theory and Applications). arXiv:1910.08260

[20] Irie, K.: Symplectic homology of fiberwise convex sets and homology of loop spaces. arXiv:1907.09749

[21] Latschev, J., McDuff, D., Schlenk, F.: The Gromov width of 4-dimensional tori. Geom. Topol. **17**, 2813–1853 (2013)

[22] Mahler, K.: Ein Übertragungsprinzip für konvexe Körper. Cas. Mat. Fys. **68**, 93–102 (1939)

[23] Ramos, V.G.B.: Symplectic embeddings and the Lagrangian bidisk. Duke Math. J. **166**, 1703–1738 (2017)

[24] Schlenk, F.: On symplectic folding. arXiv:math/9903086

[25] Schlenk, F.: Symplectic embedding problems, old and new. Bull. Am. Math. Soc. **55**, 139–182 (2017)

[26] Shi, K., Lu, G.: Some cases of the Viterbo conjecture and the Mahler one. arXiv:2008.04000

[27] Siegel, K.: Higher symplectic capacities. Israel J. Math. **245**(1), 27–38 (2021)

[28] Siegel, K.: Computing higher symplectic capacities I. arXiv:1911.06466

[29] Traynor, L.: Symplectic packing constructions. J. Differ. Geom. **42**, 411–429 (1995)

[30] Viterbo, C.: Capacités symplectiques et applications (d'après Ekeland-Hofer, Gromov), Séminaire Bourbaki, volume 1988/89. Astérisque No. **177**–**178**, 345–362 (1989)

[31] Viterbo, C.: Functors and computations in Floer homology with applications I Geom. Funct. Anal. **9**, 985–1033 (1999)

[32] Viterbo, C.: Metric and isoperimetric problems in symplectic geometry. J. Am. Math. Soc. **13**, 411–431 (2000)

Jean Gutt
Université Toulouse III - Paul Sabatier
118 route de Narbonne
31062 Toulouse Cedex 9
France

and

Institut National Universitaire Champollion
Place de Verdun
81012 Albi
France
e-mail: `jean.gutt@univ-jfc.fr`

Michael Hutchings
University of California
Berkeley CA
USA

Vinicius G. B. Ramos
Instituto de Matemática Pura e Aplicada
Estrada Dona Castorina, 110
Rio de Janeiro RJ22460-320
Brasil

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:45
https://doi.org/10.1007/s11784-022-00950-z
Published online May 24, 2022

Journal of Fixed Point Theory and Applications

Global surfaces of section with positive genus for dynamically convex Reeb flows

Umberto L. Hryniewicz, Pedro A. S. Salomão and Richard Siefring

Dedicated to Prof. Claude Viterbo on the occasion of his 60th birthday.

Abstract. We establish some new existence results for global surfaces of section of dynamically convex Reeb flows on the three-sphere. These sections often have genus, and are the result of a combination of pseudoholomorphic methods with some elementary ergodic methods.

Mathematics Subject Classification. 32Q65, 37J55, 37C27 and 53D42.

1. Introduction and main results

Let (z_0, z_1) be complex coordinates in $\mathbb{C}^2$, $S^3 = \{|z_0|^2 + |z_1|^2 = 1\}$, and α_0 be the standard Liouville form $(-i/4)\ \Sigma_j\ \bar{z}_j \mathrm{d}z_j - z_j \mathrm{d}\bar{z}_j$. The standard symplectic form on $\mathbb{C}^2$ is $\omega_0 = \mathrm{d}\alpha_0$. The fibers of the Hopf fibration are the periodic Reeb orbits of the contact form λ_0 on S^3 induced by α_0. Let us call the Reeb flow of λ_0 the *Hopf flow*. The contact structure $\xi_0 = \ker \lambda_0$ is called *standard*.

The contact form λ_0 is the first example of a *dynamically convex* contact form. In S^3, a contact form λ is said to be dynamically convex if all periodic orbits have Conley–Zehnder index ≥ 3 when computed in a global $\mathrm{d}\lambda$-symplectic frame of $\ker \lambda$. This notion was introduced by Hofer, Wysocki and Zehnder (HWZ) in [21].

One can show quite explicitly that all finite collections of periodic orbits of the Hopf flow span some global surface of section, see [2]. It is natural to ask if this property remains true for all dynamically convex Reeb flows on S^3, in particular for all strictly convex energy levels in $(\mathbb{C}^2, \omega_0)$ ([21, Theorem 3.4]). This might be too ambitious to try to prove, and one may be led to naively think that it is easy to find a counterexample. There is, however, another

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

natural way to generalize this property of the Hopf flow to the Reeb flows of all dynamically convex contact forms on S^3. Since the Hopf fibers are unknotted with self-linking number -1, one might ask if all finite collections of periodic Reeb orbits of this kind span some global surface of section. This is our first result.

Theorem 1.1. *Let L be any link formed by periodic Reeb orbits of a dynamically convex contact form on S^3 whose components are unknotted with self-linking number -1. Then, L bounds a global surface of section for the Reeb flow.*

There are no hidden genericity assumptions on the contact form. The genus of these sections will typically explode with the number of boundary orbits. Moreover, there is no need to specify the contact structure since only the standard one can be defined by a dynamically convex contact form on S^3, see [19].

A proof relying exclusively on pseudoholomorphic curves would be complicated by the fact, originally observed in [14], that transversality fails for curves with genus which are everywhere transverse to the flow. The solution proposed to this problem in [14] is to consider a perturbation of the holomorphic curve equation which corrects the transversality problem, but seriously complicates the compactness theory (see [1,6,7]). However, dealing with genus is unavoidable since the links covered by Theorem 1.1 typically have positive Seifert genus. A proof without pseudoholomorphic curves seems out of reach since dynamical convexity is an assumption only on the periodic orbits, and holomorphic curve techniques have proven to be one of the very few—if not the only—effective methods for finding surfaces of section under assumptions of this kind.

Let us outline the argument. The main step is the result from [24] stating that every component of a link L as in Theorem 1.1 bounds a disk-like global surface of section. At this point ergodic methods come to aid via asymptotic cycles. We use the statement from [26] refining a celebrated result due to Fried [11]. The disks can be used to check the hypotheses of [26, Theorem 1.3]. Each disk has uniformly bounded return time, hence all invariant measures in $S^3\setminus L$ positively hit the sum of the cohomology classes dual to each disk. Finally, positivity of rotation numbers follows from dynamical convexity.

Remark 1.2. It was explained to us by Colin, Dehornoy and Rechtman that the input from pseudoholomorphic curves from [24] can be used in a more elementary way, avoiding asymptotic cycles. One can take the union of the disks and "resolve intersections" to construct the desired sections. This idea is extensively used in [4].

In [12], Ghys introduced the notion of *right- (and left-) handed* vector field on a homology three-sphere, and explained that all finite collections of periodic orbits of such a vector field span a global surface of section. The Hopf flow is the simplest example of a right-handed vector field. Examples of left-handed geodesic flows on negatively curved two-dimensional orbifolds are presented by Dehornoy [5]. Right-handedness provides deep insight on the dynamics. For instance, it follows that every finite collection of periodic orbits is a fibered link, hence there are strong knot theoretical restrictions. Moreover, as soon as such a collection is "misplaced", then Nielsen–Thurston theory might be used to obtain entropy via the study of the isotopy class of the return map.

Question 1. *Is the Reeb flow of every dynamically convex contact form on S^3 right-handed?*

A positive answer is probably very hard to obtain, even in finite-dimensional families of interesting flows such as those appearing in Celestial Mechanics. One is then tempted to look for examples to give a negative answer, but they might not exist. In the context of the 3-body problem, we refer to [35] for a discussion of a version of this question, and to the book [10] by Frauenfelder and van Koert for a discussion on global surfaces of section, including a related conjecture of Birkhoff. The existence of genus zero global surfaces of section with prescribed binding orbits has been clarified in [31]. In [9], it is shown that geodesic flows on S^2 with curvatures pinched by some explicit constant lift to right-handed Reeb flows on S^3.

Remark 1.3. The dynamical convexity assumption is essential in Question 1, as one can easily check. It is, however, more subtle to rule out specific types of global surfaces of sections when dynamical convexity is dropped. For instance, in [33] one finds examples of contact forms on S^3 without disk-like global surfaces of section. The situation in higher dimensions is still wide open, but in [34], there are interesting new constructions for the spatial circular restricted 3-body problem.

Our second result is closely connected to the following question.

Question 2. *(HWZ [21]) Is the minimal period among closed Reeb orbits of a dynamically convex contact form on S^3 equal to the contact area of some disk-like global surface of section ?*

Starting from a nondegenerate dynamically convex contact form on S^3, Hutchings and Nelson [32] were able to implement the construction of the chain complex of Cylindrical Contact Homology (CCH), originally introduced by Eliashberg, Givental and Hofer in [8]. The arguments from [32] rely on elementary pseudoholomorphic curve methods. Invariance of the resulting homology is delicate and requires sophisticated technology, for instance, one can use the Polyfold Theory introduced by Hofer, Wysocki and Zehnder; see [15] for a survey. It will be shown in [30] that elementary methods are still enough to get invariance of CCH in its lowest degree. This is enough to get the first spectral invariant c_1^{CCH} well defined. Also in [30], it will be shown that c_1^{CCH} is the action of some periodic orbit with Conley–Zehnder index 3 realized as the asymptotic limit of a pseudoholomorphic plane. Hence we get the following consequence of a combination of Corollary 1.5 below with some of the results from [30]: *"The spectral invariant c_1^{CCH} of a nondegenerate dynamically convex contact form on S^3 is the contact area of some global surface of section."*

Theorem 1.4. *Let λ be a contact form on S^3 that is both nondegenerate and dynamically convex up to action C. Suppose that a periodic Reeb orbit $P = (x, T)$ satisfies $T \leq C$ and is the asymptotic limit of a fast finite-energy plane. Then, the knot $x(\mathbb{R})$ spans a global surface of section for the Reeb flow.*

In the theorem above and the corollary below a periodic Reeb orbit is a pair $P = (x, T)$, where x is a periodic trajectory of the Reeb flow and $T > 0$ is a period, not necessarily the primitive one.

Corollary 1.5. *Let λ be a contact form on S^3 that is both nondegenerate and dynamically convex up to action C. Suppose that a periodic Reeb orbit $P = (x, T)$ satisfies $T \leq C$, $\mathrm{CZ}(P) = 3$, and is the asymptotic limit of a finite-energy plane. Then the knot $x(\mathbb{R})$ spans a global surface of section for the Reeb flow.*

Proof. The equality $\mathrm{CZ}(P) = 3$ implies that any finite-energy plane asymptotic to P is fast. Now apply Theorem 1.4. □

The proof of Theorem 1.4 is based on a certain class of pseudoholomorphic planes called *fast*, but also uses ergodic methods (asymptotic cycles) [11,12,26,39,40]. Fast planes were originally introduced in [25] and later used in [23,24,27–29] to prove several existence results on global surfaces of section. Roughly speaking, an end of a plane is in some sense a gradient trajectory of the action functional, and the results from [17] basically say that the approach to the periodic orbit is governed by an eigenvector of an operator that plays the role of the Hessian of the action, the so-called *asymptotic operator*. The term "fast" refers to the fact that the eigenvalue of this asymptotic eigenvector has the same winding number of the most negative eigenvalue allowed, hence the approach is roughly the fastest it can be; see Definition 2.8.

Remark 1.6. The global sections obtained from Theorem 1.4 and Corollary 1.5 may have genus. Note that P is not assumed to be simply covered, but still the global sections obtained are Seifert surfaces for the knot $x(\mathbb{R})$.

2. Preliminaries

Let λ be a contact form on a 3-manifold M. The contact structure is denoted by $\xi = \ker \lambda$.

2.1. Periodic orbits, asymptotic operators and Conley–Zehnder indices

The Reeb vector field X_λ of λ is implicitly defined by

$$i_{X_\lambda} \mathrm{d}\lambda = 0, \qquad i_{X_\lambda} \lambda = 1.$$

Its flow ϕ^t is called the Reeb flow. Let us fix a marked point on every periodic trajectory of ϕ^t. A *periodic Reeb orbit* is a pair $P = (x, T)$, where $x : \mathbb{R} \to M$ is a periodic trajectory of ϕ^t such that $x(0)$ is the marked point, and $T > 0$ is a period. It is not required that T is the primitive period. The set of periodic orbits will be denoted by $\mathcal{P}(\lambda)$. If $T_0 > 0$ is the primitive period of x then

$k = T/T_0 \in \mathbb{N}$ is called the *covering multiplicity* of P. The contact form λ is said to be *nondegenerate up to action* $C \in (0,+\infty]$ if 1 is not in the spectrum of $\mathrm{d}\phi^T|_{x(0)} : \xi|_{x(0)} \to \xi|_{x(0)}$, for all $P = (x,T) \in \mathcal{P}(\lambda)$ such that $T \leq C$. When $C = +\infty$, we simply say that λ is *nondegenerate*.

There is an unbounded operator on $L^2(x(T\cdot)^*\xi)$

$$\eta \mapsto J(-\nabla_t \eta + T\nabla_\eta X_\lambda)$$

associated with a pair (P, J), where $P = (x,T) \in \mathcal{P}$ and $J : \xi \to \xi$ is a $\mathrm{d}\lambda$-compatible complex structure. Here ∇ is a symmetric connection on TM and ∇_t denotes the associated covariant derivative along the loop $t \in \mathbb{R}/\mathbb{Z} \mapsto x(Tt)$. This is called the *asymptotic operator*. It does not depend on the choice of ∇. It is self-adjoint when $L^2(x(T\cdot)^*\xi)$ is equipped with the inner product

$$(\eta, \zeta) \mapsto \int_{\mathbb{R}/\mathbb{Z}} \mathrm{d}\lambda(x(Tt))(\eta(t), J(x(Tt))\zeta(t))\ \mathrm{d}t.$$

Its spectrum is discrete, consists of eigenvalues whose geometric and algebraic multiplicities coincide, and accumulates at $\pm\infty$. It turns out that λ is nondegenerate if, and only if, 0 is never an eigenvalue of an asymptotic operator. The eigenvectors are nowhere vanishing sections of $x(T\cdot)^*\xi$ since they solve linear ODEs. Hence they have well-defined winding numbers with respect to a $\mathrm{d}\lambda$-symplectic trivialization σ of $x(T\cdot)^*\xi$. The winding number is independent of the choice of eigenvector of a given eigenvalue. This allows us to talk about the winding number

$$\mathrm{wind}_\sigma(\nu) \in \mathbb{Z}$$

of an eigenvalue ν with respect to σ. For every $k \in \mathbb{Z}$, there are precisely two eigenvalues satisfying $\mathrm{wind}_\sigma = k$, multiplicities counted and, moreover, $\nu_1 \leq \nu_2 \Rightarrow \mathrm{wind}_\sigma(\nu_1) \leq \mathrm{wind}_\sigma(\nu_2)$. These properties are independent of σ. These properties of the asymptotic operator have been established in [18]. Given any $\delta \in \mathbb{R}$ we set

$$\begin{aligned}
\alpha_\sigma^{<\delta}(P) &= \max\,\{\mathrm{wind}_\sigma(\nu) \mid \nu \text{ eigenvalue},\ \nu < \delta\},\\
\alpha_\sigma^{\geq\delta}(P) &= \min\,\{\mathrm{wind}_\sigma(\nu) \mid \nu \text{ eigenvalue},\ \nu \geq \delta\},\\
p^\delta(P) &= \alpha_\sigma^{\geq\delta}(P) - \alpha_\sigma^{<\delta}(P).
\end{aligned}$$

Finally, we consider the constrained Conley–Zehnder index

$$\mathrm{CZ}_\sigma^\delta(P) = 2\alpha_\sigma^{<\delta}(P) + p^\delta(P). \tag{1}$$

Note that this is defined also in degenerate situations.

A contact form λ is *dynamically convex up to action* $C \in (0,+\infty]$ if $c_1(\xi, \mathrm{d}\lambda)$ vanishes on $\pi_2 \hookrightarrow H_2$, and every contractible[1] $P = (x,T) \in \mathcal{P}(\lambda)$ satisfying $T \leq C$ also satisfies $\mathrm{CZ}^0_{\sigma_{\mathrm{disk}}}(P) \geq 3$. Here, σ_{disk} is a trivialization that extends to a capping disk. If $C = +\infty$, then we say that λ is *dynamically convex*.

[1]This means that the loop $t \in \mathbb{R}/\mathbb{Z} \mapsto x(Tt)$ is contractible.

Remark 2.1. Dynamical convexity was introduced by HWZ in [21]. The assumption that $c_1(\xi, \mathrm{d}\lambda)$ vanishes on spheres implies that the homotopy class of σ_{disk} does not depend on the choice of a capping disk.

2.2. Pseudoholomorphic curves in symplectizations

From now on, we assume M is closed. Let J be a compatible complex structure on the symplectic vector bundle $(\xi, \mathrm{d}\lambda)$. Hofer [13] considers an almost complex structure $\widetilde{J}$ defined on $\mathbb{R} \times M$ by

$$\widetilde{J} : \partial_a \mapsto X_\lambda \qquad\qquad \widetilde{J}|_\xi = J, \tag{2}$$

where X_λ and ξ are seen as $\mathbb{R}$-invariant objects in $\mathbb{R} \times M$. Then $\widetilde{J}$ is $\mathbb{R}$-invariant. Consider a closed Riemann surface (S, j), a finite set $\Gamma \subset S$ and a pseudoholomorphic map

$$\widetilde{u} = (a, u) : (S \setminus \Gamma, j) \to (\mathbb{R} \times M, \widetilde{J}),$$

satisfying a finite-energy condition

$$0 < E(\widetilde{u}) = \sup_{\phi} \int_{S\setminus\Gamma} \widetilde{u}^* \mathrm{d}(\phi\lambda) < \infty,$$

where the supremum is taken over the set of $\phi : \mathbb{R} \to [0, 1]$ satisfying $\phi' \geq 0$. The number $E(\widetilde{u})$ is called the *Hofer energy*. Such a map is called a *finite-energy* map. Points in Γ are called *punctures*. A puncture $z \in \Gamma$ is *positive* or *negative* if $a(w) \to +\infty$ or $a(w) \to -\infty$ when $w \to z$, respectively. It is called *removable* if $\limsup |a(w)| < \infty$ when $w \to z$. It turns out that every puncture is positive, negative or removable, and that $\widetilde{u}$ can be smoothly extended across a removable puncture; see [13].

Let $z \in \Gamma$ and let K be a conformal disk centered at z, i.e., there is a biholomorphic map $\varphi : (K, j, z) \to (\mathbb{D}, i, 0)$. Then $K \setminus \{z\}$ admits positive holomorphic polar coordinates $(s, t) \in [0, +\infty) \times \mathbb{R}/\mathbb{Z}$ defined by $(s, t) \simeq \varphi^{-1}(e^{-2\pi(s+it)})$, and negative holomorphic polar coordinates $(s, t) \in (-\infty, 0] \times \mathbb{R}/\mathbb{Z}$ defined by $(s, t) \simeq \varphi^{-1}(e^{2\pi(s+it)})$.

Theorem 2.2. (Hofer [13]) *Let $z \in \Gamma$ be a nonremovable puncture, and (s, t) be positive holomorphic polar coordinates at z. For every sequence $s_n \to +\infty$ there exist a subsequence s_{n_j} and $P = (x, T) \in \mathcal{P}$ such that $u(s_{n_j}, t) \to x(\epsilon T t + d)$ in $C^\infty(\mathbb{R}/\mathbb{Z}, M)$, for some $d \in \mathbb{R}$, where $\epsilon = \pm 1$ is the sign of the puncture.*

From now on, we denote by

$$\pi_\lambda : TM \to \xi \tag{3}$$

the projection along X_λ.

Theorem 2.3. (HWZ [17]) *Suppose that λ is nondegenerate up to action C, and that z is a nonremovable puncture of a finite-energy curve $\widetilde{u} = (a, u)$ in $(\mathbb{R} \times M, \widetilde{J})$ with Hofer energy $E(\widetilde{u}) \leq C$. Let (s, t) be positive holomorphic polar coordinates at z. There exist $P = (x, T) \in \mathcal{P}$, $d \in \mathbb{R}$ such that $u(s, t) \to x(\epsilon T t + d)$ in $C^\infty(\mathbb{R}/\mathbb{Z}, M)$ as $s \to +\infty$, where $\epsilon = \pm 1$ is the sign of the puncture.*

Remark 2.4. The orbit P is called the *asymptotic limit* of $\widetilde{u}$ at z.

Consider the space $\mathbb{R}/\mathbb{Z}\times\mathbb{C}$ equipped with coordinates $(\vartheta, z = x_1+ix_2)$ and contact form $\beta_0 = \mathrm{d}\vartheta + x_1\mathrm{d}x_2$.

Definition 2.5. A Martinet tube for $P = (x,T) \in \mathcal{P}$ is a smooth diffeomorphism $\Psi : \mathcal{N} \to \mathbb{R}/\mathbb{Z}\times\mathbb{D}$ defined on a smooth compact neighborhood $\mathcal{N}$ of $x(\mathbb{R})$ such that:

- $\Psi(x(T\vartheta/k)) = (\vartheta, 0)$ for all $\vartheta \in \mathbb{R}/\mathbb{Z}$, where $k\in\mathbb{N}$ is the covering multiplicity of P.
- $\lambda|_{\mathcal{N}} = \Psi^*(g\beta_0)$, where $g : \mathbb{R}/\mathbb{Z}\times\mathbb{D}\to(0,+\infty)$ is smooth and satisfies $g(\vartheta,0) = T/k$, $\mathrm{d}g(\vartheta,0) = 0$ for all $\vartheta\in\mathbb{R}/\mathbb{Z}$.

Theorem 2.6. (HWZ [17], Mora-Donato [36], Siefring [38]) *Suppose that λ is nondegenerate up to action $C>0$, and that z is a nonremovable puncture of sign $\epsilon=\pm1$ of a finite-energy curve $\widetilde{u}=(a,u)$ with Hofer energy $E(\widetilde{u})\le C$. Let (s,t) be positive or negative holomorphic polar coordinates at z when $\epsilon = +1$ or $\epsilon=-1$, respectively. Consider any Martinet tube $\Psi:\mathcal{N}\to\mathbb{R}/\mathbb{Z}\times\mathbb{D}$ for the asymptotic limit P of $\widetilde{u}$ at z, and $s_0\gg1$ such that $|s|\ge s_0 \Rightarrow u(s,t)\in\mathcal{N}$. Write $\Psi(u(s,t)) = (\vartheta(s,t), z(s,t))$ for $|s|\ge s_0$. Up to a rotation, we can assume $u(s,0)\to x(0)$ as $\epsilon s\to+\infty$.*

If $z(s,t)$ does not vanish identically, then the following holds. There exist $r>0$ and an eigenvalue ν of the asymptotic operator of (P,J) satisfying $\epsilon\nu<0$, such that:

- *There exist $c,d\in\mathbb{R}$ and a lift $\tilde{\vartheta}:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ of $\vartheta(s,t)$ such that*
$$\lim_{\epsilon s\to+\infty}\sup_{t\in\mathbb{R}/\mathbb{Z}} e^{r\epsilon s}\Big(|D^\beta[a(s,t)-Ts-c]| + |D^\beta[\tilde{\vartheta}(s,t)-kt-d]|\Big) = 0$$
holds for every partial derivative $D^\beta = \partial_s^{\beta_1}\partial_t^{\beta_2}$, where k is the covering multiplicity of P.
- *There exists an eigenvector of ν, represented as a nowhere-vanishing vector field $v(t)$ in the frame $\{\partial_{x_1},\partial_{x_2}\}$ along P, such that*
$$z(s,t) = e^{\nu s}(v(t)+R(s,t))$$
for some $R(s,t)$ satisfying $|D^\beta R(s,t)|\to0$ in $C^0(\mathbb{R}/\mathbb{Z})$ as $\epsilon s\to+\infty$, for every partial derivative $D^\beta = \partial_s^{\beta_1}\partial_t^{\beta_2}$.

The alternative $z(s,t)\equiv0$ can be expressed independent of coordinates as saying that the end of the domain of $\widetilde{u}$ corresponding to the puncture is mapped into the trivial cylinder over the asymptotic limit. In this case, we say that $\widetilde{u}$ has *trivial* asymptotic behavior at the puncture. Otherwise, the asymptotic behavior is said to be *nontrivial* at the puncture.

Remark 2.7. The eigenvalue ν provided by Theorem 2.6 is called the asymptotic eigenvalue of $\widetilde{u}$ at the puncture z.

Let us recall some of the invariants introduced in [18] in the $\mathbb{R}$-invariant case. Let $\widetilde{u}=(a,u)$ be a finite-energy curve on $(\mathbb{R}\times M,\widetilde{J})$, defined on a connected domain. Assume that λ is nondegenerate up to action $E(\widetilde{u})$. It can be shown that if $\pi_\lambda \circ \mathrm{d}u$ does not vanish identically then its zeros are isolated and count positively. Theorem 2.6 further implies that there are finitely many zeros in this case. HWZ [18] define

$$\mathrm{wind}_\pi(\widetilde{u}) \geq 0 \tag{4}$$

to be the algebraic count of zeros in case $\pi_\lambda \circ \mathrm{d}u$ does not vanish identically. Fix a $\mathrm{d}\lambda$-symplectic trivialization σ of $u^*\xi$. Let z be a puncture of $\widetilde{u}$ with asymptotic limit $P = (x, T)$. The asymptotic behavior described in Theorem 2.6 allows one to deform σ so that it extends to a trivialization of $x(T\cdot)^*\xi$. Let $\mathrm{wind}_\infty(\widetilde{u}, z, \sigma) \in \mathbb{Z}$ be defined to be the winding of the asymptotic eigenvalue of $\widetilde{u}$ at z with respect to the extension of σ to $x(T\cdot)^*\xi$. Finally, we consider

$$\mathrm{wind}_\infty(\widetilde{u}) = \sum\nolimits_+ \mathrm{wind}_\infty(\widetilde{u}, z, \sigma) - \sum\nolimits_- \mathrm{wind}_\infty(\widetilde{u}, z, \sigma),$$

where Σ_+ denotes a sum over the positive punctures, and Σ_- is a sum over the negative punctures. Standard degree theory shows that

$$\mathrm{wind}_\pi(\widetilde{u}) = \mathrm{wind}_\infty(\widetilde{u}) - \chi + \#\{\text{punctures}\} \tag{5}$$

holds provided $\int u^*\mathrm{d}\lambda > 0$. Note that $\mathrm{wind}_\infty(\widetilde{u})$ does not depend on the choice of trivialization σ of $u^*\xi$.

Denote by $(\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}, i)$ the Riemann sphere. For the next two definitions consider a finite-energy plane $\widetilde{u} = (a, u) : (\mathbb{C}, i) \to (\mathbb{R} \times M, \widetilde{J})$ and assume that λ is nondegenerate up to action $E(\widetilde{u})$. By Stokes' theorem, ∞ must be a positive puncture, and the similarity principle implies that $\int_{\mathbb{C}} u^*\mathrm{d}\lambda > 0$.

Definition 2.8. The plane $\widetilde{u}$ is said to be *fast* if $\mathrm{wind}_\infty(\widetilde{u}) = 1$.

Definition 2.9. The covering multiplicity $\mathrm{cov}(\widetilde{u})$ of the plane $\widetilde{u}$ is the covering multiplicity of its asymptotic limit.

Fast planes in symplectizations were originally introduced in [23].

Lemma 2.10. *If $\widetilde{u} = (a, u)$ is a fast plane, then $\widetilde{u}$ is somewhere injective and the map $u : \mathbb{C} \to M$ is an immersion transverse to X_λ.*

Proof. That u is an immersion transverse to X_λ follows from (4) and (5). If $\widetilde{u}$ is not somewhere injective, then it covers another plane via a polynomial map of degree ≥ 2, but this forces $\widetilde{u}$ to have critical points, in contradiction to u being an immersion; here, we used that the Cauchy–Riemann equations force a critical point to be a zero of the derivative of $\widetilde{u}$. $\square$

2.3. Asymptotic cycles

Here, we explain the basics on asymptotic cycles, and state the main result from [26]. Let ϕ^t be a smooth flow on a smooth, closed, oriented and connected 3-manifold M, and let L be a link consisting of (nonconstant) periodic orbits. The set of ϕ^t-invariant Borel probability measures on $M \setminus L$ is denoted by $\mathscr{P}_\phi(M \setminus L)$. Fix an auxiliary Riemannian metric g on M. If $p \in M \setminus L$ is recurrent and the sequence $T_n \to +\infty$ satisfies $\phi^{T_n}(p) \to p$, then we denote by $k(T_n, p)$ loops obtained by concatenating to $\phi^{[0,T_n]}(p)$ a g-shortest path from $\phi^{T_n}(p)$ to p. With $\mu \in \mathscr{P}_\phi(M \setminus L)$ and $y \in H^1(M \setminus L; \mathbb{R})$ fixed, one can use the Ergodic Theorem to show that μ-almost all points $p \in M \setminus L$ have the following properties: p is recurrent, and the limits

$$\lim_{n \to +\infty} \frac{\langle y, k(T_n, p)\rangle}{T_n}$$

exist independent of T_n and g, and define a μ-integrable function $f_{\mu,y}$. The integral

$$\mu \cdot y := \int_{M \setminus L} f_{\mu,y} \, \mathrm{d}\mu$$

is, by definition, the *intersection number* of μ and y.

If γ is the periodic orbit given by a connected component of L, then $\xi_\gamma = TM|_\gamma / T\gamma$ is a rank-2 vector bundle over γ. It carries an orientation induced by the ambient orientation and the flow orientation on γ. A positive frame of ξ_γ allows one to identify $\xi_\gamma \simeq \gamma \times \mathbb{C} \simeq \mathbb{R}/T_\gamma\mathbb{Z} \times \mathbb{C}$, where $T_\gamma > 0$ is the primitive period. If t is the coordinate on $\mathbb{R}/T_\gamma\mathbb{Z}$ (given by the flow) and $\theta \in \mathbb{R}/2\pi\mathbb{Z}$ is the polar angle on $\mathbb{C}^*$ then $\{\mathrm{d}t, \mathrm{d}\theta\}$ is a basis of $H^1((\xi_\gamma \setminus 0)/\mathbb{R}_+; \mathbb{R})$. With the aid of any exponential map the class y induces a class in this homology group that can be written as $p\mathrm{d}t + q\mathrm{d}\theta$. The coefficients $p, q \in \mathbb{R}$ depend only on y and on the chosen frame. If u is a nonzero vector in ξ_γ, then using the frame we can write $\mathrm{d}\phi^t \cdot u \simeq r(t)e^{i\theta(t)}$ with smooth functions $r(t) > 0, \theta(t)$. The rotation number

$$\rho^y(\gamma) := \frac{T_\gamma}{2\pi} \left(p + q \lim_{t \to +\infty} \frac{\theta(t)}{t} \right) \tag{6}$$

is independent of the choice of frame, and of the vector u.

The following statement is a refinement of a result due to Fried [11], see also Sullivan [40].

Theorem 2.11. ([26]) *Let $b \in H_2(M, L; \mathbb{Z})$ be induced by an oriented Seifert surface with boundary L, and denote by $b^* \in H^1(M \setminus L; \mathbb{R})$ the class dual to b. Consider the following assertions:*

(i) *L bounds a global surface of section representing b.*
(ii) *L binds an open book decomposition whose pages are global surfaces of section representing b.*
(iii) *The following hold:*
   (a) *$\rho^{b^*}(\gamma) > 0$ for every connected component $\gamma \subset L$.*
   (b) *$\mu \cdot b^* > 0$ for every $\mu \in \mathscr{P}_\phi(M \setminus L)$.*

Then (iii) $\Rightarrow$ (ii) $\Rightarrow$ (i) holds. Moreover, (i) $\Rightarrow$ (iii) holds C^∞-generically.

3. Proof of Theorem 1.1

The main input in the proof is the following statement proved with pseudo-holomorphic curves.

Theorem 3.1. ([24]) *Let λ be any dynamically convex contact form on (S^3, ξ_0). Then a periodic Reeb orbit bounds a disk-like global surface of section if, and only if, it is unknotted and has self-linking number -1.*

Here, there are no hidden genericity assumptions, the only assumption is that of dynamical convexity. A disk-like global surface of section D spanned by some unknotted, self-linking number -1 periodic orbit $\gamma = \partial D$ obtained from the above result has the following property: the first return time

$$\tau : D \setminus \gamma \to (0, +\infty) \qquad\qquad \tau(p) \;=\; \inf\,\{t > 0 \mid \phi^t(p) \in D\}$$

is bounded, i.e.,

$$\sup_{p\in D\setminus\gamma} \tau(p) \;<+\infty. \tag{7}$$

Since D is a global surface of section, it follows from (7) that there exists $L > 0$ such that $\phi^{[0,L]}(q) \cap D \neq \emptyset$ for every $q \in S^3 \setminus \gamma$.

Let $\gamma_1, \ldots, \gamma_N$ be a collection of unknotted, self-linking number -1 periodic Reeb orbits. These orbits are taken as knots, i.e., primitive orbits, oriented by the flow. Consider a disk-like global surface of section D_i spanned by γ_i, provided by Theorem 3.1, oriented in such a way that the identity $\partial D_i = \gamma_i$ takes orientations into account. Algebraically counting intersections with D_i induces a cohomology class $y_i \in H^1(S^3 \setminus \gamma_i; \mathbb{R})$. Denoting inclusion maps by $\iota_j : S^3 \setminus \cup_i\gamma_i \to S^3 \setminus \gamma_j$ we get a cohomology class

$$y = \sum_i \iota_i^* y_i \;\in\; H^1(S^3 \setminus \cup_i\gamma_i; \mathbb{R}). \tag{8}$$

Denote also

$$\ell_{ij} = \operatorname{link}(\gamma_i, \gamma_j) \geq 1, \tag{9}$$

which are positive integers since all D_i are global surfaces of section.

Let T_i denote the primitive period of γ_i. With i fixed consider a small smooth compact neighborhood $\mathcal{N}_i$ and a smooth, orientation preserving, diffeomorphism $\Psi_i : \mathcal{N}_i \to \mathbb{R}/T_i\mathbb{Z} \times \mathbb{D}$ such that $\Psi_i(\gamma_i) = \mathbb{R}/T_i\mathbb{Z} \times \{0\}$ and $\Psi_i \circ \phi^t \circ \Psi_i^{-1}(0, 0) = (t, 0)$. Here $\mathbb{D} \subset \mathbb{C}$ denotes the unit disk oriented by the complex orientation. Up to twisting, we may assume that Ψ_i is aligned with D_i, i.e., if $\epsilon > 0$ is small then the linking number of the loop $t \mapsto \Psi_i^{-1}(t, \epsilon)$ with γ_i is equal to zero. Denote by $re^{i\theta}$ the polar coordinates on D_i. It follows that with respect to the basis $\{\mathrm{d}t/T_i, \mathrm{d}\theta/2\pi\}$ of $\mathbb{R}/T_i\mathbb{Z} \times (\mathbb{D} \setminus \{0\})$ we can write

$$(\Psi_i)_* y = \left(\sum_{j\neq i} \ell_{ij}\right) \frac{\mathrm{d}t}{T_i} + \frac{\mathrm{d}\theta}{2\pi}\,.$$

It follows from this and from the definition of the rotation number (6) that

$$\begin{aligned} 2\pi\rho^y(\gamma_i) &= T_i \left(\sum_{j\neq i} \frac{\ell_{ij}}{T_i} + \frac{1}{2\pi} \lim_{t\to+\infty} \frac{\theta(t)}{t}\right) \\ &= \sum_{j\neq i} \ell_{ij} + \lim_{t\to+\infty} \frac{\theta(t)/2\pi}{t/T_i} \\ &\geq \lim_{t\to+\infty} \frac{\theta(t)/2\pi}{t/T_i}, \end{aligned} \tag{10}$$

where (9) was used in the third line. We claim that this limit is strictly positive. This will follow from $\mathrm{CZ}(\gamma_i) \geq 3$ together with $\mathrm{sl}(\gamma_i) = -1$. Here we write CZ for the Conley–Zehnder index in a global $\mathrm{d}\lambda$-symplectic frame of $(\xi_0, \mathrm{d}\lambda)$. In fact, the global $\mathrm{d}\lambda$-symplectic frame of ξ_0 rotates $\mathrm{sl}(\gamma_i) = -1$ turns with respect to a $\mathrm{d}\lambda$-symplectic of $\xi|_{\gamma_i}$ aligned with D_i. It turns out that there exists $\alpha_i \in \mathbb{R}$ such that $\mathrm{CZ}(\gamma_i^k) = 2\lfloor k\alpha_i \rfloor + p(\gamma_i^k)$ for every $k \geq 1$, where $|p(\gamma_i^k)| \leq 1$, and that if $\mathrm{CZ}(\gamma_i) \geq 3$ then $\alpha_i > 1$. Hence,

$$\lim_{t\to+\infty} \frac{\theta(t)/2\pi}{t/T_i} = \lim_{k\to+\infty} \frac{\mathrm{CZ}^{D_i}(\gamma_i^k)}{2k} = \lim_{k\to+\infty} \frac{\mathrm{CZ}(\gamma_i^k) - 2k}{2k} = \alpha_i - 1 > 0.$$

Hence, we are done checking

$$\rho^y(\gamma_i) > 0 \qquad \forall i \tag{11}$$

which is (iii-a) in Theorem 2.11.

Now, we check (iii-b). Let $\mu \in \mathscr{P}_\phi(S^3 \setminus \cup_i \gamma_i)$ be arbitrary. As explained in subsection 2.3, there exists a Borel set $E \subset M \setminus \cup_i \gamma_i$ contained in the set of recurrent points such that $\mu(E) = 1$, and for all $p \in E$ the limits $\lim_{n\to\infty} \langle y, k(T_n, p)\rangle / T_n$ exist independent of the sequence $T_n \to +\infty$ satisfying $\phi^{T_n}(p) \to p$ and define a function $f_{\mu,y} \in L^1(\mu)$ whose integral is $\mu \cdot y$. Since each $D_i \setminus \gamma_i$ is transverse to the flow, we conclude that $\mu(E \setminus \cup_i D_i) = 1$. Fix $p \in E \setminus \cup_i D_i$ and a sequence $T_n \to +\infty$ satisfying $\phi^{T_n}(p) \to p$. Then using the (positive) transversality of the flow with all the surfaces $D_i \setminus \gamma_i$

$$n \gg 1 \qquad \Rightarrow \qquad \langle y, k(T_n, p)\rangle = \sum_i \#\{t \in [0, T_n] \mid \phi^t(p) \in D_i\}. \tag{12}$$

But

$$\#\{t \in [0, T_n] \mid \phi^t(p) \in D_i\} \geq \frac{T_n}{\sup\ \tau_i} - 1, \tag{13}$$

where τ_i is the return time function of D_i. Recall that $\sup \tau_i < +\infty$ (7). Plugging (13) into (12) we obtain

$$\frac{\langle y, k(T_n, p)\rangle}{T_n} \geq \sum_i \left(\frac{1}{\sup\ \tau_i} - \frac{1}{T_n} \right). \tag{14}$$

Taking the limit as $n \to \infty$

$$f_{\mu,y} \geq \sum_i \frac{1}{\sup\ \tau_i}\ (\mu\text{-almost everywhere}) \quad \Rightarrow \quad \mu \cdot y \geq \sum_i \frac{1}{\sup\ \tau_i} > 0, \tag{15}$$

and we are done checking (iii-b). A direct application of Theorem 2.11 concludes the proof of Theorem 1.1.

4. Proof of Theorem 1.4

Let $P = (x, T = m_0 T_0)$ be a periodic Reeb orbit, with multiplicity m_0, of a defining contact form λ on (S^3, ξ_0). Here T_0 denotes the primitive period of x. Throughout this section we denote by τ a global $\mathrm{d}\lambda$-symplectic trivialization of ξ_0. Assume that λ is dynamically convex up to action T, and also that λ is nondegenerate up to action T.

Proposition 4.1. *If $\widetilde{u}$ is a fast plane asymptotic to P then there exists $a > 0$ such that*

$$\#\{t \in [0,T] \mid \phi^t(p) \in u(\mathbb{C})\} \geq \left\lfloor \frac{T}{a} \right\rfloor \tag{16}$$

holds for every $p \in S^3 \setminus x(\mathbb{R})$ and every $T \geq 0$.

We first show that Proposition 4.1 can be used to check the hypothesis of Theorem 2.11 for the periodic orbit $x(\mathbb{R})$ and the cohomology class counting linking numbers with it. Theorem 1.4 follows as a consequence.

Proof that Theorem 1.4 follows from Proposition 4.1. Let $y \in H^1(S^3 \setminus x(\mathbb{R}); \mathbb{R})$ be the cohomology class that counts linking number of loops in $S^3 \setminus x(\mathbb{R})$ with the loop $t \in \mathbb{R}/\mathbb{Z} \mapsto x(T_0 t) = x(Tt/m_0)$. Here, we ignore $\mathbb{Z}$-coefficients and work with $\mathbb{R}$-coefficients. If we compactify $\mathbb{C}$ to a disk $D_{\mathbb{C}}$ by adding a circle at ∞ then u induces a capping disk $\bar{u} : D_{\mathbb{C}} \to S^3$ for P such that the class in $H^1(S^3 \setminus x(\mathbb{R}))$ dual to $\bar{u}_*[D_{\mathbb{C}}] \in H_2(S^3, x(\mathbb{R}))$ is precisely $m_0 y$. Here, $[D_{\mathbb{C}}]$ is the fundamental class in $H_2(D_{\mathbb{C}}, \partial D_{\mathbb{C}}; \mathbb{Z})$ induced by the complex orientation. Observe that $u(\mathbb{C}) \setminus x(\mathbb{R})$ has measure zero with respect to any $\mu \in \mathscr{P}_\phi(S^3 \setminus x(\mathbb{R}))$ since it is transverse to the flow. Hence, in view of the discussion in Sect. 2.3, we get a Borel set $E \subset S^3 \setminus x(\mathbb{R})$ such that $\mu(E) = 1$ and every point $p \in E$ has the following properties:

(a) p is recurrent.

(b) The limits

$$\lim_{n \to +\infty} \frac{\mathrm{link}(k(T_n, p), x(\mathbb{R}))}{T_n} = \lim_{n \to +\infty} \frac{\langle y, k(T_n, p)\rangle}{T_n}$$

exist independent of $T_n \to +\infty$ satisfying $\phi^{T_n}(p) \to p$ (and of auxiliary Riemannian metrics), and define a function in $L^1(\mu)$ whose integral is equal to the intersection number $\mu \cdot y$.

(c) $p \notin u(\mathbb{C})$.

Hence, using the transversality between u and the Reeb vector field, for every $p \in E$ we can estimate

$$\begin{aligned} m_0 \ \mathrm{link}(k(T_n, p), x(\mathbb{R})) &= \langle m_0 y, k(T_n, p)\rangle \\ &= \sum_{t \in [0,T_n],\ \phi^t(p) \in u(\mathbb{C})} \#\{z \in \mathbb{C} \mid u(z) = \phi^t(p)\} \\ &\geq \#\{t \in [0, T_n] \mid \phi^t(p) \in u(\mathbb{C})\} \end{aligned} \tag{17}$$

for all n large enough, where $T_n \to +\infty$ satisfies $\phi^{T_n}(p) \to p$. With the aid of Proposition 4.1 we can estimate from (17)

$$\lim_{n \to +\infty} \frac{\mathrm{link}(k(T_n, p), x(\mathbb{R}))}{T_n} \geq \lim_{n \to +\infty} \frac{1}{m_0 T_n} \left\lfloor \frac{T_n}{a} \right\rfloor = \frac{1}{m_0 a} \qquad \forall p \in E, \tag{18}$$

which implies, by definition of intersection numbers, that

$$\mu \cdot y \geq \frac{1}{m_0 a} > 0 \qquad \forall \mu \in \mathscr{P}_\phi(S^3 \setminus x(\mathbb{R})) \tag{19}$$

Condition $\rho^y(x(\mathbb{R})) > 0$ follows immediately from $\mathrm{CZ}(P_0) \geq 3$ where $P_0 = (x, T_0)$ is the simply covered periodic orbit underlying P. Theorem 1.4 now follows from a direct application of Theorem 2.11 since y is dual to any Seifert surface for $x(\mathbb{R})$; here, it was used that the ambient space is S^3. $\square$

To complete the proof of Theorem 1.4 we need to establish Proposition 4.1. The rest of this section is concerned with the proof of Proposition 4.1.

Let us denote by $P_0 = (x, T_0)$ the simply covered periodic orbit underlying P, and recall that m_0 denotes the covering multiplicity of $P = (x, T = m_0 T_0)$. For every $k \geq 1$, we denote $P_0^k = (x, kT_0)$. In particular $P = P_0^{m_0}$.

Consider the set $\mathcal{M}^{\text{fast}}(P, J)$ of equivalence classes of fast finite-energy planes asymptotic to P, where two planes $\widetilde{u}, \widetilde{v}$ are equivalent if there exist $A \in \mathbb{C}^*$, $B \in \mathbb{C}$ and $c \in \mathbb{R}$ such that $\widetilde{v}_c(z) = \widetilde{u}(Az + B)$ holds for every $z \in \mathbb{C}$. Here $\widetilde{v}_c$ denotes the translation of $\widetilde{v}$ by c in the $\mathbb{R}$-component. Equivalence classes are denoted by $[\widetilde{u}]$.

It is possible to build a Fredholm theory for $\mathcal{M}^{\text{fast}}(P, \bar{J})$ modeled on sections of the normal bundle, using Sobolev or Hölder spaces [3]. Fix a number $\delta < 0$ in the spectral gap of the asymptotic operator associated to (P, J) between eigenvalues with winding number 1 and 2 with respect to τ. This is possible since $\mathrm{CZ}_\tau^0(P) \geq 3$. Note that $\alpha_\tau^{<\delta}(P) = 1$ and $\alpha_\tau^{\geq\delta}(P) = 2$. Let

$$\widetilde{u} = (a, u) : (\mathbb{C}, i) \to (\mathbb{R} \times S^3, \widetilde{J}) \tag{20}$$

be a fast plane representing an element of $\mathcal{M}^{\text{fast}}(P, J)$. Consider the space of sections of the normal bundle of $\widetilde{u}(\mathbb{C})$ with exponential decay faster than δ. The Fredholm index of the linearization $D_{\widetilde{u}}$ of the Cauchy–Riemann equations at $\widetilde{u}$ restricted to this space of sections is

$$\mathrm{ind}_\delta(\widetilde{u}) = \mathrm{CZ}_\tau^\delta(P) - 1 = 3 - 1 = 2. \tag{21}$$

An important fact is that *automatic transversality* holds, i.e., $D_{\widetilde{u}}$ at a fast plane $\widetilde{u}$ is always a surjective Fredholm operator. Let us prove this fact. There is no loss of generality to deform the normal bundle so that it coincides with $u^*\xi_0$ over $\mathbb{C} \setminus B_R(0)$, $R \gg 1$. A $\mathrm{d}\lambda$-symplectic trivializing frame of the normal bundle induces, up to homotopy, a $\mathrm{d}\lambda$-symplectic trivialization σ_N of $x(T\cdot)^*\xi_0$ which winds $+1$ with respect to the global frame τ. Moreover, a nontrivial section $\zeta \in \ker D_{\widetilde{u}}$ admits an asymptotic behavior governed by an eigensection of the asymptotic operator associated with an eigenvalue $\nu < \delta$, see [23, Theorem 6.1], or [38, Theorem A.1]. Hence, ζ does not vanish near ∞ and the total algebraic count of zeros of ζ is equal to the winding number of ν with respect to σ_N, which is equal to

$$\mathrm{wind}_\tau(\nu) - 1 \leq \alpha_\tau^{<\delta}(P) - 1 = 1 - 1 = 0.$$

But the equation $D_{\widetilde{u}}\zeta = 0$ allows us to use Carleman's similarity principle to say that zeros are isolated and count positively. The important conclusion is that ζ never vanishes. Since the Fredholm index is 2, we would find 3 linearly independent sections of the kernel if $D_{\widetilde{u}}$ were not surjective. But the normal bundle is two-dimensional, and hence a nontrivial linear combination of them would have to vanish at some point, which gives the desired contradiction.

Remark 4.2. Arguments like the one used above to prove automatic transversality statements were explored in [16, 41], see also [25].

It follows from the above discussed automatic transversality that $\mathcal{M}^{\text{fast}}(P, J)$ can be given the structure of a smooth, Hausdorff and second-countable one-dimensional manifold.

Remark 4.3. Under our standing assumption that λ is nondegenerate up to action T one can show that the topology on $\mathcal{M}^{\text{fast}}(P, J)$ inherited from the functional analytic set-up used for the Fredholm theory coincides with the topology of C^∞_{loc}-convergence. There are situations where this can also be proved dropping nondegeneracy [24, 29].

Consider a sequence $\widetilde{u}_n : (\mathbb{C}, i) \to (\mathbb{R} \times S^3, \widetilde{J})$ of fast finite-energy planes asymptotic to P. Since λ is assumed to be nondegenerate up to action T we can apply the SFT compactness theorem [3] to get, up to selection of a subsequence, that $\widetilde{u}_n$ SFT-converges to a stable holomorphic building $\mathbf{u}$. Since $\mathbf{u}$ is a limit of planes it can be conveniently described as a directed, rooted tree $\mathcal{T}$. Each vertex v corresponds to a finite-energy map

$$\widetilde{u}_v = (a_v, u_v) : (\mathbb{C} \setminus \Gamma_v, i) \to (\mathbb{R} \times S^3, \widetilde{J})$$

with a unique positive puncture ∞. The finite set Γ_v consists of the negative punctures of $\widetilde{u}_v$. The top level of this building corresponds to the root r, and consists of a single finite-energy map $\widetilde{u}_r$ which is asymptotic to P at its positive puncture ∞. Edges are always assumed oriented as going away from the root. An edge e from the vertex v to the vertex v' corresponds to a negative puncture of $\widetilde{u}_v$. The asymptotic limit $\widetilde{u}_v$ at the negative puncture corresponding to e is equal to the asymptotic limit of $\widetilde{u}_{v'}$ at its positive puncture. The leaves correspond precisely to the vertices v such that $\widetilde{u}_v$ is a plane ($\Gamma_v = \emptyset$).

Lemma 4.4. *If v is a vertex of $\mathcal{T}$ such that $\int u_v^* \mathrm{d}\lambda > 0$ then* wind_∞ $(\widetilde{u}_v, \infty, \tau) \leq 1$.

Proof. SFT compactness allows us to find $A_n \in \mathbb{C}^*$, $B_n \in \mathbb{C}$ and $c_n \in \mathbb{R}$ such that the planes $\widetilde{w}_n(z) = c_n \cdot \widetilde{u}_n(A_n z + B_n)$ converge to $\widetilde{u}_v$ in $C^\infty_{\text{loc}}(\mathbb{C} \setminus \Gamma_v)$. Here, $c_n \cdot \widetilde{u}_n$ denotes the translation by c_n in the $\mathbb{R}$-component.

Consider components $\widetilde{w}_n = (d_n, w_n)$ and $\widetilde{u}_v = (a_v, u_v)$ in $\mathbb{R} \times S^3$. Write $\widetilde{w}_n(s,t) = (d_n(s,t), w_n(s,t))$ instead of $\widetilde{w}_n(e^{2\pi(s+it)})$, and similarly $\widetilde{u}_v(s,t) = (a_v(s,t), u_v(s,t))$. Fix s_0 such that $z \in \Gamma_v \Rightarrow |z| < e^{2\pi s_0}$. By Theorem 2.6, we can find $s_1 > s_0$ such that $\pi_\lambda(\partial_s u_v)$ does not vanish on $[s_1, +\infty) \times \mathbb{R}/\mathbb{Z}$ and the winding number $\text{wind}(\pi_\lambda(\partial_s u_v)(s_1, \cdot))$ of $t \mapsto \pi_\lambda(\partial_s u_v)(s_1, t)$ in the global frame τ is equal to $\text{wind}_\infty(\widetilde{u}_v, \infty, \tau)$. Since $\pi_\lambda(\partial_s w_n) \to \pi_\lambda(\partial_s u_v)$ in C^∞_{loc} we find n_0 such that if $n \geq n_0$ then $\pi_\lambda(\partial_s w_n)$ does not vanish on $\{s_1\} \times \mathbb{R}/\mathbb{Z}$ and

$$\text{wind}(\pi_\lambda(\partial_s w_n)(s_1, \cdot)) = \text{wind}(\pi_\lambda(\partial_s u_v)(s_1, \cdot)) = \text{wind}_\infty(\widetilde{u}_v, \infty, \tau).$$

The frame τ can be used to represent the maps $(s,t) \mapsto \pi_\lambda(\partial_s w_n)$ by smooth maps $\zeta_n : [s_0, +\infty) \times \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ satisfying a Cauchy–Riemann type equation. Carleman's similarity principle implies that either ζ_n vanishes identically on

$[s_0,+\infty)\times\mathbb{R}/\mathbb{Z}$, or its zeros are isolated and count positively. It can not vanish identically since the $\widetilde{w}_n$ are planes. By Theorem 2.6 $\zeta_n(s,t)$ does not vanish when s is large enough and for every n we have

$$\lim_{s\to+\infty}\mathrm{wind}(\zeta_n(s,\cdot))=\lim_{s\to+\infty}\mathrm{wind}(\pi_\lambda(\partial_s w_n)(s,\cdot))=\mathrm{wind}_\infty(\widetilde{w}_n)=\mathrm{wind}_\infty(\widetilde{u}_n).$$

If $s>s_1$ is large enough then $\mathrm{wind}(\zeta_n(s,\cdot))-\mathrm{wind}(\zeta_n(s_1,\cdot))$ is the algebraic count of zeros of ζ_n on $[s_1,s]\times\mathbb{R}/\mathbb{Z}$. Since this count is nonnegative we get

$$\mathrm{wind}(\pi_\lambda(\partial_s w_n)(s_1,\cdot))\le\mathrm{wind}_\infty(\widetilde{u}_n)$$

for all $n\ge n_0$. Hence,

$$n\ge n_0\Rightarrow\mathrm{wind}_\infty(\widetilde{u}_v,\infty,\tau)\le\mathrm{wind}_\infty(\widetilde{u}_n)=1$$

as desired. □

Lemma 4.5. *If the vertex v is not a leaf, then $\int u_v^*\mathrm{d}\lambda=0$, i.e., $\widetilde{u}_v$ is a possibly branched cover of a trivial cylinder over a periodic orbit.*

Proof. Suppose that $\int u_v^*\mathrm{d}\lambda>0$. At the negative punctures $z\in\Gamma_v$ of $\widetilde{u}_v$, we have $\mathrm{wind}_\infty(\widetilde{u}_v,z,\tau)\ge 2$ since the asymptotic limits at these punctures are periodic Reeb orbits with action less than T and hence, by assumption, satisfy $\mathrm{CZ}_\tau^0\ge 3$. By the previous lemma together with (4) and (5) we arrive at

$$0\le\mathrm{wind}_\pi(\widetilde{u}_v)=\mathrm{wind}_\infty(\widetilde{u}_v)-1+\#\Gamma_v\le 1-2\#\Gamma_v-1+\#\Gamma_v=-\#\Gamma_v. \tag{22}$$

Thus, $\Gamma_v=\emptyset$ and v is a leaf. □

Corollary 4.6. *The following dichotomy holds for every vertex v of $\mathcal{T}$:*
- (i) *v is not a leaf, $\int u_v^*\mathrm{d}\lambda=0$ and $\widetilde{u}_v$ is a (possibly branched) cover of a trivial cylinder.*
- (ii) *v is a leaf, $\int u_v^*\mathrm{d}\lambda>0$ and $\widetilde{u}_v$ is a fast plane asymptotic to a covering of P_0.*

Proof. Case (i) is handled by the previous lemma. Let us now argue for (ii). By Lemma 4.4 if v is a leaf then it is a plane satisfying $\mathrm{wind}_\infty(\widetilde{u}_v)\le 1$. Hence, $0\le\mathrm{wind}_\pi(\widetilde{u}_v)=\mathrm{wind}_\infty(\widetilde{u}_v)-1\le 1-1=0$, i.e., $\mathrm{wind}_\infty(\widetilde{u}_v)=1$ and $\widetilde{u}_v$ is a fast plane, and by (i) it is clearly asymptotic to a cover of P_0. □

For every $1\le k\le m_0$, we consider $\mathcal{M}^{\mathrm{fast}}(P_0^k,J)$ the moduli space of fast finite-energy planes asymptotic to P_0^k, defined as before. For each k there is a suitable choice of negative weight placed precisely at the spectral gap between eigenvalues of the asymptotic operator associated to (P_0^k,J) with winding 1 and 2 in a global frame. With these weights one builds a Fredholm theory as before, and there is automatic transversality. The space $\mathcal{M}^{\mathrm{fast}}(P_0^k,J)$ becomes a 1-dimensional smooth, second countable Hausdorff manifold. Moreover, the induced topology coincides with the topology induced by C^∞_{loc}-convergence.

Corollary 4.7. *There exists $m\in\{1,\dots,m_0\}$ such that $\mathcal{M}^{\mathrm{fast}}(P_0^m,J)$ is nonempty and compact.*

Proof. We work under the assumption of Theorem 1.4 that $\mathcal{M}^{\text{fast}}(P=P_0^{m_0},J)$ is nonempty. If $\mathcal{M}^{\text{fast}}(P,J)$ is compact then there is nothing to be proved. If $\mathcal{M}^{\text{fast}}(P,J)$ is not compact then some sequence in $\mathcal{M}^{\text{fast}}(P,J)$, represented by fast planes $\widetilde{u}_n$, will SFT-converge to a building $\mathbf{u}$ with more than one level. This means that the corresponding tree $\mathcal{T}$ does not consist of a single vertex (the root), and by Corollary 4.6 every leaf v must be a fast finite-energy plane asymptotic to $P_0^{k_v}$, for some $k_v \in \{1,\dots,m_0-1\}$. The reason for the strict inequality $k_v < m_0$ is that the root must have at least two negative punctures: otherwise the root corresponds to a trivial cylinder, which is ruled out by stability of the limiting building. Pick any leaf v, denote $m_1 = k_v$. Hence the moduli space $\mathcal{M}^{\text{fast}}(P_0^{m_1},J)$ of fast planes asymptotic to $P_0^{m_1}$ is not empty. If $\mathcal{M}^{\text{fast}}(P_0^{m_1},J)$ is compact, then we are done with the proof. If not we proceed just as above to find $1 \le m_2 \le m_1 - 1$ such that $\mathcal{M}^{\text{fast}}(P_0^{m_2},J)$ is non empty. After a finite number of steps $k \ge 0$ this process stops and we find $1 \le m_k \le m_0$ such that $\mathcal{M}^{\text{fast}}(P_0^{m_k},J)$ is nonempty and compact. □

From now on m is given by the previous lemma, that is, $\mathcal{M}^{\text{fast}}(P_0^m,J)$ is a nonempty, compact, smooth and Hausdorff 1-dimensional manifold, i.e., a finite collection of circles.

Consider the space $\mathcal{M}_1^{\text{fast}}(P_0^m,J)$ of equivalence classes of pairs $(\widetilde{u},z)$ where $\widetilde{u}$ is a fast plane asymptotic to P_0^m and $z \in \mathbb{C}$. Two pairs $(\widetilde{u}_0,z_0)$, $(\widetilde{u}_1,z_1)$ are equivalent if there exist $A \in \mathbb{C}^*$, $B \in \mathbb{C}$ such that $\widetilde{u}_1(Az+B) = \widetilde{u}_0(z)$ for all $z \in \mathbb{C}$ and $z_1 = Az_0 + B$. Note that $(\mathbb{R},+)$ acts freely on $\mathcal{M}_1^{\text{fast}}(P_0^m,J)$ by translations in the symplectization direction. Hence, $\mathcal{M}_1^{\text{fast}}(P_0^m,J)/\mathbb{R}$ is a smooth three-dimensional manifold. The map

$$\text{ev} : \mathcal{M}_1^{\text{fast}}(P_0^m,J)/\mathbb{R} \to S^3 \qquad \text{ev}([\widetilde{u}=(a,u),z]/\mathbb{R}) \mapsto u(z) \tag{23}$$

is smooth.

Lemma 4.8. *The map* ev *is a submersion.*

Proof. For every $\widetilde{u} \in \mathcal{M}^{\text{fast}}(P_0^m,J)$ nontrivial sections in the kernel of the linearized Cauchy–Riemann operator at $\widetilde{u}$, with the appropriate weighted exponential decay, which represent elements in the tangent space, never vanish and u is an immersion. □

Lemma 4.9. *If $K \subset S^3 \setminus x(\mathbb{R})$ is compact then* $\text{ev}^{-1}(K)$ *is compact.*

Proof. Suppose that $[\widetilde{u}_n,z_n]$ represents a sequence in $\text{ev}^{-1}(K)$. Up to reparametrization, translation in the $\mathbb{R}$-component, and selection of a subsequence, we may assume that $\widetilde{u}_n$ converges in C^∞_{loc} to some plane $\widetilde{u}$ representing an element of $\mathcal{M}^{\text{fast}}(P_0^m,J)$. Let $\mathcal{N}$ be a neighborhood of $x(\mathbb{R})$ such that $K \cap \mathcal{N} = \emptyset$. One can then invoke results on cylinders of small contact area from [22] to conclude that there exists R and n_0 such that if $n \ge n_0$ and $|z| \ge R$ then $\widetilde{u}_n(z) \in \mathbb{R}\times\mathcal{N}$. This implies that $\sup_n |z_n| \le R$. Hence one can assume, up to selection of a subsequence, that $z_n \to z$ for some z. It follows that $[\widetilde{u}_n,z_n]/\mathbb{R} \to [\widetilde{u},z]/\mathbb{R}$. □

Lemma 4.10. *The image of the map* ev *contains $S^3 \setminus x(\mathbb{R})$.*

Proof. By Lemma 4.8 the image is open in S^3, hence its intersection with $S^3 \setminus x(\mathbb{R})$ is an open subset of $S^3 \setminus x(\mathbb{R})$. By Lemma 4.9, the intersection of the image of ev with $S^3 \setminus x(\mathbb{R})$ is a closed subset of $S^3 \setminus x(\mathbb{R})$. The conclusion follows from connectedness of $S^3 \setminus x(\mathbb{R})$. □

Consider $[\widetilde{u} = (a, u)] \in \mathcal{M}^{\text{fast}}(P_0^m, J)$ and the function

$$\tau : S^3 \setminus x(\mathbb{R}) \to [0, +\infty] \tag{24}$$

defined by

$$\tau(p) \;=\; \inf\{t > 0 \mid \phi^t(p) \in u(\mathbb{C})\} \tag{25}$$

with the convention that the infimum of the empty set is $+\infty$.

Lemma 4.11. *τ takes values on $(0, +\infty)$, and $\sup \tau < +\infty$.*

Proof. From the transversality of u to the Reeb flow, and the asymptotic formula from Theorem 2.6, we conclude that given any $[\widetilde{v} = (b, v)] \in \mathcal{M}^{\text{fast}}$ (P_0^m, J) and $p \in S^3$, the set $v^{-1}(p) \subset \mathbb{C}$ is finite, and also that τ takes values on $(0, +\infty]$.

Suppose that $p \notin x(\mathbb{R})$ and $\omega(p) \cap x(\mathbb{R}) \neq \emptyset$. By invariance of $x(\mathbb{R})$ under the Reeb flow, the trajectory $\phi^t(p)$ will spend arbitrarily long times in the future arbitrarily and uniformly close to $x(\mathbb{R})$. Hence, the way in which it rotates around $x(\mathbb{R})$ is governed by the linearized Reeb flow along x. Every plane $\widetilde{v} = (b, v)$ representing an element in $\mathcal{M}^{\text{fast}}(P_0^m, J)$ is asymptotic to P_0^m according to an eigenvector of a negative eigenvalue of the asymptotic operator with winding $+1$ in a global frame; this information is encoded in $\text{wind}_\infty(\widetilde{v}) = 1$. Hence, in transverse polar coordinates aligned with the global frame the plane rotates 2π. After one period $T = mT_0$ the linearized flow rotates every transverse vector by an angle larger than $2\pi + \Delta$ for some uniform $\Delta > 0$. This information is encoded in $\text{CZ}^0_\tau(P_0^m) \geq 3$. Hence after flow time of about $\lfloor \frac{2\pi}{\Delta} + 1 \rfloor T$ any point nearby P_0 already returned once back to the plane. It follows that the return time is bounded from above for points near P_0.

If $\omega(p) \cap x(\mathbb{R}) = \emptyset$, then it follows from compactness of $\mathcal{M}^{\text{fast}}(P_0^m, J)$ and transversality of the planes to the Reeb flow that for every $[\widetilde{v}] \in \mathcal{M}^{\text{fast}}(P_0^m, J)$ the trajectory $\phi^t(p)$ will hit $v(\mathbb{C})$ in finite time.

So far we have proved that τ takes values on $(0, +\infty)$, and that τ is bounded near $x(\mathbb{R})$. To conclude, we note that if

$$\{w_1, \ldots, w_N\} = u^{-1}(\phi^{\tau(p)}(p)),$$

then there are N local smooth germs of hitting times $\tau_1, \ldots, \tau_N$ near p. Then, τ can be locally bounded in terms of these germs. □

Proposition 4.1 is a consequence of Lemma 4.11.

Remark 4.12. We observe that the finite energy planes produced by Corollary 4.7 can themselves be thought of as sorts of generalized surfaces of section where we allow for the possibility that the surface is an immersion rather than embedding. Indeed our proof shows that the projection to S^3 of every such plane is an immersion, transverse to the Reeb flow, and that the flow line through any given point in $S^3 \setminus P_0$ will hit the surface in forward and backward time. In the case that the plane is not an embedding, it follows from results in [18,37] that it must intersect its asymptotic limit, and thus in this case the plane will intersect the flow line through any given point in S^3 including points in P_0.

We observe further that, since our proof shows that the evaluation map

$$\mathrm{ev} : \mathcal{M}_1^{\mathrm{fast}}(P_0^m, J)/\mathbb{R} \to S^3$$

is an immersion between manifolds of the same dimension, it is also a local diffeomorphism, so we can use ev^{-1} to lift the flow to the moduli space $\mathcal{M}_1^{\mathrm{fast}}(P_0^m, J)/\mathbb{R}$, each component of which is diffeomorphic to $\mathbb{C} \times S^1$. Moreover, since each plane in $\mathcal{M}_1^{\mathrm{fast}}(P_0^m, J)/\mathbb{R}$ is transverse to the flow, the resulting flow on $\mathcal{M}_1^{\mathrm{fast}}(P_0^m, J)/\mathbb{R}$ will be transverse to the disk-like fibers of the forgetful map $\mathcal{M}_1^{\mathrm{fast}}(P_0^m, J)/\mathbb{R} \mapsto \mathcal{M}^{\mathrm{fast}}(P_0^m, J)/\mathbb{R}$. So although the surface of section provided by our theorem will in general have genus, the fast finite energy planes that we construct in the proof can themselves be used to visualize the dynamics as a return map on a disk.

Acknowledgements

UH would like to thank V. Colin, P. Dehornoy and A. Rechtman for enlightening discussions leading to Remark 1.2 during the *104e rencontre entre mathématiciens et physiciens théoriciens* held at IRMA (2019), the organisers of this event for the invitation, and IRMA for the hospitality. UH also thanks Kai Cieliebak and Urs Frauenfelder for helpful conversations, and the hospitality during a visit to Universität Augsburg. RS thanks Jungsoo Kang and Urs Frauenfelder for helpful discussions. We thank the referee for interesting and helpful feedback. RS was partially supported by the SFB/TRR 191 "Symplectic Structures in Geometry, Algebra and Dynamics" funded by the DFG (Projektnummer 281071066 - TRR 191). PS is partially supported by FAPESP 2016/25053-8 and CNPq 306106/2016-7. PS acknowledges the support of NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai.

Funding Open Access funding enabled and organized by Projekt DEAL.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Abbas, C.: Holomorphic open book decompositions. Duke Math. J. **158**(1), 29–82 (2011). https://doi.org/10.1215/00127094-1276301

[2] Albers, P., Geiges, H., Zehmisch, K.: A symplectic dynamics proof of the degree-genus formula. arXiv:1905.03054

[3] Bourgeois, F., Eliashberg, Y., Hofer, H., Wysocki, K., Zehnder, E.: Compactness results in symplectic field theory. Geom. Topol. **7**, 799–888 (2003). https://doi.org/10.2140/gt.2003.7.799

[4] Colin, V., Dehornoy, P., Rechtman, A.: On the existence of supporting broken book decompositions for contact forms in dimension 3. arXiv:2001.01448v3

[5] Dehornoy, P.: Which geodesic flows are left-handed? Groups Geom. Dyn. **11**(4), 1347–1376 (2017). https://doi.org/10.4171/GGD/431

[6] Doicu, A., Fuchs, U.: Asymptotic behaviour for H-holomorphic cylinders of small area. J. Fixed Point Theory Appl. **21**(4), 51 (2019). https://doi.org/10.1007/s11784-019-0735-6

[7] Doicu, A., Fuchs, U.: Asymptotic behaviour for H-holomorphic cylinders of small area. arXiv:1802.05573v2

[8] Eliashberg, Y., Givental, A., Hofer, H.: Introduction to symplectic field theory. Geom. Funct. Anal. Special Volume, Part II (2000), pp. 560–673. GAFA 2000 (Tel Aviv, 1999)

[9] Florio, A., Hryniewicz, U.: Quantitative conditions for right-handedness. arXiv:2106.12512

[10] Frauenfelder, U., van Koert, O.: The Restricted Three-body Problem and Holomorphic Curves, Pathways in Mathematics. Birkhäuser/Springer, Cham (2018). (**MR3837531**)

[11] Fried, D.: The geometry of cross sections to flows. Topology **21**(4), 353–371 (1982). https://doi.org/10.1016/0040-9383(82)90017-9

[12] Ghys, É.: Right-handed vector fields & the Lorenz attractor. Jpn. J. Math. **4**(1), 47–61 (2009). https://doi.org/10.1007/s11537-009-0854-8

[13] Hofer, H.: Pseudoholomorphic curves in symplectizations with applications to the Weinstein conjecture in dimension three. Invent. Math. **114**(3), 515–563 (1993). https://doi.org/10.1007/BF01232679

[14] Hofer, H.: Holomorphic curves and real three-dimensional dynamics. Geom. Funct. Anal. Special Volume, Part II (2000), 674–704. GAFA 2000 (Tel Aviv, 1999)

[15] Hofer, H.: A general Fredholm theory and applications. Curr. Dev. Math. **2006**, 1–71 (2004)

[16] Hofer, H., Lizan, V., Sikorav, J.-C.: On genericity for holomorphic curves in four-dimensional almost-complex manifolds. J. Geom. Anal. **7**(1), 149–159 (1997). https://doi.org/10.1007/BF02921708

[17] Hofer, H., Wysocki, K., Zehnder, E.: Properties of pseudoholomorphic curves in symplectisations. I. asymptotics. Ann. Inst. H. Poincaré Anal. Non Linéaire 13(3), 337–379 (1996). https://doi.org/10.1016/s0294-1449(16)30108-1

[18] Hofer, H., Wysocki, K., Zehnder, E.: Properties of pseudo-holomorphic curves in symplectisations. II. Embedding controls and algebraic invariants. Geom. Funct. Anal. **5**(2), 270–328 (1995). https://doi.org/10.1007/BF01895669

[19] Hofer, H., Wysocki, K., Zehnder, E.: A characterization of the tight 3-sphere II. Commun. Pure Appl. Math. **52**(9), 1139–1177 (1999). https://doi.org/10.1002/(SICI)1097-0312(199909)52:9⟨3c1139::AID-CPA5?⟩3e3.3.CO;2-C

[20] Hofer, H., Wysocki, K., Zehnder, E.: Properties of Pseudoholomorphic Curves in Symplectizations. Topics in Nonlinear Analysis, III. Fredholm Theory, pp. 381–475 (1999)

[21] Hofer, H., Wysocki, K., Zehnder, E.: The dynamics on three-dimensional strictly convex energy surfaces. Ann. Math. (2) **148**(1), 197–289 (1998). https://doi.org/10.2307/120994

[22] Hofer, H., Wysocki, K., Zehnder, E.: Finite energy cylinders of small area. Ergod. Theory Dyn. Syst. **22**(5), 1451–1486 (2002). https://doi.org/10.1017/S0143385702001013

[23] Hryniewicz, U.: Fast finite-energy planes in symplectizations and applications. Trans. Am. Math. Soc. **364**(4), 1859–1931 (2012). https://doi.org/10.1090/S0002-9947-2011-05387-0

[24] Hryniewicz, U.: Systems of global surfaces of section for dynamically convex Reeb flows on the 3-sphere. J. Symplectic Geom. **12**(4), 791–862 (2014)

[25] Hryniewicz, U.: Finite energy foliations of convex sets in R4, 176. Thesis (Ph.D.), New York University (2008)

[26] Hryniewicz, U.: A note on Schwartzman–Fried–Sullivan theory, with an application. J. Fixed Point Theory Appl. **22**(1), 25 (2020)

[27] Hryniewicz, U., Licata, J.E., Salomão, P.A.S.: A dynamical characterization of universally tight lens spaces. Proc. Lond. Math. Soc. (3) **110**(1), 213–269 (2015)

[28] Hryniewicz, U., Salomão, P.A.S.: On the existence of disk-like global sections for Reeb flows on the tight 3-sphere. Duke Math. J. **160**(3), 415–465 (2011)

[29] Hryniewicz, U., Salomão, P.A.S.: Elliptic bindings for dynamically convex Reeb flows on the real projective three-space. Calc. Var. Partial Differ. Equ. **55**(2), 43 (2016)

[30] Hryniewicz, U., Hutchings, M., Ramos, V.: (**In preparation**)

[31] Hryniewicz, U., Salomão, P.A.S., Wysocki, K.: Genus zero global surfaces of section for Reeb flows and a result of Birkhoff. J. Eur. Math. Soc. arXiv:1912.01078v3 (**To appear**)

[32] Hutchings, M., Nelson, J.: Cylindrical contact homology for dynamically convex contact forms in three dimensions. J. Symplectic Geom. **14**(4), 983–1012 (2016)

[33] van Koert, O.: A Reeb flow on the three-sphere without a disk-like global surface of section. Qual. Theory Dyn. Syst. **19**(1), 36 (2020)

[34] van Koert, O., Moreno, A.: Global hypersurfaces of section in the spatial restricted three-body problem. arXiv:2011.10386

[35] Lhuissier, M.: Le probl'eme mathématique des trois corps, abordé simultanément sous l'angle de la recherche théorique et celui de la diffusion aupr'es de publics variés, Th'ese de doctorat (2019)

[36] Mora-Donato, E.: Pseudoholomorphic cylinders in symplectisations, 101. Thesis (Ph.D.), New York University (2003)

[37] Siefring, R.: Intersection theory of punctured pseudoholomorphic curves. Geom. Topol. **15**(4), 2351–2457 (2011)

[38] Siefring, R.: Relative asymptotic behavior of pseudoholomorphic half-cylinders. Commun. Pure Appl. Math. **61**(12), 1631–1684 (2008)

[39] Schwartzman, S.: Asymptotic cycles. Ann. Math. (2) **66**, 270–284 (1957)

[40] Sullivan, D.: Cycles for the dynamical study of foliated manifolds and complex manifolds. Invent. Math. **36**, 225–255 (1976). https://doi.org/10.1007/BF01390011

[41] Wendl, C.: Automatic transversality and orbifolds of punctured holomorphic curves in dimension four. Comment. Math. Helv. **85**(2), 347–407 (2010)

Umberto L. Hryniewicz
RWTH Aachen
Pontdriesch 10-12
Aachen 52062
Germany
e-mail: hryniewicz@mathga.rwth-aachen.de

Pedro A. S. Salomão
Departamento de Matemática, Instituto de Matemática e Estatística
Universidade de São Paulo
Rua do Matão, 1010, Cidade Universitária
São Paulo SP05508-090
Brazil
e-mail: psalomao@ime.usp.br;
pas383@nyu.edu

and

NYU-ECNU Institute of Mathematical Sciences at NYU Shanghai
3663 Zhongshan Road North
Shanghai 200062
China

Richard Siefring
Mathematisches Institut
Ruprecht-Karls-Universität Heidelberg
Im Neuenheimer Feld 205
69120 Heidelberg
Germany
e-mail: siefring@mathi.uni-heidelberg.de

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:50
https://doi.org/10.1007/s11784-022-00968-3
Published online June 11, 2022

Journal of Fixed Point Theory and Applications

ECH capacities and the Ruelle invariant

Michael Hutchings

Abstract. The ECH capacities are a sequence of real numbers associated to any symplectic four-manifold, which are monotone with respect to symplectic embeddings. It is known that for a compact star-shaped domain in $\mathbb{R}^4$, the ECH capacities asymptotically recover the volume of the domain. We conjecture, with a heuristic argument, that generically the error term in this asymptotic formula converges to a constant determined by a "Ruelle invariant" which measures the average rotation of the Reeb flow on the boundary. Our main result is a proof of this conjecture for a large class of toric domains. As a corollary, we obtain a general obstruction to symplectic embeddings of open toric domains with the same volume. For more general domains in $\mathbb{R}^4$, we bound the error term with an improvement on the previously known exponent from 2/5 to 1/4.

Mathematics Subject Classification. 37E99, 53D42.

1. Introduction

1.1. Asymptotics of ECH capacities

Given a symplectic 4-manifold (X, ω), possibly noncompact or with boundary, there is associated a sequence of real numbers

$$0 = c_0(X, \omega) < c_1(X, \omega) \leq c_2(X, \omega) \leq \cdots \leq \infty, \tag{1.1}$$

called the *ECH capacities* of (X, ω). These were defined in [19] using embedded contact homology; see [20] for a survey. Some basic properties of ECH capacities proved in [19] are:

- (Monotonicity) If there exists a symplectic embedding of (X, ω) into (X', ω') then

$$c_k(X, \omega) \leq c_k(X', \omega') \tag{1.2}$$

for all k.

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder and Felix Schlenk..

- (Conformality) If $r > 0$ then

$$c_k(X, r\omega) = rc_k(X, \omega). \tag{1.3}$$

- (Disjoint unions) Given a (possibly finite) sequence[1] of symplectic 4-manifolds $\{(X_i, \omega_i)\}$, we have

$$c_k\left(\coprod_i (X_i, \omega_i)\right) = \sup_{\sum_i k_i = k} \sum_i c_{k_i}(X_i, \omega_i). \tag{1.4}$$

- (Balls) If $a > 0$, define the ball

$$B(a) = \left\{z \in \mathbb{C}^2 \;\middle|\; \pi|z|^2 \le a\right\}.$$

Then

$$c_k(B(a)) = da \tag{1.5}$$

where d is the unique nonnegative integer such that

$$d^2 + d \le 2k \le d^2 + 3d.$$

- (Volume property) If X is a compact domain in $\mathbb{R}^4$ with piecewise smooth boundary, then

$$\lim_{k\to\infty} \frac{c_k(X)^2}{k} = 4\,\mathrm{vol}(X). \tag{1.6}$$

Here for domains in $\mathbb{R}^4 = \mathbb{C}^2$ we always take the restriction of the standard symplectic form

$$\omega = \sum_{i=1}^{2} \mathrm{d}x_i\, \mathrm{d}y_i.$$

The symplectic embedding obstructions resulting from the monotonicity property (1.2) are sharp in some cases, for example when X and X' are ellipsoids in $\mathbb{R}^4$, as shown by McDuff [25], or more generally when X is a "concave toric domain" and X' is a "convex toric domain", as shown by Cristofaro-Gardiner [8].

Define a "nice star-shaped domain" to be a compact domain in $\mathbb{R}^4$ whose boundary is smooth and transverse to the radial vector field. If X is a nice star-shaped domain, then the asymptotic formula (1.6) is a special case of a more general result about the asymptotics of the "ECH spectrum" of a contact three-manifold, which was proved in [11] using Seiberg–Witten theory. The formula (1.6) for nice star-shaped domains corresponds to the case when the contact three-manifold is the boundary of X, which of course is diffeomorphic to S^3, together with an induced contact form (see (1.12) below) whose kernel is the tight contact structure.

The ECH spectrum of a contact three-manifold is defined in terms of the periods of certain Reeb orbits, and as a result the asymptotic formula for the ECH spectrum has various applications to dynamics. In particular,

[1] In [19] it was assumed that the sequence of symplectic manifolds $\{(X_i, \omega_i)\}$ is finite, and in that case one has 'max' instead of 'sup' in (1.4). The countable case follows directly from the finite case using the definition of ECH capacities in [19].

[9] deduces the existence of at least two simple Reeb orbits; [10] proves the existence of either two or infinitely many simple Reeb orbits under certain hypotheses; [4,23] obtain C^∞ generic density of Reeb orbits and periodic orbits of Hamiltonian surface diffeomorphisms, see also the survey [18]; and [22,31] obtain relations between periodic orbits of area preserving disk or annulus diffeomorphisms and the Calabi invariant.

Returning to symplectic embedding problems, the asymptotic formula (1.6) implies that for k large, the symplectic embedding obstruction (1.2) recovers the obvious volume constraint $\operatorname{vol}(X) \le \operatorname{vol}(X')$. Additional embedding obstructions arise from the deviation of $c_k(X)$ from the asymptotics in (1.6). More precisely, define the "error term"

$$e_k(X) = c_k(X) - 2\sqrt{k \operatorname{vol}(X)} \tag{1.7}$$

It is then interesting to try to understand the size of this error term and its geometric significance.

A result of Sun [29] implies that if X is a nice star-shaped domain, then

$$e_k(X) = O\left(k^{125/252}\right).$$

The exponent was improved by Cristofaro-Gardiner and Savale [13] to 2/5. Both of these results for nice star-shaped domains are special cases of general results on the asymptotics of the ECH spectrum of a contact three-manifold, proved using Seiberg–Witten theory.

We use more elementary arguments to further improve the exponent for domains in $\mathbb{R}^4$:

Theorem 1.1. (proved in Sect. 4) *If X is a compact domain in $\mathbb{R}^4$ with smooth boundary* (*not necessarily star-shaped*), *then*

$$e_k(X) = O\left(k^{1/4}\right).$$

In fact, $e_k(X)$ is $O(1)$ in all examples for which it has been computed.

Example 1.2. Let X be the ball $B(a)$. We have $\operatorname{vol}(B(a)) = a^2/2$, see (1.14) below. By (1.5), we then have

$$e_k(B(a)) = \left(d - \sqrt{2k}\right) a,$$

where d is the unique nonnegative integer such that

$$d^2 + d \le 2k \le d^2 + 3d.$$

It follows from the above two lines that

$$\begin{aligned} \liminf_{k\to\infty} e_k(B(a)) &= -\frac{3}{2}a, \\ \limsup_{k\to\infty} e_k(B(a)) &= -\frac{1}{2}a. \end{aligned} \tag{1.8}$$

More generally, [32, Thm. 1.1] implies that for certain "lattice convex toric domains", e_k is also $O(1)$ with a more complicated oscillating behavior.

1.2. The Ruelle invariant

We now formulate a general conjecture about the limiting behavior of the error term e_k. This requires a digression to define the "Ruelle invariant" of a contact form on a homology three-sphere, which can be regarded as a measure of the average rotation rate of the Reeb flow. (One can also define the Ruelle invariant more generally for volume-preserving vector fields.)

Let $\widetilde{\mathrm{Sp}}(2)$ denote the universal cover of the group $\mathrm{Sp}(2)$ of 2×2 symplectic matrices. There is a standard "rotation number" function

$$\mathrm{rot} : \widetilde{\mathrm{Sp}}(2) \longrightarrow \mathbb{R}$$

defined as follows. Let $A \in \mathrm{Sp}(2)$, and let $\widetilde{A} \in \widetilde{\mathrm{Sp}}(2)$ be a lift of A, represented by a path $\{A_t\}_{t\in[0,1]}$ in $\mathrm{Sp}(2)$ with $A_0 = I$ and $A_1 = A$. If v is a nonzero vector in $\mathbb{R}^2$, then the path of vectors $\{A_t v\}_{t\in[0,1]}$ rotates by some angle which we denote by $2\pi\rho(v) \in \mathbb{R}$. We then define

$$\mathrm{rot}\left(\widetilde{A}\right) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \rho\left(A^{k-1}v\right).$$

This does not depend on the choice of nonzero vector v. For example, if A is conjugate to rotation by angle $2\pi\theta$, then $\mathrm{rot}\left(\widetilde{A}\right)$ is a lift of θ from $\mathbb{R}/2\pi\mathbb{Z}$ to $\mathbb{R}$. The rotation number is a quasimorphism: if $\widetilde{B}$ is another element of $\widetilde{\mathrm{Sp}}(2)$, then

$$\left|\mathrm{rot}\left(\widetilde{A}\widetilde{B}\right) - \mathrm{rot}\left(\widetilde{A}\right) - \mathrm{rot}\left(\widetilde{B}\right)\right| < 1. \tag{1.9}$$

Now let Y be a homology three-sphere, and let λ be a contact form on Y with associated contact structure ξ and Reeb vector field R. For $t \in \mathbb{R}$, let $\phi_t : Y \to Y$ denote the diffeomorphism given by the time t Reeb flow. For each $y \in Y$, the derivative of ϕ_t restricts to a linear map

$$d\phi_t : \xi_y \longrightarrow \xi_{\phi_t(y)} \tag{1.10}$$

which is symplectic with respect to $d\lambda$. Now fix a symplectic trivialization of ξ, consisting of a symplectic linear map $\tau : \xi_y \to \mathbb{R}^2$ for each $y \in Y$. Then for $y \in Y$ and $t \in \mathbb{R}$, the composition

$$\mathbb{R}^2 \xrightarrow{\tau^{-1}} \xi_y \xrightarrow{d\phi_t} \xi_{\phi_t(y)} \xrightarrow{\tau} \mathbb{R}^2$$

is a symplectic matrix which we denote by $A^\tau_{y,t}$. In particular, if $y \in Y$ and $T \geq 0$, then the path of symplectic matrices $\{A^\tau_{y,t}\}_{t\in[0,T]}$ defines an element of $\widetilde{\mathrm{Sp}}(2)$. We denote its rotation number by

$$\mathrm{rot}_\tau(y,T) = \mathrm{rot}\left(\{A^\tau_{y,t}\}_{t\in[0,T]}\right) \in \mathbb{R}.$$

As explained by Ruelle [28], see also [14, §3.2], one can use the quasimorphism property (1.9) to show that for almost all $y \in Y$, the limit

$$\rho(y) = \lim_{T\to\infty} \frac{1}{T}\,\mathrm{rot}_\tau(y,T)$$

is well defined and independent of τ, and the function ρ is integrable.

Definition 1.3. If Y is a homology three-sphere and λ is a contact form on Y, define the *Ruelle invariant*

$$\mathrm{Ru}(Y,\lambda) = \int_Y \rho\,\lambda\wedge \mathrm{d}\lambda. \tag{1.11}$$

If X is a nice star-shaped domain in $\mathbb{R}^4$, then the standard Liouville form

$$\lambda_0 = \frac{1}{2}\sum_{i=1}^{2}(x_i\,\mathrm{d}y_i - y_i\,\mathrm{d}x_i) \tag{1.12}$$

restricts to a contact form on ∂X.

Definition 1.4. If X is a nice star-shaped domain in $\mathbb{R}^4$, then we define

$$\mathrm{Ru}(X) = \mathrm{Ru}\left(\partial X, \lambda_0|_{\partial X}\right).$$

We can now state our main conjecture:

Conjecture 1.5. *If X is a generic nice star-shaped domain in $\mathbb{R}^4$, then*

$$\boxed{\lim_{k\to\infty} e_k(X) = -\frac{1}{2}\mathrm{Ru}(X).} \tag{1.13}$$

Example 1.6. The ball $B(a)$ from Example 1.2 does not satisfy the above conjecture (hence the word "generic" in the conjecture), since $e_k(B(a))$ does not converge. However we will see below that $\mathrm{Ru}(B(a)) = 2a$, so it is still true that $(-1/2)\,\mathrm{Ru}(B(a))$ is between the lim inf and lim sup of $e_k(B(a))$. One might conjecture that for any nice star-shaped domain, not necessarily generic, e_k is $O(1)$ and the Ruelle invariant is between the lim inf and the lim sup.

1.3. Results for toric domains

Given a domain Ω in the nonnegative quadrant of $\mathbb{R}^2$, we define an associated *toric domain*

$$X_\Omega = \left\{z\in\mathbb{C}^2 \;\middle|\; \pi(|z_1|^2,|z_2|^2)\in\Omega\right\}.$$

The factor of π ensures among other things that

$$\mathrm{vol}(X_\Omega) = \mathrm{area}(\Omega). \tag{1.14}$$

Definition 1.7. A *nice toric domain* is a toric domain X_Ω which is also a nice star-shaped domain, meaning that ∂X_Ω is a smooth hypersurface transverse to the radial vector field. This implies that $\partial\Omega$ consists of the line segment from $(0,0)$ to $(a,0)$ for some $a>0$, the line segment from $(0,0)$ to $(0,b)$ for some $b>0$, and a smooth curve from $(0,b)$ to $(a,0)$ which is transverse to the radial vector field on $\mathbb{R}^2$. We denote the numbers a and b by $a(\Omega)$ and $b(\Omega)$, and the smooth curve from $(0,b)$ to $(a,0)$ by $\partial_+\Omega$.

Example 1.8. Suppose Ω is the triangle with vertices $(0,0)$, $(a,0)$, and $(0,b)$. Then X_Ω is the ellipsoid

$$E(a,b) = \left\{z\in\mathbb{C}^2 \;\middle|\; \frac{\pi|z_1|^2}{a}+\frac{\pi|z_2|^2}{b}\le 1\right\}.$$

This is a nice toric domain.

Definition 1.9. A *strictly convex toric domain* is a nice toric domain X_Ω in which $\partial_+\Omega$ is the graph of a function $f : [0,a] \to [0,b]$ with $f(0) = b$, $f'(0) < 0$, $f'' < 0$ everywhere, and $f(a) = 0$.

A *strictly concave toric domain* is a nice toric domain X_Ω in which $\partial_+\Omega$ is the graph of a function $f : [0,a] \to [0,b]$ with $f(0) = b$, $f'' > 0$ everywhere, and $f(a) = 0$.

We can now state one of the main results of this paper:

Theorem 1.10. (proved in Sect. 3) *Equation* (1.13) *holds whenever X is a strictly convex or strictly concave toric domain.*[2]

To clarify what this theorem says, we have:

Proposition 1.11. (proved in Sect. 2) *Let X_Ω be a nice toric domain such that $\partial_+\Omega$ has negative slope*[3] *everywhere. Then*

$$\mathrm{Ru}(X_\Omega) = a(\Omega) + b(\Omega).$$

Remark 1.12. Equation (1.13) also holds for ellipsoids $E(a, b)$ with a/b irrational, by [12, Lem. 2.2].

It is quite possible that Eq. (1.13) is special to toric domains and that Conjecture 1.5 is false more generally. Nonetheless, the toric case already gives an application to symplectic embedding problems:

Corollary 1.13. *Let X_Ω and $X_{\Omega'}$ be nice toric domains satisfying* (1.13), *e.g. strictly convex or strictly concave toric domains, or irrational ellipsoids. Suppose that* $\mathrm{vol}(X_\Omega) = \mathrm{vol}(X_{\Omega'})$ *and that there exists a symplectic embedding* $\mathrm{int}(X_\Omega) \to X_{\Omega'}$. *Then*

$$a(\Omega) + b(\Omega) \geq a(\Omega') + b(\Omega').$$

Proof. The interior of X_Ω has the same ECH capacities as X_Ω; see [19, §4.2]. Thus, by the monotonicity of the ECH capacities (1.2), the definition of the error term (1.7), and the hypothesis that $\mathrm{vol}(X_\Omega) = \mathrm{vol}(X_{\Omega'})$, we have

$$e_k(X_\Omega) \leq e_k(X_{\Omega'})$$

for all k. Since X_Ω and $X_{\Omega'}$ satisfy (1.13), it follows from Proposition 1.11 that

$$\frac{-(a(\Omega) + b(\Omega))}{2} \leq \frac{-(a(\Omega') + b(\Omega'))}{2}.$$

□

[2] It is shown in [33] that Theorem 1.10 generalizes to (not necessarily strictly) convex and concave toric domains such that $\partial_+\Omega$ has no edges of rational slope.

[3] For nice toric domains in $\mathbb{R}^4$, the condition that $\partial_+\Omega$ has negative slope is equivalent to dynamical convexity by [16, Prop. 1.8]. In fact, the negative slope hypothesis can be removed from Proposition 1.11 by a more careful argument [17].

Remark 1.14. Corollary 1.13 is not vacuous; there are examples of symplectic embeddings of an open toric domain into another (nonsymplectomorphic) toric domain of the same volume, including many cases when the domains are ellipsoids. For example, it is shown in [27] that if $a \geq (17/6)^2$, then the interior of the ellipsoid $E(1,a)$ symplectically embeds into a ball[4] of the same volume, namely $E(\sqrt{a},\sqrt{a})$.

Remark 1.15. The examples of nice star-shaped domains X discussed here seem to have $e_k(X)$ negative for all $k > 0$. However there also exist examples of nice star-shaped domains $X \subset \mathbb{R}^4$ with $e_1(X)$ positive. The reason is that if X is a nice star-shaped domain, then by the definition of ECH capacities, $c_1(X) \geq \mathcal{A}_{\min}(X)$, where $\mathcal{A}_{\min}(X)$ denotes the minimum symplectic action (period) of a Reeb orbit on ∂X. Now define the *systolic ratio*

$$\operatorname{sys}(X) = \frac{\mathcal{A}_{\min}(X)^2}{2\operatorname{vol}(X)}.$$

It then follows from (1.7) that

$$e_1(X) \leq 0 \Longrightarrow \operatorname{sys}(X) \leq 2.$$

However it is shown in [1] that there exist nice star-shaped domains with systolic ratio greater than 2 (in fact arbitrarily large), so these must have e_1 positive.

On the other hand, in the *dynamically convex* case, the best known examples [2] have systolic ratio $2-\epsilon$. A reasonable conjecture would be that if X is dynamically convex then $e_k(X) < 0$ for all $k > 0$.

1.4. Outline of the rest of the paper

In Sect. 2 we prove Proposition 1.11, computing the Ruelle invariant of some toric domains, by direct calculation.

In Sect. 3 we prove the main result, Theorem 1.10. To do so, we use two formulas for the ECH capacities of concave toric domains proved in [7]: one in terms of the "weight expansion", and one in terms of lattice paths. We also use two similar formulas for the ECH capacities of convex toric domains from [8]. By carefully estimating using all four of these formulas and combining the results with Proposition 1.11, we obtain the theorem.

In Sect. 4 we prove Theorem 1.1. The idea is to estimate the ECH capacities of a region by packing it with cubes in a naive way. The estimates we get in this case are not as good as in the case of toric domains, because concave toric domains can be packed "more efficiently" with balls coming from the weight expansion.

In Sect. 5 we give a heuristic discussion of why we expect Conjecture 1.5 to be true, by comparing the definition of the ECH index to Arnold's asymptotic linking number and relating this to a conjecture by Irie on equidistribution properties of ECH capacities. While this is far from a proof, we do see the volume and Ruelle invariant emerge naturally.

[4] Although Corollary 1.13 is not applicable here because the ball does not satisfy (1.13), the conclusion of Corollary 1.13 is still true in this example since $1 + a \geq 2\sqrt{a}$.

2. The Ruelle invariant of toric domains

We now prove Proposition 1.11, computing the Ruelle invariant of a nice toric domain X_Ω such that $\partial_+\Omega$ has everywhere negative slope.

To start, we denote the Euclidean coordinates on the plane in which Ω lives by μ_1 and μ_2. Define two functions

$$\alpha, \beta : \partial_+\Omega \longrightarrow \mathbb{R}$$

as follows: Given $(\mu_1, \mu_2) \in \partial_+\Omega$, the tangent line to $\partial_+\Omega$ through (μ_1, μ_2) intersects the axes at the points $(\alpha(\mu_1, \mu_2), 0)$ and $(0, \beta(\mu_1, \mu_2))$.

Proposition 1.11 now follows from the two lemmas below:

Lemma 2.1. *If X_Ω is a nice toric domain such that $\partial_+\Omega$ has everywhere negative slope, then*

$$\mathrm{Ru}(X_\Omega) = \int_{\partial_+\Omega} \frac{\alpha+\beta}{\alpha\beta}(\mu_1\,\mathrm{d}\mu_2 - \mu_2\,\mathrm{d}\mu_1) \tag{2.1}$$

where $\partial_+\Omega$ is oriented as a curve from $(a(\Omega), 0)$ to $(0, b(\Omega))$.

Lemma 2.2. *If γ is a differentiable plane curve from $(a, 0)$ to $(0, b)$ with everywhere negative slope, where $a, b > 0$, and if α and β are defined as above, then*

$$\int_\gamma \frac{\alpha+\beta}{\alpha\beta}(\mu_1\,\mathrm{d}\mu_2 - \mu_2\,\mathrm{d}\mu_1) = a + b.$$

Proof. Write $Y = \partial X_\Omega \subset \mathbb{C}^2$, and let Y_0 denote the set of $z \in Y$ such that $z_1, z_2 \neq 0$. For $z = (z_1, z_2) \in Y_0$, write $\mu_i = \pi|z_i|^2$, and let θ_i denote the argument of z_i. In these coordinates, the standard Liouville form (1.12) is given by

$$\lambda_0 = \frac{1}{2\pi}\left(\mu_1\,\mathrm{d}\theta_1 + \mu_2\,\mathrm{d}\theta_2\right). \tag{2.2}$$

We have

$$T_zY = \mathrm{span}\left(\partial_{\theta_1}, \partial_{\theta_2}, \alpha\partial_{\mu_1} - \beta\partial_{\mu_2}\right).$$

Thus the contact plane ξ_z is spanned by the vectors

$$V = \mu_2\partial_{\theta_1} - \mu_1\partial_{\theta_2},$$
$$W = \alpha\partial_{\mu_1} - \beta\partial_{\mu_2}.$$

The Reeb vector field is then given by

$$R = \frac{2\pi\left(\beta\partial_{\theta_1} + \alpha\partial_{\theta_2}\right)}{\alpha\beta}. \tag{2.3}$$

Note here that $\lambda_0(R) = 1$ because

$$\beta\mu_1 + \alpha\mu_2 = \alpha\beta \tag{2.4}$$

by the definition of α and β. Equation (2.4) also implies that we have a symplectic trivialization τ' of $\xi|_{Y_0}$ given by

$$(\tau')^{-1} = \left(V, \frac{-2\pi W}{\alpha\beta}\right).$$

Since R preserves μ_1 and μ_2, we have $[R, V] = 1$, so in the notation (1.10) we have $\mathrm{d}\phi_t V = V$. This implies that

$$\mathrm{rot}_{\tau'}(y, T) = 0$$

for all $y \in Y_0$ and $T \geq 0$. However we cannot use the trivialization τ' to compute the Ruelle invariant because this trivialization does not extend over $Y \backslash Y_0$. In particular, if τ is a trivialization of ξ over all of Y, then as one moves around a circle in Y_0 in which either θ_1 or θ_2 rotates once around S^1, the vector V rotates once around S^1 with respect to τ. It follows that on Y_0 we have

$$\rho = \frac{1}{2\pi} R(\theta_1 + \theta_2).$$

By Eq. (2.3), we conclude that

$$\rho = \frac{\alpha + \beta}{\alpha\beta}. \tag{2.5}$$

Now by Eq. (2.2), we have

$$\lambda_0 \wedge \mathrm{d}\lambda_0 = \frac{1}{4\pi^2}(\mu_1 \,\mathrm{d}\mu_2 - \mu_2 \,\mathrm{d}\mu_1)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2$$

on Y_0. So by Eqs. (1.11) and (2.5) we have

$$\mathrm{Ru}(X_\Omega) = \frac{1}{4\pi^2} \int_{Y_0} \frac{\alpha + \beta}{\alpha\beta}(\mu_1 \,\mathrm{d}\mu_2 - \mu_2 \,\mathrm{d}\mu_1)\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2.$$

Integrating out θ_1 and θ_2 then gives (2.1). □

Proof of Lemma 2.2. Choose an oriented parametrization of the curve γ as $(\mu_1(t), \mu_2(t))$ for $t \in [t_0, t_1]$. Then

$$\int_\gamma \frac{\alpha + \beta}{\alpha\beta}(\mu_1 \,\mathrm{d}\mu_2 - \mu_2 \,\mathrm{d}\mu_1) = \int_{t_0}^{t_1} \frac{\alpha + \beta}{\alpha\beta} \Delta \mathrm{d}t \tag{2.6}$$

where we use the notation

$$\Delta = \mu_1 \mu_2' - \mu_1' \mu_2.$$

By the definition of α and β, we have

$$\alpha = \Delta/\mu_2',$$
$$\beta = -\Delta/\mu_1'.$$

The integrand in (2.6) is then

$$\frac{\alpha + \beta}{\alpha\beta}\Delta = -\mu_1' + \mu_2'.$$

The lemma now follows from the fundamental theorem of calculus. □

3. Bounds on ECH capacities of toric domains

3.1. The Ruelle invariant and the weight expansion

To relate the Ruelle invariant to ECH capacities, we need to recall the definition of the "weight expansion" of a concave toric domain following [7].

Definition 3.1. A *concave toric domain* is a toric domain X_Ω such that

$$\Omega = \{(\mu_1, \mu_2) \mid 0 \le \mu_1 \le a,\ 0 \le \mu_2 \le f(\mu_1)\}$$

where $f : [0, a] \to [0, b]$ is a convex function[5] for some $a, b > 0$ with $f(0) = b$ and $f(a) = 0$. Write $a(\Omega) = a$ and $b(\Omega) = b$, and denote the graph of f by $\partial_+\Omega$.

For $c > 0$, let $\Delta(c)$ denote the triangle in the plane with vertices $(0, 0)$, $(c, 0)$, and $(0, c)$. Also, define an *integral affine transformation* to be a map $\mathbb{R}^2 \to \mathbb{R}^2$ given by the composition of an element of $\mathrm{SL}_2\,\mathbb{Z}$ with a translation. We say that two sets in $\mathbb{R}^2$ are *integral affine equivalent* if one is the image of the other under an integral affine transformation.

Definition 3.2. If X_Ω is a concave toric domain, we inductively define a canonical countable set $\mathcal{T}(\Omega)$ of triangles in $\mathbb{R}^2$ such that:

(i) Each triangle in $\mathcal{T}(\Omega)$ is integral affine equivalent to $\Delta(c)$ for some c.
(ii) Two different triangles in $\mathcal{T}(\Omega)$ intersect only along their boundaries.
(iii) $\overline{\bigcup_{T \in \mathcal{T}(\Omega)} T} = \Omega$.

To start defining $\mathcal{T}(\Omega)$, let c be the largest real number such that the triangle $\Delta(c) \subset \Omega$.

Now $\partial_+\Delta(c)$ coincides with $\partial_+\Omega$ on the line segment from $(t', c - t')$ to $(t'', c - t'')$ for some $t' \le t''$. If $t' > 0$, let Ω' denote the closure of the component of $\Omega \setminus \Delta(c)$ with $\mu_1 \le t'$; otherwise let $\Omega' = \emptyset$. If $t'' < c$, let Ω'' denote the closure of the component of $\Omega \setminus \Delta(c)$ with $\mu_1 \ge t''$; otherwise let $\Omega'' = \emptyset$.

Let $\phi' : \mathbb{R}^2 \to \mathbb{R}^2$ denote the integral affine transformation defined by

$$\phi'(\mu_1, \mu_2) = (\mu_1, \mu_1 + \mu_2 - c).$$

If Ω' is nonempty, then $X_{\phi'(\Omega')}$ is a concave toric domain. Likewise, let ϕ'' denote the integral affine transformation defined by

$$\phi''(\mu_1, \mu_2) = (\mu_1 + \mu_2 - c, \mu_2).$$

If Ω'' is nonempty then $X_{\phi''(\Omega'')}$ is a concave toric domain.

We now inductively define

$$\mathcal{T}(\Omega) = \{\Delta(c)\} \cup \bigsqcup_{T \in \mathcal{T}(\phi'(\Omega'))} (\phi')^{-1}(T) \cup \bigsqcup_{T \in \mathcal{T}(\phi''(\Omega''))} (\phi'')^{-1}(T).$$

Here we interpret the terms involving Ω' or Ω'' to be the empty set when Ω' or Ω'' are empty.

---

[5] This is more general than a strictly concave toric domain as in Definition 1.9. For a strictly concave toric domain, the function f must furthermore be smooth and strictly convex, and must satisfy additional conditions near 0 and a to ensure that ∂X_Ω is smooth.

Properties (i) and (ii) above are immediate from the construction. It also follows from the construction that each triangle in $\mathcal{T}(\Omega)$ is a subset of Ω. One can prove the rest of property (iii) by elementary arguments with a bit more work; or as overkill one can use Eq. (3.1) below and the volume property of ECH capacities (1.6).

Definition 3.3. If X_Ω is a concave toric domain, choose an ordering $\mathcal{T}(\Omega) = \{T_1, T_2, \ldots\}$ where T_i is integral affine equivalent to $\Delta(a_i)$ and $a_i \geq a_{i+1}$ for each i. The (possibly finite) sequence $(a_1, a_2, \ldots)$ is the *weight expansion* of X_Ω, which we denote by $W(\Omega)$.

The significance of the weight expansion is:

Theorem 3.4. [7, Thm. 1.4 and Rmk. 1.6] *If X_Ω is a concave toric domain with weight expansion $W(\Omega) = (a_1, \ldots)$, then its ECH capacities are given by*

$$c_k(X_\Omega) = c_k\left(\bigsqcup_i B(a_i)\right). \tag{3.1}$$

Note that by properties (i)–(iii) above, we have

$$\operatorname{vol}(X_\Omega) = \operatorname{area}(\Omega) = \frac{1}{2}\sum_i a_i^2.$$

It turns out that $\sum_i a_i$ is also finite, and can be described explicitly as follows.

Definition 3.5. Given a line segment L in the plane, define its *affine length* $\ell_{\mathrm{Aff}}(L) \in \mathbb{R}$ as follows. Let $v = (a, b)$ be the vector given by the difference between the endpoints of L.

- If $a/b \notin \mathbb{Q} \cup \{\infty\}$, define $\ell_{\mathrm{Aff}}(L) = 0$.
- If $a/b \in \mathbb{Q} \cup \{\infty\}$, let d be the largest real number such that $(a/d, b/d) \in \mathbb{Z}^2$, and define $\ell_{\mathrm{Aff}}(L) = d$.

If γ is an injective continuous path in the plane including line segments $L_1, \ldots$, define its affine length

$$\ell_{\mathrm{Aff}}(\gamma) = \sum_i \ell_{\mathrm{Aff}}(L_i).$$

Lemma 3.6. [26] *If X_Ω is a concave toric domain with weight expansion $W(\Omega) = (a_1, \ldots)$, then*

$$\sum_i a_i = a(\Omega) + b(\Omega) - \ell_{\mathrm{Aff}}(\partial_+\Omega). \tag{3.2}$$

Proof. Following the construction in Definition 3.2, we inductively define a sequence of domains Ω_k for $k \geq 1$ such that X_{Ω_k} is a concave toric domain, $\Omega_k \subset \Omega_{k+1}$, and $\overline{\bigcup_k \Omega_k} = \Omega$, as follows. Using the notation of Definition 3.2:

- $\Omega_1 = \Delta(c)$.

- If $k > 1$, then
$$\Omega_k = \Delta(c) \cup (\phi')^{-1}(\phi'(\Omega')_{k-1}) \cup (\phi'')^{-1}(\phi''(\Omega'')_{k-1}).$$
Here we omit the terms corresponding to Ω' or Ω'' when those domains are empty.

Observe that X_{Ω_k} has a finite weight expansion with at most $2^k - 1$ terms. Moreover these are all terms in the weight expansion of X_Ω; and if $S(\Omega)$ denotes the sum of the terms in the weight expansion $W(\Omega)$, then $\lim_{k\to\infty} S(\Omega_k) = S(\Omega)$.

We will prove by induction on k that for every concave toric domain X_Ω, we have
$$S(\Omega_k) = a(\Omega_k) + b(\Omega_k) - \ell_{\mathrm{Aff}}(\partial_+\Omega_k). \tag{3.3}$$
The lemma then follows by fixing Ω and taking the limit of (3.3) as $k \to \infty$.

If $k = 1$, then both sides of Eq. (3.3) are equal to c above.

Now suppose that $k > 1$. For simplicity we assume that both Ω' and Ω'' are nonempty; the other cases work similarly. By induction we can assume that
$$S(\Omega'_{k-1}) = a(\Omega'_{k-1}) + b(\Omega'_{k-1}) - \ell_{\mathrm{Aff}}(\partial_+\Omega'_{k-1}),$$
$$S(\Omega''_{k-1}) = a(\Omega''_{k-1}) + b(\Omega''_{k-1}) - \ell_{\mathrm{Aff}}(\partial_+\Omega''_{k-1}).$$
By construction we have
$$\begin{aligned} S(\Omega_k) &= c + S(\Omega'_{k-1}) + S(\Omega''_{k-1}),\\ a(\Omega_k) &= c + a(\Omega''_{k-1}),\\ b(\Omega_k) &= c + b(\Omega'_{k-1}). \end{aligned}$$
Combining the above equations, we obtain
$$\begin{aligned} S(\Omega_k) - a(\Omega_k) - b(\Omega_k) = -c &+ a(\Omega'_{k-1}) + b(\Omega''_{k-1}) - \ell_{\mathrm{Aff}}(\partial_+\Omega'_{k-1})\\ &-\ell_{\mathrm{Aff}}(\partial_+\Omega''_{k-1}). \end{aligned} \tag{3.4}$$

Now observe that $\partial_+\Omega_k$ consists of the following:

- The curve $(\phi')^{-1}(\partial_+\Omega'_{k-1})$ from $(0, c+b(\Omega'_{k-1}))$ to $(a(\Omega'_{k-1}), c-a(\Omega'_{k-1}))$.
- The line segment from the latter point to $(c - b(\Omega''_{k-1}), b(\Omega''_{k-1}))$.
- The curve $(\phi'')^{-1}(\partial_+\Omega''_{k-1})$ from the latter point to $(c + a(\Omega''_{k-1}), 0)$.

Since affine length is invariant under integral affine transformations, it follows that
$$\ell_{\mathrm{Aff}}(\partial_+\Omega_k) = \ell_{\mathrm{Aff}}\left(\partial_+\Omega'_{k-1}\right) + \left(c - a\left(\Omega'_{k-1}\right) - b\left(\Omega''_{k-1}\right)\right) + \ell_{\mathrm{Aff}}\left(\partial_+\Omega''_{k-1}\right).$$
Combining this last equation with (3.4) proves (3.3). □

As a corollary, we obtain a relation between the weight expansion and the Ruelle invariant in the strictly concave case:

Corollary 3.7. *If X_Ω is a strictly concave toric domain (or more generally any concave toric domain such that $\partial_+\Omega$ does not contain any line segments of rational slope) with weight expansion $W(\Omega) = (a_1, \ldots)$, then*
$$\sum_i a_i = a(\Omega) + b(\Omega). \tag{3.5}$$

Proof. This follows from Lemma 3.6 because $\partial_+\Omega$ contains no line segments of rational slope, so its affine length is zero. $\square$

3.2. An estimate from the weight expansion

Lemma 3.8. *Let $(a_i)_{i=1,\ldots}$ be a (possibly finite) sequence of positive real numbers with $\sum_i a_i < \infty$. Write $X = \coprod_i B(a_i)$ and $V = \mathrm{vol}(X) = \frac{1}{2}\sum_i a_i^2$. Then*

$$\limsup_{k\to\infty}\left(c_k(X) - 2\sqrt{kV}\right) \le -\frac{1}{2}\sum_i a_i.$$

Corollary 3.9. *If X_Ω is a concave toric domain such that $\partial_+\Omega$ does not contain any line segments of rational slope, then*

$$\limsup_{k\to\infty} e_k(X_\Omega) \le -\frac{a(\Omega)+b(\Omega)}{2}.$$

Proof. This follows from Lemma 3.8 by plugging in Eqs. (1.7), (3.1), (1.14), and (3.5). $\square$

Proof of Lemma 3.8. By Eqs. (1.4) and (1.5), we have

$$c_k(X) = \sup\left\{\sum_i a_i d_i \;\middle|\; \sum_i (d_i^2 + d_i) \le 2k\right\} \tag{3.6}$$

where the d_i are nonnegative integers. Now if we put the sequence (a_i) in nonincreasing order, then in the above supremum, we can restrict to the case where $d_i = 0$ for $i > k$. There are then only finitely many possibilities, so we can write 'max' instead of 'sup' in (3.6).

For each k, choose a sequence $d(k) = \{d(k)_i\}_{i=1,\ldots}$ realizing the maximum in (3.6). In particular, we have

$$\sum_i a_i d(k)_i = c_k(X), \tag{3.7}$$

$$\sum_i (d(k)_i^2 + d(k)_i) \le 2k. \tag{3.8}$$

By (3.8) and the Cauchy-Schwarz inequality, for each k we have

$$\sum_i a_i\sqrt{d(k)_i^2 + d(k)_i} \le \sqrt{2V}\sqrt{2k}.$$

Combining this with (3.7), we have

$$c_k(X) - 2\sqrt{kV} \le -\sum_i a_i\left(\sqrt{d(k)_i^2 + d(k)_i} - d(k)_i\right). \tag{3.9}$$

To complete the proof, it is enough to show that for fixed i we have

$$\lim_{k\to\infty} d(k)_i = \infty, \tag{3.10}$$

so that

$$\lim_{k\to\infty}\left(\sqrt{d(k)_i^2 + d(k)_i} - d(k)_i\right) = \frac{1}{2}.$$

To prove (3.10), suppose to the contrary that $\liminf_{k\to\infty} d(k)_i < \infty$. Then it follows similarly to (3.9) that

$$\lim\inf_{k\to\infty}\left(c_k(X) - 2\sqrt{k\left(V - \frac{1}{2}a_i^2\right)}\right) \le 0.$$

Thus

$$\lim\inf_{k\to\infty}\frac{c_k(X)^2}{k} \le 4\,\mathrm{vol}\,(X\backslash B(a_i)).$$

However the argument in [19, Prop. 8.4] shows that X satisfies the volume property (1.6), which is a contradiction. □

3.3. Lattice point estimates

If Ω is a domain in the nonnegative quadrant of $\mathbb{R}^2$, define

$$\widehat{\Omega} = \{(\mu_1,\mu_2)\in\mathbb{R}^2 \mid (|\mu_1|,|\mu_2|)\in\Omega\}.$$

Definition 3.10. A *convex toric domain* is a toric domain X_Ω such that $\widehat{\Omega}$ is compact and convex with nonempty interior. Let $a(\Omega)$ and $b(\Omega)$ denote the intersections of $\partial\widehat{\Omega}$ with the positive μ_1-axis and positive μ_2-axis, and let $\partial_+\Omega$ denote the closure of the part of ∂_Ω not on the axes; this is a path from $(0,b(\Omega))$ to $(a(\Omega),0)$.

We now prove the following estimate, which is similar to Corollary 3.9 but proved by different methods:

Lemma 3.11. *Let X_Ω be a convex toric domain such that $\partial_+\Omega$ is the graph of a strictly concave C^2 function.*[6] *Then*

$$\lim\sup_{k\to\infty} e_k(X_\Omega) \le -\frac{a(\Omega)+b(\Omega)}{2}.$$

To prove this lemma, we need to recall some material from [21]. Let Ω be a domain as in Definition 3.10. If v is a vector in $\mathbb{R}^2$, define

$$\|v\|_\Omega^* = \max\left\{\langle v,w\rangle \mid w\in\widehat{\Omega}\right\}.$$

Note that $\|\cdot\|_\Omega^*$ is a norm; it is the dual of the norm with unit ball $\widehat{\Omega}$. If $\gamma:[\alpha,\beta]\to\mathbb{R}^2$ is a continuous, piecewise differentiable parametrized curve, define its Ω-*length* by

$$\ell_\Omega(\gamma) = \int_\alpha^\beta \|J\gamma'(t)\|_\Omega^*\mathrm{d}t \tag{3.11}$$

where $J = \begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$. The Wulff isoperimetric inequality [5,34] implies that if γ is the boundary of a compact region R, then

$$\ell_\Omega(\gamma)^2 \ge 4\,\mathrm{Area}(\widehat{\Omega})\,\mathrm{Area}(R), \tag{3.12}$$

[6]This is slightly more general than a "strictly convex toric domain", because ∂X_Ω might not be smooth.

with equality if and only if R is a scaling and translation of $\widehat{\Omega}$. Below we just need to know that equality holds in (3.12) when R is a scaling of $\widehat{\Omega}$, which follows by direct calculation.

Definition 3.12. A *convex integral path* is a polygonal path Λ in the nonnegative quadrant from the point $(0,b)$ to the point $(a,0)$, for some nonnegative integers a and b, with vertices at lattice points, such that if R denotes the region bounded by Λ and the line segments from $(0,0)$ to $(a,0)$ and from $(0,0)$ to $(0,b)$, then $\widehat{R}$ is convex. Define $\mathcal{L}(\Lambda)$ to be the number of lattice points in R, including lattice points on the boundary.

We now have the following theorem,[7] proved in [21, Prop. 5.6], as a special case of [8, Cor. A.5]:

Theorem 3.13. *Let X_Ω be a convex toric domain. Then*

$$c_k(X) = \min\{\ell_\Omega(\Lambda) \mid \mathcal{L}(\Lambda) \geq k+1\}. \tag{3.13}$$

Here the minimum is over convex integral paths Λ.

Proof of Lemma 3.11. Given a positive integer k, let r be the smallest real number such that the scaling $r\Omega$ contains at least $k+1$ lattice points. The boundary of the convex hull of $r\Omega \cap \mathbb{Z}^2$ consists of a segment on the μ_1-axis, a segment on the μ_2-axis, and a convex integral path Λ with $\mathcal{L}(\Lambda) \geq k+1$. Thus by Theorem 3.13, we have

$$c_k(X_\Omega) \leq \ell_\Omega(\Lambda). \tag{3.14}$$

Next observe that

$$\ell_\Omega(\Lambda) \leq \ell_\Omega(\partial_+(r\Omega)). \tag{3.15}$$

The reason is that Λ can be obtained from $\partial_+(r\Omega)$ by a finite sequence of operations, each of which replaces a portion of a curve by a line segment with the same endpoints. These operations do not increase Ω-length since $\|\cdot\|_\Omega^*$ is a norm.

By the equality case of Wulff's isoperimetric inequality (3.12), we have

$$\ell_\Omega(\partial_+(r\Omega)) = 2\sqrt{\operatorname{Area}(\Omega)\operatorname{Area}(r\Omega)}.$$

By (1.14), we can rewrite the above as

$$\ell_\Omega(\partial_+(r\Omega)) = 2\sqrt{\operatorname{vol}(X_\Omega)\operatorname{Area}(r\Omega)}. \tag{3.16}$$

Next, a classical result of van der Korput, see the refinement by Chaix [6], asserts that if R is a region in the plane with C^2 strictly convex boundary, then

$$\left||R \cap \mathbb{Z}^2| - \operatorname{Area}(R)\right| \leq 10000(1+M)^{2/3},$$

where M denotes the maximum radius of curvature of ∂R. Taking $\epsilon > 0$ small and applying this result to $R = (r-\epsilon)\widehat{\Omega}$, with the intersections with the axes

[7] The statement in [21] looks slightly different, writing $\mathcal{L}(\Lambda) = k+1$ instead of $\mathcal{L}(\Lambda) \geq k+1$ in (3.13). However this makes no difference, as any convex integral path Λ with $\mathcal{L}(\Lambda) > k+1$ can be "shrunk" to a convex integral path with $\mathcal{L}(\Lambda) = k+1$ without increasing Ω-length; see the proof of Lemma 3.11 below.

appropriately smoothed, we find that there is a constant C, depending only on Ω and not on the positive integer k, such that

$$\mathrm{Area}(r\Omega) \le k - \frac{r}{2}(a(\Omega) + b(\Omega)) + Cr^{2/3}.$$

In particular, since $\mathrm{Area}(r\Omega) = r^2 \,\mathrm{vol}(X_\Omega)$, we get

$$r = \sqrt{\frac{k}{\mathrm{vol}(X_\Omega)}} + o(\sqrt{k}).$$

Putting this into the previous inequality, we get

$$\mathrm{Area}(r\Omega) \le k - \left(\frac{a(\Omega) + b(\Omega)}{2\sqrt{\mathrm{vol}(X_\Omega)}}\right)\sqrt{k} + o(\sqrt{k}). \tag{3.17}$$

Combining (3.14), (3.15), (3.16), and (3.17), we obtain

$$\begin{aligned} c_k(X_\Omega) &\le 2\sqrt{\mathrm{vol}(X_\Omega)\left(k - \left(\frac{a(\Omega) + b(\Omega)}{2\sqrt{\mathrm{vol}(X_\Omega)}}\right)\sqrt{k} + o(\sqrt{k})\right)} \\ &= 2\sqrt{\mathrm{vol}(X_\Omega)k} - \frac{a(\Omega) + b(\Omega)}{2} + o(1). \end{aligned}$$

By Eq. (1.7), the lemma follows. □

We also have a "dual" version of Lemma 3.11 for concave toric domains.

Lemma 3.14. *Let X_Ω be a concave toric domain (see Definition 3.1) such that $\partial_+\Omega$ is the graph of a strictly convex C^2 function.*[8] *Then*

$$\liminf_{k\to\infty} e_k(X_\Omega) \ge -\frac{a(\Omega) + b(\Omega)}{2}.$$

Proof. This is proved similarly to Lemma 3.11, but with inequalities going in the reverse direction.

To start, there is a counterpart of Theorem 3.13, proved in [7, Thm. 1.21], which reads

$$c_k(X_\Omega) = \max\{\ell_\Omega(\Lambda) \mid \mathcal{L}(\Lambda) \le k\}.$$

Here Λ is a *concave integral path*, which is a polygonal path with vertices at lattice points from $(0, b)$ to $(a, 0)$ with $a, b \ge 0$ which is the graph of a convex function. In this context the Ω-length $\ell_\Omega(\Lambda)$ is defined as in (3.11), but with the norm $\|\cdot\|_\Omega^*$ replaced by the "anti-norm" given by

$$[v] = \min\{\langle v, w\rangle | w \in \partial_+\Omega\}.$$

Finally, $\mathcal{L}(\Lambda)$ now denotes the number of lattice points in the region enclosed by Λ and the axes, this time not including lattice points on Λ.

Given a positive integer k, let r be the supremum of the set of real numbers such that the scaling $r\Omega$ contains at most k lattice points. The boundary of the convex hull of the set of lattice points in the nonnegative

[8]Again, this is a bit more general than a "strictly concave toric domain".

quadrant but not in $(r-\epsilon)\Omega$ then consists of rays along the axes, together with a concave integral path Λ satisfying $\mathcal{L}(\Lambda) \le k$. Thus

$$c_k(X_\Omega) \ge \ell_\Omega(\Lambda).$$

The rest of the proof now parallels the proof of Lemma 3.11. □

3.4. Completing the proof of the main theorem

Proof of Theorem 1.10. Let X_Ω be a strictly convex or strictly concave toric domain. By Proposition 1.11, what we need to show is that

$$\lim_{k\to\infty} e_k(X_\Omega) = -\frac{a(\Omega)+b(\Omega)}{2}. \tag{3.18}$$

In the strictly concave case, this follows from Corollary 3.9 and Lemma 3.14.

In the strictly convex case, by Lemma 3.11, we just need to show that

$$\liminf_{k\to\infty} e_k(X_\Omega) \ge -\frac{a(\Omega)+b(\Omega)}{2}. \tag{3.19}$$

To do so, recall the notation $\Delta(c)$ from Sect. 3.1, and let c be the smallest positive real number such that $\Omega \subset \Delta(c)$. Then $\partial_+\Omega$ intersects $\partial_+\Delta(c)$ in a unique point $(t, c-t)$. Suppose that $0<t<c$. (The cases where $t=0$ or $t=c$ are simpler and will be omitted.)

Let Ω' denote the closure of the component of $\Delta(c)\backslash\Omega$ with $\mu_1 < t$, and let Ω'' denote the closure of the component of $\Delta(c)\backslash\Omega$ with $\mu_1 > t$. Define integral affine transformations $\phi', \phi'' : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$\begin{aligned}\phi'(\mu_1,\mu_2) &= (c-\mu_1-\mu_2, \mu_1),\\ \phi''(\mu_1,\mu_2) &= (\mu_2, c-\mu_1-\mu_2).\end{aligned}$$

Then $X' = X_{\phi'(\Omega')}$ and $X'' = X_{\phi''(\Omega'')}$ are concave toric domains satisfying the hypotheses of Corollary 3.9 and Lemma 3.14, so that they satisfy (3.18). Observe also that

$$\begin{aligned}a(\phi'(\Omega')) &= c-b(\Omega),\\ b(\phi'(\Omega')) &= t,\\ a(\phi''(\Omega'')) &= c-t,\\ b(\phi''(\Omega'')) &= c-a(\Omega).\end{aligned}$$

By [8, Thm. A.1], we have

$$c_k(X_\Omega) = \inf_{k',k''\ge 0}\left(c_{k+k'+k''}(B(c)) - c_{k'}(X') - c_{k''}(X'')\right). \tag{3.20}$$

By (3.18) for X' and X'' we get, as functions of k' and k'',

$$\begin{aligned}c_{k'}(X') &= 2\sqrt{k'\cdot \mathrm{vol}(X')} + \frac{b(\Omega)-c-t}{2} + o(1),\\ c_{k''}(X'') &= 2\sqrt{k''\cdot \mathrm{vol}(X'')} + \frac{a(\Omega)-2c+t}{2} + o(1).\end{aligned} \tag{3.21}$$

By (1.8), we have

$$c_{k+k'+k''}(B(c)) \ge 2\sqrt{(k+k'+k'')\,\mathrm{vol}(B(c))} - \frac{3c}{2} + o(1). \tag{3.22}$$

Now since $\mathrm{vol}(B(c)) = \mathrm{vol}(X_\Omega) + \mathrm{vol}(X') + \mathrm{vol}(X'')$, by the Cauchy-Schwarz inequality (for three-component vectors) we have

$$\sqrt{(k + k' + k'')\,\mathrm{vol}(B(c))} \geq \sqrt{k\,\mathrm{vol}(X_\Omega)} + \sqrt{k'\,\mathrm{vol}(X')} + \sqrt{k''\,\mathrm{vol}(X'')}. \tag{3.23}$$

Combining (3.20), (3.21), (3.22), and (3.23), we obtain

$$\begin{aligned} e_k(X_\Omega) &\geq \frac{-3c}{2} + \frac{-b(\Omega) + c + t}{2} + \frac{-a(\Omega) + 2c - t}{2} + o(1) \\ &= -\frac{a(\Omega) + b(\Omega)}{2} + o(1). \end{aligned}$$

(Note that while the $o(1)$ terms in (3.21) are as functions of k' and k'', we do get $o(1)$ terms as functions of k above, since when k is large, we must also have k' and k'' large when close to the infimum in (3.20), as in the proof of Lemma 3.8.) This proves (3.19) for our strictly convex toric domain X_Ω and thus completes the proof of the theorem. □

4. Improving the exponent in the general case

In this section we prove Theorem 1.1, estimating $e_k(X)$ for a general compact domain $X \subset \mathbb{R}^4$ with smooth boundary.

To prepare for this, if $a, b > 0$, define the *polydisk*

$$P(a, b) = \left\{ z \in \mathbb{C}^2 \;\middle|\; \pi|z_1|^2 \leq a^2,\ \pi|z_2|^2 \leq b^2 \right\}.$$

It was shown in [19] (and also follows directly from the more general Theorem 3.13) that the ECH capacities of a polydisk are given by

$$c_k(P(a, b)) = \min\left\{ am + bm \;\middle|\; (m + 1)(n + 1) \geq k + 1 \right\} \tag{4.1}$$

where m, n are nonnegative integers. We now need two simple estimates.

Lemma 4.1. *$e_k(P(a, a)) \geq -2a$ for all k.*

Proof. For each nonnegative integer k, there is a unique nonnegative integer d such that

$$d^2 \leq k \leq d^2 + 2d.$$

It follows from (4.1) that

$$c_k(P(a, a)) = \begin{cases} (2d - 1)a, & d^2 \leq k \leq d^2 + d, \\ 2da, & d^2 + d < k \leq d^2 + 2d. \end{cases} \tag{4.2}$$

On the other hand, $\mathrm{vol}(P(a, a)) = a^2$, so

$$e_k(P(a, a)) = c_k(P(a, a)) - 2a\sqrt{k}. \tag{4.3}$$

In the first line of (4.2) we have $\sqrt{k} < d + 1/2$, and in the second line of (4.2) we have $\sqrt{k} < d + 1$. The lemma then follows from (4.2) and (4.3). □

Lemma 4.2. *Let X be a bounded domain in $\mathbb{R}^4$, and suppose there are disjoint open subsets $P_1, P_2, \ldots \subset X$ such that P_i is symplectomorphic to* $\operatorname{int}(P(a_i, a_i))$. *Let k be a positive integer. Let*

$$I_k = \left\{ i \mid a_i^2 \geq \operatorname{vol}(X)/k \right\}$$

and write

$$V_k = \sum_{i \in I_k} a_i^2 = \operatorname{vol}\left(\bigcup_{i \in I_k} P_i \right).$$

Then

$$e_k(X) \geq -2\sqrt{2} \sum_{i \in I_k} a_i + 2 \frac{(V_k - \operatorname{vol}(X))}{\sqrt{\operatorname{vol}(X)}} \sqrt{k}. \tag{4.4}$$

Proof. For each i define a positive real number

$$\widehat{k_i} = \frac{a_i^2}{\operatorname{vol}(X)} k,$$

and define a nonnegative integer

$$k_i = \left\lfloor \widehat{k_i} \right\rfloor.$$

Note that $k_i > 0$ if and only if $i \in I_k$.

By the disjoint union property of ECH capacities (1.4) and the definition of the error term (1.7), we have

$$\begin{aligned} c_k(X) &\geq \sum_i c_{k_i}(P(a_i, a_i)) \\ &= \sum_{i \in I_k} \left(2a_i \sqrt{k_i} + e_k(P(a_i, a_i)) \right) \\ &= 2 \sum_{i \in I_k} a_i \sqrt{\widehat{k_i}} + \sum_{i \in I_k} \left(2a_i \left(\sqrt{k_i} - \sqrt{\widehat{k_i}} \right) + e_k(P(a_i, a_i)) \right). \end{aligned}$$

By the definition of $\widehat{k_i}$, we have

$$\sum_{i \in I_k} a_i \sqrt{\widehat{k_i}} = \frac{V_k}{\sqrt{\operatorname{vol}(X)}} \sqrt{k}.$$

And for each $i \in I_k$, by Lemma 4.1 and the fact that $k_i \geq 1$, we have

$$2a_i \left(\sqrt{k_i} - \sqrt{\widehat{k_i}} \right) + e_k(P(a_i, a_i) \geq -2\sqrt{2} a_i.$$

Combining the above three lines gives

$$c_k(X) \geq -2\sqrt{2} \sum_{i \in I_k} a_i + \frac{2V_k}{\sqrt{\operatorname{vol}(X)}} \sqrt{k}.$$

The lemma now follows from the definition of the error term (1.7). □

Proof of Theorem 1.1. We first prove the inequality

$$e_k(X) \geq -Ck^{1/4}. \tag{4.5}$$

Here and below, C denotes a positive constant which depends only on X, but which may change from one line to the next.

To do so, we inductively define a sequence $P_1, P_2, \ldots$ as in (4.4) as follows. Step 1 is to add all open cubes whose vertices are consecutive points on the half-integer lattice $\frac{1}{2}\mathbb{Z}^4$ that are contained in X. For $n > 1$, Step n is to add all open cubes whose vertices are consecutive points in the scaled lattice $2^{-n}\mathbb{Z}^4$ that are contained in X but not contained in any of the cubes added in the first $n-1$ steps. Each cube added in Step n is symplectomorphic to the open polydisk $\operatorname{int}(P(4^{-n}, 4^{-n}))$.

Let X_n denote the closure of the union of all cubes added in Steps 1 to n. Then we have

$$\operatorname{vol}(X \setminus X_n) \leq C \cdot 2^{-n}. \tag{4.6}$$

The reason is that by construction, any point in $X\setminus X_n$ is within distance 2^{1-n} of ∂X. And since ∂X is assumed smooth, it follows that the volume of the set of points within distance d of ∂X is at most $C \cdot d$ when d is small.

Let m_n denote the number of cubes obtained in Step n. Since these cubes are disjoint and each have volume 16^{-n}, it follows from (4.6) that

$$m_n \leq C \cdot 8^n. \tag{4.7}$$

Now suppose that

$$16^n \leq \frac{k}{\operatorname{vol}(X)} < 16^{n+1}. \tag{4.8}$$

Then in the notation of Lemma 4.2, the set I_k consists of the indices of the cubes added in the first n steps. By (4.7), we have

$$\sum_{i \in I_k} a_i \leq C \cdot 2^n.$$

And by (4.6), we have

$$\frac{V_k - \operatorname{vol}(X)}{\sqrt{\operatorname{vol}(X)}} \geq -C \cdot 2^{-n}.$$

Putting the above three lines into (4.4) gives

$$e_k(X) \geq -C \cdot 2^n.$$

By (4.8), we obtain (4.5).

To complete the proof of the theorem, we need to prove the reverse inequality

$$e_k(X) \leq C \cdot k^{1/4}.$$

To do so, we choose a large cube W containing X, divide the complement $W\setminus X$ into cubes as above, and use a similar argument. (Compare [19, Prop. 8.6].) $\square$

5. Heuristics for the conjecture

We now review some facts from embedded contact homology, and then use these to give a heuristic discussion of why we expect Conjecture 1.5 to be true.

5.1. Facts

We first briefly review some notions from embedded contact homology. Let Y be a homology 3-sphere, and let λ be a nondegenerate contact form on Y.

Definition 5.1. An *ECH generator* is a finite set of pairs $\alpha = \{(\alpha_i, m_i)\}$ where:

- The α_i are distinct simple Reeb orbits.
- The m_i are positive integers.
- If α_i is hyperbolic (meaning that the linearized return map of the Reeb flow along α_i has real eigenvalues) then $m_i = 1$.

Define the *symplectic action* of α to be the real number

$$\mathcal{A}(\alpha) = \sum_i m_i \mathcal{A}(\alpha_i).$$

Here $\mathcal{A}(\alpha_i)$ denotes the symplectic action, or period, of the Reeb orbit α_i.

Let τ be a trivialization of the contact structure ξ; this trivialization exists and is unique up to homotopy by our assumption that Y is a homology sphere. If γ is a Reeb orbit, define its rotation number

$$\theta(\gamma) = \operatorname{rot}_\tau(y, \mathcal{A}(\gamma)) = \mathcal{A}(\gamma)\rho(y).$$

where y is a point on the image of γ.

Definition 5.2. If $\alpha = \{(\alpha_i, m_i)\}$ is an ECH generator, define[9] its *ECH index* to be the integer

$$I(\alpha) = \sum_i m_i^2 \operatorname{sl}(\alpha_i) + \sum_{i \neq j} m_i m_j \ell(\alpha_i, \alpha_j) + \sum_i \sum_{k=1}^{m_i} \left(\lfloor k\theta(\alpha_i) \rfloor + \lceil k\theta(\alpha_i) \rceil \right). \tag{5.1}$$

Here $\ell(\alpha_i, \alpha_j)$ denotes the linking number of α_i and α_j; and $\operatorname{sl}(\alpha_i)$ denotes the self-linking number of the transverse knot α_i, which is the linking number of α_i with a pushoff in the direction τ, see [15, §3.5.2].

If (Y, ξ) is diffeomorphic to S^3 with the tight contact structure, then one can define the *ECH spectrum* of (Y, λ), which is a sequence of real numbers $c_k(Y, \lambda)$ indexed by nonnegative integers k. The relevance for our discussion is that if X is a nice star-shaped domain in $\mathbb{R}^4$, then its ECH capacities are defined by

$$c_k(X) = c_k(\partial X, \lambda_0|_{\partial X}).$$

[9] This is a special case of the general definition of the ECH index in [20, Def. 3.5]. The relative first Chern class term there is not present here because we are using a global trivialization τ.

And the key fact we need to know is that

$$c_k(Y,\lambda) = \mathcal{A}(\alpha), \tag{5.2}$$

where α is a certain ECH generator with ECH index

$$I(\alpha) = 2k,$$

selected by a "min-max" procedure using the ECH chain complex.

We now want to look at the index formula (5.1) more closely. To prepare for this we need a bit more background. Choose an auxiliary metric on Y. If $y \in Y$ and $T > 0$, we can form a loop $\eta_{y,T}$ by starting with the path given by the time t Reeb flow from y to $\phi_T(y)$, and then appending a length-minimizing geodesic from $\phi_T(y)$ back to y. (If this geodesic is not unique, pick one arbitrarily.) If y_1, y_2 are distinct, define the *asymptotic linking number* by

$$f(y_1,y_2) = \lim_{T_1,T_2\to\infty} \frac{1}{T_1T_2}\ell(\eta_{y_1,T_1},\eta_{y_2,T_2}),$$

when this limit exists. Here of course $\ell(\eta_{y_1,T_1},\eta_{y_2,T_2})$ is defined only when the loops η_{y_1,T_1} and η_{y_2,T_2} are disjoint. By a result of Arnold [3] and Vogel [30] (which applies to more general volume-preserving vector fields), the function f is defined almost everywhere on $Y \times Y$ and integrable, and

$$\int_{Y\times Y} f = \mathrm{vol}(Y,\lambda). \tag{5.3}$$

Here we are integrating with respect to the measure on $Y \times Y$ given by the product of the contact volume forms $\lambda \wedge \mathrm{d}\lambda$, and we define $\mathrm{vol}(Y,\lambda) = \int_Y \lambda \wedge \mathrm{d}\lambda$.

For example, if y_1 and y_2 are on distinct simple Reeb orbits γ_1 and γ_2, then it follows from the definition that

$$f(y_1,y_2) = \frac{1}{\mathcal{A}(\gamma_1)\mathcal{A}(\gamma_2)}\ell(\gamma_1,\gamma_2).$$

If y_1 and y_2 are on the same simple Reeb orbit γ, then $f(y_1,y_2)$ is not defined; however it is natural to extend the definition in this case to set

$$f(y_1,y_2) = \frac{1}{\mathcal{A}(\gamma)^2}\left(\mathrm{sl}(\gamma) + \theta(\gamma)\right).$$

Using the above formulas, we can rewrite the index formula (5.1) as

$$I(\alpha) = \sum_{i,j} m_i m_j \mathcal{A}_i\mathcal{A}_j f_{i,j} - \sum_i m_i^2 \mathcal{A}_i\rho_i + \sum_i \sum_{k=1}^{m_i} \left(\lfloor k\mathcal{A}_i\rho_i\rfloor + \lceil k\mathcal{A}_i\rho_i\rceil\right). \tag{5.4}$$

Here we write $\mathcal{A}_i = \mathcal{A}(\alpha_i)$; we let $f_{i,j}$ denote $f(y_i,y_j)$ for y_i in the image of α_i and y_j in the image of α_j; and ρ_i denotes $\rho(y)$ for y in the image of α_i.

5.2. A new definition

Definition 5.3. If $\alpha = \{(\alpha_i, m_i)\}$ is an ECH generator, then using the notation of (5.4), define its *approximate ECH index* to be the real number

$$I_{\mathrm{approx}}(\alpha) = \sum_{i,j} m_i m_j \mathcal{A}_i \mathcal{A}_j f_{i,j} + \sum_i m_i \mathcal{A}_i \rho_i. \tag{5.5}$$

We can bound the error in this approximation as follows:

Lemma 5.4. $|I_{\mathrm{aprox}}(\alpha) - I(\alpha)| \le \sum_i m_i$.

Proof. It follows from (5.4) and (5.5) that

$$I_{\mathrm{approx}}(\alpha) - I(\alpha) = \sum_i \sum_{k=1}^{m_i} \left(2k\mathcal{A}_i\rho_i - \lfloor k\mathcal{A}_i\rho_i \rfloor - \lceil k\mathcal{A}_i\rho_i \rceil\right).$$

The lemma then follows since

$$\left| 2x - \lfloor x \rfloor - \lceil x \rceil \right| < 1$$

for every real number x. □

We can now suggestively rewrite (5.5) as

$$I_{\mathrm{approx}}(\alpha) = \int_{\alpha\times\alpha} f + \int_\alpha \rho \tag{5.6}$$

where the integral is with respect to the measure given by the Reeb vector field, multiplied by m_i on each orbit α_i.

5.3. Heuristics

A conjecture of Irie [24], of which a version has been verified for convex and concave toric domains, asserts that if λ is generic, then ECH generators α realizing $c_k(Y,\lambda)$ as in (5.2) are *equidistributed* in Y as $k \to \infty$. This means that if $U \subset Y$ is an open set, then the symplectic action of $\alpha \cap U$ divided by the symplectic action of α converges to $\mathrm{vol}(U)/\mathrm{vol}(Y)$. If we assume a very favorable version of this equidistribution, then by Lemma 5.4 and Eq. (5.6) we can approximate

$$2k = I(\alpha) \approx I_{\mathrm{approx}}(\alpha) \approx \frac{\mathcal{A}(\alpha)^2}{\mathrm{vol}(Y,\lambda)^2}\int_{Y\times Y} f + \frac{\mathcal{A}(\alpha)}{\mathrm{vol}(Y,\lambda)}\int_Y \rho.$$

Here we are not discussing the size of the error in the approximation since this is just a heuristic. Comparing with (1.11) and (5.3), we obtain

$$2k\cdot \mathrm{vol}(Y,\lambda) \approx \mathcal{A}(\alpha)^2 + \mathcal{A}(\alpha)\,\mathrm{Ru}(Y,\lambda).$$

Since $\mathcal{A}(\alpha) = c_k(Y,\lambda)$, we then get

$$c_k(Y,\lambda) \approx \sqrt{2k\cdot\mathrm{vol}(Y,\lambda)} - \frac{1}{2}\mathrm{Ru}(Y,\lambda).$$

When X is a nice star-shaped domain, we have $\mathrm{vol}(\partial X, \lambda_0|_{\partial X}) = 2\,\mathrm{vol}(X)$ by Stokes's theorem, so we obtain

$$c_k(X) \approx 2\sqrt{k\cdot\mathrm{vol}(X)} - \frac{1}{2}\mathrm{Ru}(X).$$

Acknowledgements

Thanks to Alberto Abbondandolo, Julian Chaidez, and Umberto Hryniewicz for explaining the Ruelle invariant; to Dusa McDuff for explaining Lemma 3.6; and to Dan Cristofaro-Gardiner for discussions about the asymptotics of ECH capacities. Partially supported by NSF grant DMS-1708899 and a Humboldt Research Award.

References

[1] Abbondandolo, A., Bramham, B., Hryniewicz, U., Salomão, P.: Sharp systolic inequalities for Reeb flows on the three-sphere. Invent. Math. **211**, 687–778 (2018)

[2] Abbondandolo, A., Bramham, B., Hryniewicz, U., Salomão, P.: Systolic ratio, index of closed orbits and convexity for tight contact forms on the three-sphere. Compos. Math. **154**, 2643–2680 (2018)

[3] Arnold, V.: The asymptotic Hopf invariant and its applications. Sel. Math. Sov. **5**, 327–345 (1986)

[4] Asaoka, M., Irie, K.: A C^∞ closing lemma for Hamiltonian diffeomorphism of closed surfaces. GAFA **26**, 1245–1254 (2016)

[5] Brothers, J., Morgan, F.: The isoperimetric theorem for general integrands. Mich. Math. J. **41**, 419–431 (1994)

[6] Chaix, H.: Démonstration élémentaire d'un théorème de Van der Corput. C.R. Acad. Sc. Paris **275**, 883–885 (1972)

[7] Choi, K., Cristofaro-Gardiner, D., Frenkel, D., Hutchings, M., Ramos, V.: Symplectic embeddings into four-dimensional concave toric domains. J. Topol. **7**, 1054–1076 (2014)

[8] Cristofaro-Gardiner,D.: Symplectic embeddings of concave toric domains into convex ones. J. Differ. Geom. **112**, 199–232 (2019) **(appendix with K. Choi)**

[9] Cristofaro-Gardiner, D., Hutchings, M.: From one Reeb orbit two two. J. Differ. Geom. **102**, 25–36 (2016)

[10] Cristofaro-Gardiner, D., Hutchings, M., Pomerleano, D.: Torsion contact forms in three dimensions have two or infinitely many Reeb orbits. Geom. Topol. **23**, 3601–4645 (2019)

[11] Cristofaro-Gardiner, D., Hutchings, M., Ramos, V.: The asymptotics of ECH capacities. Invent. Math. **199**, 187–214 (2015)

[12] Cristofaro-Gardiner, D., Li, T., Stanley, R.: Irrational triangles with polynomial Ehrhart functions. Discrete Comput. Geom. **61**, 227–246 (2019)

[13] Cristofaro-Gardiner, D., Savale, N.: Sub-leading asymptotics of ECH capacities. Sel. Math. New Ser. **26**, 65 (2020)

[14] Gambaudo, J., Ghys, E.: Enlacements asymptotiques. Topology **36**, 1355–1379 (1997)

[15] Geiges, v: An Introduction to Contact Topology, Cambridge Studies in Advanced Mathematics, vol. 109. Cambridge University Press, Cambridge (2008)

[16] Gutt, J., Hutchings, M., Ramos, V.: Examples around the strong Viterbo conjecture. J. Fixed Point Theory Appl. arXiv:2003.10854 **(to appear)**

[17] Gutt, J., Zhang, J.: Private communication (2021)

[18] Humilière, V.: Un lemme de fermeture C^∞ [d'après Asaoka et Irie], Séminaire Bourbaki, nov 2019. Astérisque **430** (2021)

[19] Hutchings, M.: Quantitative embedded contact homology. J. Differ. Geom. **88**, 231–266 (2011)

[20] Hutchings, M.: Lecture notes on embedded contact homology, Contact and symplectic topology, Bolyai Soc. Math. Stud., vol. 26, pp. 389–484. Springer (2014)

[21] Hutchings, M.: Beyond ECH capacities. Geom. Topol. **20**, 1085–1126 (2016)

[22] Hutchings, M.: Mean action and the Calabi invariant. J. Mod. Dyn. **10**, 511–539 (2016)

[23] Irie, K.: Dense existence of periodic orbits and ECH spectral invariants. J. Mod. Dyn. **9**, 357–363 (2015)

[24] Irie, K.: Equidistributed periodic orbits of C^∞-generic three-dimensional Reeb flows. J. Sympl. Geom. **19**, 531–566 (2021)

[25] McDuff, D.: The Hofer conjecture on embedding symplectic ellipsoids. J. Differ. Geom. **88**, 519–532 (2011)

[26] McDuff, D.: Private communication (2018)

[27] McDuff, D., Schlenk, F.: The embedding capacity of 4-dimensional symplectic ellipsoids. Ann. Math. **175**, 1191–1282 (2012)

[28] Ruelle, D.: Rotation numbers for diffeomorphisms and flows. Ann. Inst. H. Poincaré Phyhs. Théor. **42**, 109–115 (1985)

[29] Sun, W.: An estimate on energy of min-max Seiberg–Witten Floer generators. Math. Res. Lett. **26**, 1807–1827 (2019)

[30] Vogel, T.: On the asymptotic linking number. Proc. Am. Math. Soc. **131**, 2289–2297 (2003)

[31] Weiler, M.: Mean action of periodic orbits of area-preserving annulus diffeomorphisms. J. Topol. Anal. **13**, 1013–1074 (2021)

[32] Wormleighton, B.: ECH capacities, Ehrhart theory, and toric varieties. J. Symplectic Geom. **19**, 465–506 (2021)

[33] Wormleighton, B.: Towers of Looijenga pairs and asymptotics of ECH capacities. arXiv:2101.06153

[34] Wulff, G.: Zur Frage der Geschwindigkeit des Wachstums und der Auflösung der Krystallflächen. Zeitshrift für Krystallographie and Mineralogie **34**, 449–530 (1901)

Michael Hutchings
Department of Mathematics
University of California, Berkeley
Berkeley CA 94720-3840
USA
e-mail: hutching@math.berkeley.edu

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:51
https://doi.org/10.1007/s11784-022-00952-x
Published online June 13, 2022

Journal of Fixed Point Theory and Applications

Capacities of billiard tables and S^1-equivariant loop space homology

Kei Irie

Abstract. We introduce a sequence of "capacities" of Riemannian manifolds (with corners). These capacities are defined as min–max values associated with a variational problem concerning periodic billiard trajectories. After establishing basic properties of these capacities, we discuss a conjectural relation between our notion of capacities and symplectic capacities defined from S^1-equivariant symplectic homology. Then we compute capacities of rectangles, and check that the result is consistent with known results (by Gutt–Hutchings and Ramos–Sepe) in symplectic geometry.

Mathematics Subject Classification. 70H12, 37C83, 53D35.

1. Introduction

In [11], the author introduced the notion of "capacity" of compact Riemannian manifolds with boundaries. This capacity is defined as a min–max value associated with a variational problem concerning periodic billiard trajectories. The author conjectures that the capacity of any compact Riemannian manifold (with boundary) is equal to symplectic capacity (which is defined from symplectic homology) of its unit disk cotangent bundle, and confirmed this conjecture for domains in Euclidean spaces ([10]).

The aim of this paper is to introduce a sequence of capacities of Riemannian manifolds (with corners), using a natural S^1-symmetry of free loop spaces. Our definition of these capacities is inspired by the definition of (equivariant) Ekeland–Hofer capacities ([4]) and their analogue in Floer theory ([16], [9]).

This paper is organized as follows. In Sect. 2, we define capacities of Riemannian manifolds with corners. In Sect. 3, we establish basic properties of these capacities. In particular, we show that for any compact Riemannian manifold M with $\partial M \neq \emptyset$, each capacity of M is equal to the length

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

of a periodic billiard trajectory on M. In Sect. 4, we discuss a conjectural relation between our notion of capacities and symplectic capacities defined from equivariant symplectic homology. In Sect. 5, we compute capacities of rectangles, and check that the result is consistent with known results ([9], [14]) in symplectic geometry. In Sect. 6, we briefly discuss a few questions.

Convention. In this paper, we consider only (co)homology with coefficients in $\mathbb{Q}$, unless otherwise specified.

2. Definition of capacities

The aim of this section is to define capacities c and $(c_k^{S^1})_{k\geq 0}$.
We first recall the definition of S^1-equivariant homology. Let

$$S^1 \to ES^1 \to BS^1 \cong \mathbb{C}P^\infty$$

be the universal S^1-bundle. For any topological pair (X, Y) which admits a continuous S^1-action, we define

$$H_*^{S^1}(X,Y) := H_*((X,Y) \times_{S^1} ES^1).$$

If the S^1-action is trivial, there exists a canonical isomorphism

$$H_*^{S^1}(X,Y) \cong H_*(X,Y) \otimes H_*(\mathbb{C}P^\infty).$$

Let us define $e \in H^2(\mathbb{C}P^\infty)$ by $\langle e, [\mathbb{C}P^1]\rangle = -1$. For any (X, Y) as above, there exists a canonical long exact sequence

$$\cdots \longrightarrow H_*(X,Y) \longrightarrow H_*^{S^1}(X,Y) \xrightarrow[-\cap e]{} H_{*-2}^{S^1}(X,Y) \longrightarrow H_{*-1}(X,Y) \longrightarrow \cdots . \tag{1}$$

Remark 2.1. By (1), one can easily check the following: for any integer $k \geq 0$, if $H_k^{S^1}(X,Y) \neq 0$ then $H_{\leq k}(X,Y) := \bigoplus_{0\leq i\leq k} H_i(X,Y) \neq 0$.

Let M be an oriented n-dimensional Riemannian manifold such that $\partial M = \emptyset$. For any $p \in M$, $H_n(M, M \setminus \{p\}) \cong \mathbb{Q}$ has a canonical generator, which we denote by μ_p. We define $\hat{H}_*(M) := \varprojlim_{K\subset M} H_*(M, M \setminus K)$ where K runs over all compact subsets of M. There exists a unique element $\mu_M \in \hat{H}_n(M)$ such that the natural map $\hat{H}_n(M) \to H_n(M, M \setminus \{p\})$ sends μ_M to μ_p for any $p \in M$. For any compact set $K \subset M$, we define $\mu_{M,K} \in H_n(M, K)$ to be the image of μ_M by the natural map $\hat{H}_n(M) \to H_n(M, M \setminus K)$.

Let ΩM denote the space of $L^{1,2}$-free loops on M. Namely, ΩM consists of absolutely continuous maps $\gamma : S^1 \to M$ such that γ' is square-integrable. ΩM is equipped with the natural $L^{1,2}$-topology. Moreover, it has a structure of a smooth Hilbert manifold. ΩM admits a natural S^1-action, which is defined as follows:

$$S^1 \times \Omega M \to \Omega M;\ (t,\gamma) \mapsto \gamma^t, \qquad \gamma^t(\theta) := \gamma(\theta - t).$$

Let us introduce some notations.

- The energy functional $\mathbb{E} : \Omega M \to \mathbb{R}$ is defined by $\mathbb{E}(\gamma) := \int_{S^1} \frac{|\gamma'|^2}{2} (\forall \gamma \in \Omega M)$.
- For any $a > 0$, let $\Omega^a M := \{\gamma \in \Omega M \mid \mathbb{E}(\gamma) < a\}$.

- The length functional len : $\Omega M \to \mathbb{R}$ is defined by $\mathrm{len}(\gamma) := \int_{S^1} |\gamma'| (\forall \gamma \in \Omega M)$.
- For any compact set $K \subset M$, let $\Omega_K M := \{\gamma \in \Omega M \mid \gamma(S^1) \not\subset K\}$.
- For any $p \in M$, let c_p denote the constant loop at p, i.e. $c_p(\theta) := p \quad (\forall \theta \in S^1)$.

For any $a > 0$ and $K \subset M$, let us consider the map

$$i^a_K : (M, M \setminus K) \to (\Omega^a M \cup \Omega_K M, \Omega_K M); \quad p \mapsto c_p.$$

Note that if $H_*(i^a_K)(\mu_{M,K}) = 0$ then $H_*(i^{a'}_K)(\mu_{M,K}) = 0$ for any $a' \geq a$, since $i^{a'}_K$ factors i^a_K. Now let us define

$$c(M, K) := \inf\{\sqrt{2a} \mid H_*(i^a_K)(\mu_{M,K}) = 0\}.$$

Let us consider the trivial S^1-action on M. Then the map i^a_K is S^1-equivariant, and there exists an isomorphism $H^{S^1}_*(M, M \setminus K) \cong H_*(M, M \setminus K) \otimes H_*(\mathbb{C}P^\infty)$. For any $k \in \mathbb{Z}_{\geq 0}$, let us define

$$c^{S^1}_k(M, K) := \begin{cases} \inf\{\sqrt{2a} \mid H^{S^1}_*(i^a_K)(\mu_{M,K} \otimes [\mathbb{C}P^{k-1}]) = 0\} & (k \geq 1), \\ 0 & (k = 0). \end{cases}$$

Remark 2.2. For any a and K as above, let

$$\Omega^a_K M := \Omega^a M \cap \Omega_K M, \qquad j^a_K : (M, M \setminus K) \to (\Omega^a_M, \Omega^a_K M); \, p \mapsto c_p.$$

The inclusion map $(\Omega^a M, \Omega^a_K M) \to (\Omega^a M \cup \Omega_K M, \Omega_K M)$ induces an isomorphism on (S^1-equivariant) homology since $\{\Omega^a M, \Omega_K M\}$ is an open covering of $\Omega^a M \cup \Omega_K M$. Thus, in the above definitions of $c(M, K)$ and $c^{S^1}_k(M, K)$, one may replace i^a_K with j^a_K.

Let K and K' be compact subsets of M such that $K \subset K'$. Consider the following commutative diagram (the right vertical map is an inclusion):

$$\begin{tikzcd}
(M, M \setminus K') \arrow[r, "i^a_{K'}"] \arrow[d, "\mathrm{id}_M"'] & (\Omega^a M \cup \Omega_{K'} M, \Omega_{K'} M) \arrow[d] \\
(M, M \setminus K) \arrow[r, "i^a_K"'] & (\Omega^a M \cup \Omega_K M, \Omega_K M).
\end{tikzcd}$$

The left vertical map induces a map on homology $H_*(M, M \setminus K') \to H_*(M, M \setminus K)$, which sends $\mu_{M,K'}$ to $\mu_{M,K}$. Hence, we obtain

$$K \subset K' \implies c(M, K) \leq c(M, K'), \quad c^{S^1}_k(M, K) \leq c^{S^1}_k(M, K') \quad (\forall k \geq 0).$$

Let us define capacities c and $(c^{S^1}_k)_{k \geq 0}$ by

$$c(M) := \sup_K c(M, K), \qquad c^{S^1}_k(M) := \sup_K c_k(M, K) \quad (\forall k \geq 0),$$

where K runs over all compact subsets of M.

So far we have assumed that $\partial M = \emptyset$. When M is an oriented Riemannian manifold with corners, we define

$$c(M) := c(\mathrm{int}\, M), \qquad c^{S^1}_k(M) := c^{S^1}_k(\mathrm{int}\, M) \quad (\forall k \geq 0),$$

where $\mathrm{int}\, M$ denotes the interior of M.

Remark 2.3. By 'n-dimensional manifold with corners', we mean a Hausdorff and second-countable topological space which is locally modeled on open sets of $(\mathbb{R}_{\geq 0})^n$, such that all coordinate changes are C^∞.

Now we have finished the definition of capacities c and $(c_k^{S^1})_{k\geq 0}$. Let us conclude this section with the following remarks.

Remark 2.4. Let M be an *orientable* manifold with corners. It is easy to check that $c(M)$ and $c_k^{S^1}(M)\,(k \geq 0)$ do not depend on choice of orientations on M. Thus, one can define these capacities for orientable manifolds.

Remark 2.5. Let M be an orientable manifold, and let $(M_i)_{i\in I}$ be the set of connected components of M. Then, it is easy to check that

$$c(M) = \sup_{i\in I} c(M_i), \qquad c_k^{S^1}(M) = \sup_{i\in I} c_k^{S^1}(M_i) \quad (\forall k \geq 0).$$

Remark 2.6. The capacity c was already introduced in [11]. For any compact and connected Riemannian manifold Q with $\partial Q \neq \emptyset$, the capacity $c(Q)$ defined in this section is equal to $c^\Omega(Q : [Q, \partial Q])$ defined in [11] pp. 243. On the other hand, it seems that the capacities $(c_k^{S^1})_k$ have not appeared in the literature.

Remark 2.7. Although we work with $\mathbb{Q}$-coefficients (see our convention), it is straightforward to define the capacities c and $(c_k^{S^1})_{k\geq 0}$ with an arbitrary coefficient ring. The author is not aware if the capacities depend on the choice of coefficient rings.

3. Properties of capacities

In this section, we state and prove some basic properties of the capacities c and $(c_k^{S^1})_{k\geq 0}$. These properties are analogous to the corresponding properties of symplectic capacities defined by symplectic homology.

3.1. Conformality and monotonicity

In this subsection, we state and prove the conformality (Lemma 3.1) and monotonicity (Proposition 3.3) properties. These properties imply that our capacities reflect "sizes" of Riemannian manifolds.

Lemma 3.1. *Let M be an orientable manifold with corners, equipped with a Riemannian metric g. Then, for any $\alpha \in \mathbb{R}_{>0}$, there holds*

$$c(M, \alpha g) = \alpha \cdot c(M, g), \qquad c_k^{S^1}(M, \alpha g) = \alpha c_k^{S^1}(M, g) \quad (\forall k \geq 0).$$

Proof. Let $\mathbb{E}_g$ (resp. $\mathbb{E}_{\alpha g}$) denote the energy functional with respect to the metric g (resp. αg). Then $\mathbb{E}_{\alpha g}(\gamma) = \alpha^2 \cdot \mathbb{E}_g(\gamma)$ for any $\gamma \in \Omega M$. Then, this lemma follows from the definition of the capacities. This lemma also immediately follows from Proposition 3.3 (i) and (ii). $\square$

To state the monotonicity property, we first need the following definition.

Definition 3.2. Let M_0 and M_1 be oriented and connected manifolds with $\partial M_0 = \partial M_1 = \emptyset$ and $\dim M_0 = \dim M_1$, and let $f : M_0 \to M_1$ be a proper continuous map. Then, for any compact set $K \subset M_1$, $f^{-1}(K) \subset M_0$ is also compact. Thus one can define a natural map $\hat{H}_*(f) : \hat{H}_*(M_0) \to \hat{H}_*(M_1)$, and $d \in \mathbb{Z}$ by $\hat{H}_*(f)(\mu_{M_0}) = d \cdot \mu_{M_1}$. This integer d is called the *degree* of the map f, and denoted by $\deg(f)$.

Now let us state the monotonicity property. For simplicity, we consider only connected manifolds without boundaries.

Proposition 3.3. *Let (M_0, g_0) and (M_1, g_1) be oriented and connected Riemannian manifolds such that $\partial M_0 = \partial M_1 = \emptyset$ and $\dim M_0 = \dim M_1$. Let C be a positive real number.*

(i): *If there exists a proper C^∞-map $f : M_0 \to M_1$ such that $\|df(v)\|_{g_1} \leq C\|v\|_{g_0}$ $(\forall v \in TM_0)$ and $\deg(f) \neq 0$, then*

$$c(M_1, g_1) \leq C \cdot c(M_0, g_0), \qquad c_k^{S^1}(M_1, g_1) \leq C \cdot c_k^{S^1}(M_0, g_0) \quad (\forall k \geq 0).$$

(ii): *If there exists a C^∞-embedding $f : M_0 \to M_1$ such that $\|df(v)\|_{g_1} \geq C\|v\|_{g_0}$ $(\forall v \in TM_0)$, then*

$$c(M_1, g_1) \geq C \cdot c(M_0, g_0), \qquad c_k^{S^1}(M_1, g_1) \geq C \cdot c_k^{S^1}(M_0, g_0) \quad (\forall k \geq 0).$$

Proof of Proposition 3.3. (i): We only prove that $c(M_1) \leq C \cdot c(M_0)$, since the proof of $c_k^{S^1}(M_1) \leq C \cdot c_k^{S^1}(M_0)$ is similar. Let us define $\Omega f : \Omega M_0 \to \Omega M_1$ by $\Omega f(\gamma) := f \circ \gamma$. By the assumption, $\Omega f(\Omega^a M_0) \subset \Omega^{Ca} M_1$ for any $a > 0$. For any compact set $K_1 \subset M_1$ let $K_0 := f^{-1}(K_1) \subset M_0$, and consider the following commutative diagram:

$$\begin{array}{ccc} H_*(M_0, M_0 \setminus K_0) & \longrightarrow & H_*(\Omega^a M_0 \cup \Omega_{K_0} M_0, \Omega_{K_0} M_0) \\ \Big\downarrow{\scriptstyle H_*(f)} & & \Big\downarrow{\scriptstyle H_*(\Omega f)} \\ H_*(M_1, M_1 \setminus K_1) & \longrightarrow & H_*(\Omega^{Ca} M_1 \cup \Omega_{K_1} M_1, \Omega_{K_1} M_1). \end{array}$$

Since $\mu_{M_1,K_1} = \frac{H_*(f)(\mu_{M_0,K_0})}{\deg(f)}$, we obtain $c(M_1, K_1) \leq C \cdot c(M_0, K_0) \leq C \cdot c(M_0)$. Since this holds for any compact subset $K_1 \subset M_1$, we obtain $c(M_1) \leq C \cdot c(M_0)$.

(ii): We only prove that $C \cdot c(M_0) \leq c(M_1)$. Let $U := f(M_0)$. Then $C \cdot c(M_0) \leq c(U, g_1|_U)$ by (i). Thus it is sufficient to prove $c(U) \leq c(M_1)$. We can prove this inequality by $c(U) = \sup_{K \subset U} c(U, K) = \sup_{K \subset U} c(M, K) \leq c(M)$, where the first equality holds by definition, and the second equality follows from Lemma 3.4 below. $\square$

Lemma 3.4. *Let M be an oriented Riemannian manifold with $\partial M = \emptyset$. Let K be a compact subset of M, and let U be an open neighborhood of K. Then, there holds $c(U, K) = c(M, K)$ and $c_k^{S^1}(U, K) = c_k^{S^1}(M, K)$ $(\forall k \geq 0)$.*

Proof. We only prove that $c(U, K) = c(M, K)$. For any $a > 0$, consider the following commutative diagram, where vertical maps are induced by inclusions:

$$\begin{array}{ccc} H_*(U, U \setminus K) & \longrightarrow & H_*(\Omega^a U \cup \Omega_K U, \Omega_K U) \\ \downarrow & & \downarrow \\ H_*(M, M \setminus K) & \longrightarrow & H_*(\Omega^a M \cup \Omega_K M, \Omega_K M). \end{array}$$

Then vertical maps are isomorphisms, since $\{U, M \setminus K\}$ is an open cover of M, and $\{\Omega^a U \cup \Omega_K U, \Omega_K M\}$ is an open cover of $\Omega^a M \cup \Omega_K M$. Since the left vertical map sends $\mu_{U,K}$ to $\mu_{M,K}$, we obtain $c(U, K) = c(M, K)$. □

We also prove the following lemma as an application of Lemma 3.4, since it will be useful later.

Lemma 3.5. *Let M be a compact Riemannian manifold with corners. Suppose that M is isometrically embedded into M_+, which is an oriented Riemannian manifold such that* $\dim M_+ = \dim M$ *and* $\partial M_+ = \emptyset$*. Then* $c(M) = c(M_+, M)$ *and* $c_k^{S^1}(M) = c_k^{S^1}(M_+, M)$ $(\forall k \geq 0)$.

Proof. We only prove that $c(M) = c(M_+, M)$. By definition, $c(M) = c(\operatorname{int} M) = \sup_K c(\operatorname{int} M, K)$ where K runs over all compact subsets of $\operatorname{int} M$. For any compact subset $K \subset \operatorname{int} M$,

$$c(\operatorname{int} M, K) = c(M_+, K) \leq c(M_+, M),$$

where the equality follows from Lemma 3.4 and the inequality follows from $K \subset M$. Hence we obtain $c(M) \leq c(M_+, M)$.

Let us prove the opposite inequality $c(M_+, M) \leq c(M)$. For any $\varepsilon > 0$, there exist a compact subset $K \subset \operatorname{int} M$ and a diffeomorphism $f : M_+ \to M_+$ such that $M \subset f(K)$ and $\|df(v)\| \leq (1 + \varepsilon)\|v\|$ for any $v \in TM$. Then we obtain

$$\begin{aligned} c(M_+, M) &\leq c(M_+, f(K)) \leq (1 + \varepsilon) c(M_+, K) = (1 + \varepsilon) c(\operatorname{int} M, K) \\ &\leq (1 + \varepsilon) c(M), \end{aligned}$$

where the second inequality follows from (the proof of) Proposition 3.3 (i). Since ε is an arbitrary positive number, we obtain $c(M_+, M) \leq c(M)$. □

3.2. Nontriviality

Let us state and prove the following "nontriviality" property.

Proposition 3.6. *Let M be an oriented Riemannian manifold with corners.*

(i): *There holds $c(M) > 0$ and $c_k^{S^1}(M) > 0$ for any $k \geq 1$.*
(ii): *If M is compact and connected, then $c(M) < \infty$ if and only if $\partial M \neq \emptyset$.*
(iii): *If M is compact, connected and $\partial M \neq \emptyset$, then $c_k^{S^1}(M) < \infty$ for any $k \geq 1$.*

Proof. (i):Let $n := \dim M$, and let

$$B^n(1) := \{x \in \mathbb{R}^n \mid |x| < 1\}, \qquad \bar{B}^n(1) := \{x \in \mathbb{R}^n \mid |x| \leq 1\}.$$

In the following arguments, we use abbreviations $B := B^n(1)$ and $\bar{B} := \bar{B}^n(1)$.

We first prove that

$$c(B) = c(\mathbb{R}^n, \bar{B}) \geq 1/2, \qquad c_k^{S^1}(B) = c_k^{S^1}(\mathbb{R}^n, \bar{B}) \geq 1/2 \quad (\forall k \geq 1). \tag{2}$$

The two equalities follow from Lemma 3.5. By Remark 2.2, to prove the two inequalities it is sufficient to show that the map

$$I : (\mathbb{R}^n, \mathbb{R}^n \setminus \bar{B}) \to (\Omega^a \mathbb{R}^n, \Omega^a_{\bar{B}} \mathbb{R}^n); \quad p \mapsto c_p$$

is a homotopy equivalence if $a < 1/8$.

If $a < 1/8$, then any $\gamma \in \Omega^a \mathbb{R}^n$ satisfies $\mathrm{len}(\gamma) < 1/2$. In particular, any $\gamma \in \Omega^a_{\bar{B}}(\mathbb{R}^n)$ satisfies $|\gamma(0)| > 1/2$. Then one can define a map

$$J : (\Omega^a \mathbb{R}^n, \Omega^a_{\bar{B}} \mathbb{R}^n) \to (\mathbb{R}^n, \mathbb{R}^n \setminus \bar{B}); \quad \gamma \mapsto 2\gamma(0).$$

Let us show that J is a homotopy inverse of I. A homotopy between $J \circ I$ and the identity map on $(\mathbb{R}^n, \mathbb{R}^n \setminus \bar{B})$ is given by

$$H_0 : (\mathbb{R}^n, \mathbb{R}^n \setminus \bar{B}) \times [0, 1] \to (\mathbb{R}^n, \mathbb{R}^n \setminus \bar{B}); \quad (x, t) \mapsto (1 + t)x.$$

To define a homotopy between $I \circ J$ and the identity map on $(\Omega^a \mathbb{R}^n, \Omega^a_{\bar{B}} \mathbb{R}^n)$, notice that if $\gamma \in \Omega^a_{\bar{B}}(\mathbb{R}^n)$ then there exists $\theta \in S^1$ such that $|\gamma(\theta)| > 1$, and there holds

$$\begin{aligned} &|(t + t^2)\gamma(0) + (1 - t^2)\gamma(\theta)| = |(t + t^2)(\gamma(0) - \gamma(\theta)) + (1 + t)\gamma(\theta)| \\ \geq &(1 + t)|\gamma(\theta)| - (t + t^2)|\gamma(0) - \gamma(\theta)| > (1 + t) - (t + t^2)/2 \geq 1. \end{aligned}$$

Then we can define a desired homotopy

$$H_1 : (\Omega^a \mathbb{R}^n, \Omega^a_{\bar{B}} \mathbb{R}^n) \times [0, 1] \to (\Omega^a \mathbb{R}^n, \Omega^a_{\bar{B}} \mathbb{R}^n)$$

by $H_1(\gamma, t)(\theta) := (t + t^2)\gamma(0) + (1 - t^2)\gamma(\theta)$. Thus, we have proved that J is a homotopy inverse of I.

Let us finish the proof of (i). For any n-dimensional Riemannian manifold M, there exist an open embedding $\varphi : B \to M$ and a positive real number $C > 0$ such that $\|d\varphi(v)\| \geq C \cdot \|v\|$ for any $v \in TB$. By Proposition 3.3 (ii), we obtain

$$c(M) \geq C \cdot c(B) > 0, \qquad c_k^{S^1}(M) \geq C \cdot c_k^{S^1}(B) > 0 \quad (\forall k \geq 1).$$

(ii): $\partial M \neq \emptyset \implies c(M) < \infty$ follows from [11] Proposition 5.5. When $\partial M = \emptyset$, then the map on homology $H_*(M) \to H_*(\Omega M)$ which is induced from the map $M \to \Omega M$; $p \mapsto c_p$ is injective. This means that $c(M) = \infty$.

(iii) follows from (ii) and the inequality $c_k^{S^1}(M) \leq c_{k-1}^{S^1}(M) + c(M)$ ($\forall k \geq 1$), which we prove in Sect. 3.4. $\square$

3.3. Periodic billiard trajectory

Let us first clarify our definition of periodic billiard trajectories.

Definition 3.7. Let M be a Riemannian manifold with boundary. A periodic billiard trajectory on M is a nonconstant continuous map $\gamma : S^1 \to M$ such that there exists a finite set $\mathscr{B}_\gamma \subset S^1$ satisfying the following conditions.

- $\nabla_{\gamma'} \gamma' \equiv 0$ on $S^1 \setminus \mathscr{B}_\gamma$.

- $\gamma(t) \in \partial M$ for any $t \in \mathscr{B}_\gamma$.
- For any $t \in \mathscr{B}_\gamma$, the limits $\gamma'_+(t) := \lim_{h\to+0}\gamma'(t+h)$ and $\gamma'_-(t) := \lim_{h\to+0}\gamma'(t-h)$ exist, and satisfy the "law of reflection":
$$\gamma'_+(t)+\gamma'_-(t) \in T_{\gamma(t)}\partial M, \qquad \gamma'_+(t)-\gamma'_-(t) \in (T_{\gamma(t)}\partial M)^\perp \setminus \{0\}.$$

Remark 3.8. We allow the case $\mathscr{B}_\gamma = \emptyset$. In that case, γ is a closed geodesic.

Now we prove the following result. This is analogous to the fact that the (equivariant) symplectic homology capacity of a Liouville domain (see Sect. 4) is equal to the period of a Reeb orbit on its boundary.

Proposition 3.9. *Let M be a compact, connected, and orientable Riemannian manifold with $\partial M \neq \emptyset$ and $\dim M = n$. There exists a periodic billiard trajectory γ on M such that $\mathrm{len}(\gamma) = c(M)$ and $\sharp\mathscr{B}_\gamma \le n+1$. Moreover, for any integer $k \ge 1$, there exists a periodic billiard trajectory γ_k on M such that $\mathrm{len}(\gamma_k) = c_k^{S^1}(M)$ and $\sharp\mathscr{B}_\gamma \le n+2k-1$.*

Proof. The first assertion follows from Lemma 5.4 (ii) of [11], where α in the statement of that lemma is the relative fundamental class of M. The proof of the second assertion is very similar to the proof of the first assertion, nevertheless we explain details in four steps, for the sake of completeness.

Step 1. Let us take an orientable Riemannian manifold M_+ such that $\dim M = \dim M_+$, $\partial M_+ = \emptyset$ and M is isometrically embedded into M_+. By Lemma 3.5, $c_k^{S^1}(M) = c_k^{S^1}(M_+, M)$. In the following, we denote $C := (c_k^{S^1}(M))^2/2$.

Step 2. By the the definition of $c_k^{S^1}$, for any real numbers A_- and A_+ satisfying $A_- < C < A_+$, there holds
$$H^{S^1}_{n+2k-1}(\Omega^{A_+}M_+ \cup \Omega_M M_+, \Omega^{A_-}M_+ \cup \Omega_M M_+) \neq 0.$$
By Remark 2.1,
$$H_{\le n+2k-1}(\Omega^{A_+}M_+ \cup \Omega_M M_+, \Omega^{A_-}M_+ \cup \Omega_M M_+) \neq 0.$$
Step 3. Let d_M denote the natural distance function on the Riemannian manifold M. For any $\delta > 0$, let $M(\delta) := \{x \in M \mid d_M(x, \partial M) \ge \delta\}$. We prove that, for any $\varepsilon > 0$, there exist A_- and A_+ satisfying $A_- \in (C-\varepsilon, C)$, $A_+ \in (C, C+\varepsilon)$ and
$$\varinjlim_{\delta\to 0} H_{\le n+2k-1}(\Omega^{A_+}M \cup \Omega_{M(\delta)}M, \Omega^{A_-}M \cup \Omega_{M(\delta)}M) \neq 0. \tag{3}$$

We first take $A'_- \in (C-\varepsilon, C)$ and $A'_+ \in (C, C+\varepsilon)$ arbitrarily. By Step 2, there holds
$$H_{\le n+2k-1}(\Omega^{A'_+}M_+ \cup \Omega_M M_+, \Omega^{A'_-}M_+ \cup \Omega_M M_+) \neq 0.$$
There exist $A_- \in (C-\varepsilon, A'_-)$ and $A_+ \in (C, A'_+)$ such that the map
$$\begin{aligned} &H_{\le n+2k-1}(\Omega^{A_+}M_+ \cup \Omega_M M_+, \Omega^{A_-}M_+ \cup \Omega_M M_+) \\ &\quad\to H_{\le n+2k-1}(\Omega^{A'_+}M_+ \cup \Omega_M M_+, \Omega^{A'_-}M_+ \cup \Omega_M M_+) \end{aligned} \tag{4}$$

is nonzero. When δ is sufficiently close to 0, there exists an isotopy $(f_t)_{t\in[0,1]}$ on M_+ which satisfies the following properties:

- $f_0 = \mathrm{id}_{M_+}$,
- $M \subset f_1(M(\delta))$,
- $M \subset f_t(M)$ for any $t \in [0,1]$,
- $\|df_t(v)\| \le \|v\| \cdot \min\left\{\sqrt{A'_-/A_-}, \sqrt{A'_+/A_+}\right\}$ for any $t \in [0,1]$ and $v \in TM_+$.

By the last property,

$$\gamma \in \Omega^{A_-}(M_+) \implies f \circ \gamma \in \Omega^{A'_-}(M_+),$$
$$\gamma \in \Omega^{A_+}(M_+) \implies f \circ \gamma \in \Omega^{A'_+}(M_+).$$

Then (4) factors the map

$$\begin{aligned} &H_{\le n+2k-1}(\Omega^{A_+}M_+ \cup \Omega_M M_+, \Omega^{A_-}M_+ \cup \Omega_M M_+) \\ &\quad \to H_{\le n+2k-1}(\Omega^{A_+}M_+ \cup \Omega_{M(\delta)}M_+, \Omega^{A_-}M_+ \cup \Omega_{M(\delta)}M_+), \end{aligned} \tag{5}$$

in particular (5) is nonzero. Hence, we obtain

$$\varprojlim_{\delta\to 0} H_{\le n+2k-1}(\Omega^{A_+}M_+ \cup \Omega_{M(\delta)}M_+, \Omega^{A_-}M_+ \cup \Omega_{M(\delta)}M_+) \ne 0. \tag{6}$$

Now (3) follows from (6) and the following isomorphisms which hold for any $\delta > 0$:

$$\begin{aligned} &H_*(\Omega^{A_+}M_+ \cup \Omega_{M(\delta)}M_+, \Omega^{A_-}M_+ \cup \Omega_{M(\delta)}M_+) \\ &\quad \cong H_*(\Omega^{A_+}\mathrm{int}\, M \cup \Omega_{M(\delta)}\mathrm{int}\, M, \Omega^{A_-}\mathrm{int}\, M \cup \Omega_{M(\delta)}\mathrm{int}\, M) \\ &\quad \cong H_*(\Omega^{A_+}M \cup \Omega_{M(\delta)}M, \Omega^{A_-}M \cup \Omega_{M(\delta)}M). \end{aligned}$$

The first isomorphism holds since $\{\Omega^{A_-}M_+ \cup \Omega_{M(\delta)}M_+, \Omega^{A_+}\mathrm{int}\, M \cup \Omega_{M(\delta)}\mathrm{int}\, M\}$ is an open covering of $\Omega^{A_+}M_+ \cup \Omega_{M(\delta)}M_+$. The second isomorphism holds since

$$H_*(\Omega^c M \cup \Omega_{M(\delta)}M, \Omega^c\,\mathrm{int}\, M \cup \Omega_{M(\delta)}\mathrm{int}\, M) = 0 \tag{7}$$

for any $c, \delta > 0$. The proof of (7) is parallel to the proof of Lemma 4.1 in [11], which is an analogue of (7) for the free path space.

Step 4. Theorem 1.2 (ii) of [11] claims the following:

> Let $0 < a < b$. If $\varprojlim_{\delta\to 0} H_j(\Omega^b M \cup \Omega_{M(\delta)}M, \Omega^a M \cup \Omega_{M(\delta)}M) \ne 0$, then there exists a periodic billiard trajectory γ on M such that $\sharp\mathscr{B}_\gamma \le j$ and $\mathrm{len}(\gamma) \in [\sqrt{2a}, \sqrt{2b}]$.

By this claim and Step 3, for any $\varepsilon > 0$ there exists a periodic billiard trajectory γ_ε on M such that $\sharp\mathscr{B}_{\gamma_\varepsilon} \le n+2k-1$ and $\mathrm{len}(\gamma_\varepsilon) \in [\sqrt{2(C-\varepsilon)}, \sqrt{2(C+\varepsilon)}]$. Then there exists a sequence $(\varepsilon_j)_j$ such that $\lim_{j\to\infty} \varepsilon_j = 0$ and $(\gamma_{\varepsilon_j})_j$ converges to a periodic billiard trajectory γ. Then γ satisfies $\sharp\mathscr{B}_\gamma \le n + 2k - 1$ and $\mathrm{len}(\gamma) = \sqrt{2C} = c_k^{S^1}(M)$. This completes the proof. $\square$

3.4. Growth of $c_k^{S^1}$

The next property is similar to the corresponding property for (equivariant) symplectic homology capacities ([8] Proposition 3.7). The following proof is also very similar to the proof of [8] Proposition 3.7.

Proposition 3.10. *Let M be any oriented Riemannian manifold. For any integer $k \geq 1$, there holds $c_{k-1}^{S^1}(M) \leq c_k^{S^1}(M) \leq c_{k-1}^{S^1}(M) + c(M)$.*

Corollary 3.11. *If $c(M) < \infty$, then $\sup_{k\geq 1} \frac{c_k^{S^1}(M)}{k} < \infty$.*

The next lemma is an analogue of Proposition 3.5 in [8].

Lemma 3.12. *Let M be any oriented Riemannian manifold without boundary, K be any compact set in M, and a and b be positive real numbers satisfying $\sqrt{2b} > \sqrt{2a} + c(M,K)$. Then, the inclusion map $i_K^{ab} : (\Omega^a M, \Omega_K^a M) \to (\Omega^b M, \Omega_K^b M)$ satisfies $H_*(i_K^{ab}) = 0$.*

Proof. Let $n := \dim M$. The first step in our proof is to define the *loop product*

$$\bullet : H_*(\Omega^A M, \Omega_K^A M) \otimes H_*(\Omega^B M, \Omega_K^B M) \to H_{*-n}(\Omega^C M, \Omega_K^C M)$$

for any positive real numbers A, B, C satisfying $\sqrt{2A} + \sqrt{2B} < \sqrt{2C}$.

Let

$$\begin{aligned} X &:= \Omega^A M \times \Omega^B M, \\ Y &:= \Omega^A M \times_M \Omega^B M := \{(\gamma, \gamma') \in \Omega^A M \times \Omega^B M \mid \gamma(0) = \gamma'(0)\}, \\ U &:= (\Omega_K^A M \times \Omega^B M) \cup (\Omega^A M \times \Omega_K^B M). \end{aligned}$$

Then $(X, U) = (\Omega^A M, \Omega_K^A M) \times (\Omega^B M, \Omega_K^B M)$. We also define the concatenation map $\mathrm{con} : Y \to \Omega^C M$ by

$$\mathrm{con}(\gamma,\gamma')(\theta) := \begin{cases} \gamma\left(\left(1 + \sqrt{\frac{B}{A}}\right)\theta\right) & \left(0 \leq \theta \leq \frac{\sqrt{A}}{\sqrt{A}+\sqrt{B}}\right), \\ \gamma'\left(\left(1 + \sqrt{\frac{A}{B}}\right)\theta - \sqrt{\frac{A}{B}}\right) & \left(\frac{\sqrt{A}}{\sqrt{A}+\sqrt{B}} \leq \theta \leq 1\right). \end{cases}$$

Then $\mathrm{con}(Y \cap U) \subset \Omega_K^C M$. Let us define the loop product $\bullet$ as the composition of the following three maps:

$$\begin{aligned} &H_*(\Omega^A M, \Omega_K^A M) \otimes H_*(\Omega^B M, \Omega_K^B M) \to H_*(X,U) \to H_{*-n}(Y, Y\cap U) \\ &\quad \to H_{*-n}(\Omega^C M, \Omega_K^C M). \end{aligned}$$

The first map is the cross product on homology, the second map is the Gysin map (which we denote by G) and the third map is induced by the concatenation map. Let

$$i_K^B : (M, M \setminus K) \to (\Omega^B M, \Omega_K^B M); \quad p \mapsto c_p.$$

For any $x \in H_*(\Omega^A M, \Omega_K^A M)$, there holds

$$G(x \times H_*(i_K^B)(\mu_{M,K})) = H_*(j)(x),$$

where $j : (\Omega^A M, \Omega^A_K M) \to (Y, U)$ is defined by $j(\gamma) := (\gamma, c_{\gamma(0)})$. Thus,

$$x \bullet H_*(i^B_K)(\mu_{M,K}) = H_*(\mathrm{con} \circ j)(x).$$

On the other hand, $H : [0,1] \times (\Omega^A M, \Omega^A_K M) \to (\Omega^C M, \Omega^C_K M)$ defined by

$$H(t,\gamma)(\theta) := \begin{cases} \gamma\left(\frac{(\sqrt{A}+\sqrt{B})\theta}{\sqrt{A}+t\sqrt{B}}\right) & \left(0 \le \theta \le \frac{\sqrt{A}+t\sqrt{B}}{\sqrt{A}+\sqrt{B}}\right), \\ \gamma(0) & \left(\frac{\sqrt{A}+t\sqrt{B}}{\sqrt{A}+\sqrt{B}} \le \theta \le 1\right) \end{cases}$$

is a homotopy from $\mathrm{con} \circ j$ to i^{AC}_K, thus $H_*(\mathrm{con} \circ j) = H_*(i^{AC}_K)$. In conclusion, we obtain

$$x \bullet H_*(i^B_K)(\mu_{M,K}) = H_*(i^{AC}_K)(x) \qquad (\forall x \in H_*(\Omega^A M, \Omega^A_K M)). \tag{8}$$

Now let us prove Lemma 3.12. By the assumption $\sqrt{2b} > \sqrt{2a} + c(M,K)$, there exists c such that $c(M,K) < \sqrt{2c} < \sqrt{2b} - \sqrt{2a}$. Then, we obtain

$$H_*(i^{ab}_K)(x) = x \bullet H_*(i^c_K)(\mu_{M,K}) = x \bullet 0 = 0 \qquad (\forall x \in H_*(\Omega^a M, \Omega^a_K M)),$$

where the first equality follows from (8) (by setting $A := a$, $B := c$, $C := b$), and the second equality follows from $\sqrt{2c} > c(M,K)$. □

Proof of Proposition 3.10. It is sufficient to show that, for any oriented Riemannian manifold M with $\partial M = \emptyset$, a compact subset $K \subset M$ and $k \ge 1$, there holds

$$c^{S^1}_{k-1}(M,K) \le c^{S^1}_k(M,K) \le c^{S^1}_{k-1}(M,K) + c(M,K).$$

For any $a > 0$, there holds

$$H^{S^1}_*(i^a_K)(\mu_{M,K} \otimes [\mathbb{C}P^{k-1}]) = H^{S^1}_*(i^a_K)(\mu_{M,K} \otimes [\mathbb{C}P^k]) \cap e,$$

where e denotes the Euler class (see Sect. 2). Thus $H^{S^1}_*(i^a_K)(\mu_{M,K} \otimes [\mathbb{C}P^k]) = 0 \implies H^{S^1}_*(i^a_K)(\mu_{M,K} \otimes [\mathbb{C}P^{k-1}]) = 0$, which implies the first inequality $c^{S^1}_{k-1}(M,K) \le c^{S^1}_k(M,K)$.

To prove the second inequality, it is enough to prove

$$\sqrt{2a} > c^{S^1}_{k-1}(M,K) + c(M,K) \implies H^{S^1}_*(i^a_K)(\mu_{M,K} \otimes [\mathbb{C}P^k]) = 0.$$

Let us take a' so that $c^{S^1}_{k-1}(M,K) < \sqrt{2a'} < \sqrt{2a} - c(M,K)$.

Since $[\mathbb{C}P^k] \cap e = [\mathbb{C}P^{k-1}]$ and $\sqrt{2a'} > c^{S^1}_{k-1}(M,K)$, there holds

$$H^{S^1}_*(i^{a'}_K)(\mu_{M,K} \otimes [\mathbb{C}P^k]) \cap e = 0.$$

By the long exact sequence (1), $H^{S^1}_*(i^{a'}_K)(\mu_{M,K} \otimes [\mathbb{C}P^k])$ is in the image of the left vertical map in the following diagram:

$$\begin{tikzcd} H_*(\Omega^{a'} M, \Omega^{a'}_K M) \arrow[r] \arrow[d] \arrow[dr] & H_*(\Omega^a M, \Omega^a_K M) \arrow[d] \\ H^{S^1}_*(\Omega^{a'} M, \Omega^{a'}_K M) \arrow[r] & H^{S^1}_*(\Omega^a M, \Omega^a_K M). \end{tikzcd}$$

Thus, $H_*^{S^1}(i_K^a)(\mu_{M,K} \otimes [\mathbb{C}P^k])$ is in the image of the diagonal map. By Lemma 3.12 and $\sqrt{2a} > \sqrt{2a'} + c(M,K)$, the upper horizontal map is zero, thus the diagonal map is zero. Therefore, $H_*^{S^1}(i_K^a)(\mu_{M,K} \otimes [\mathbb{C}P^k]) = 0$. □

4. Relation to symplectic capacities

In this section, we briefly discuss a conjectural relation between our notion of capacities and symplectic capacities defined using symplectic homology.

We call a pair (W, ω) a *Liouville domain*, if (W, ω) is a compact symplectic manifold with boundary, and there exists a vector field X on W such that $L_X\omega = \omega$ and X points strictly outwards at every point on ∂W.

Remark 4.1. Usually, Liouville domain is a triple (W, ω, X) satisfying the above conditions. However, since all invariants discussed in this section depend only on (W, ω) and do not depend on choice of X, here we call the pair (W, ω) a Liouville domain.

For any Liouville domain (W, ω) of dimension $2n$ and a positive real number a, one can define a graded $\mathbb{Q}$-vector space $\mathrm{SH}_*^a(W, \omega)$ which is called *symplectic homology*, together with a linear map

$$i^a : H_{*+n}(W, \partial W : \mathbb{Q}) \to \mathrm{SH}_*^a(W, \omega).$$

We define *symplectic homology capacity* $c(W, \omega)$ by

$$c(W, \omega) := \inf\{a \mid i^a([W, \partial W]) = 0\}.$$

The idea that one can define symplectic capacities using periodic orbits of Hamiltonian systems goes back at least to Ekeland–Hofer [3]. The above definition is very close to the one in Viterbo [16] Sect. 5.3, which uses symplectic cohomology. Essentially the same construction was introduced by Floer–Hofer–Wysocki [7] for open sets in symplectic vector spaces, using symplectic homology developed by Floer–Hofer [6].

One can also define a sequence of capacities using the S^1-symmetry of free loop spaces. This idea goes back at least to Ekeland–Hofer [4], and its analogue in Floer theory was already introduced in Viterbo [16] Sect. 5.3. More recently, Gutt–Hutchings [9] defined a sequence of capacities $(c_k^{\mathrm{GH}})_{k \ge 1}$ for Liouville domains using S^1-equivariant positive symplectic homology, which we call Gutt–Hutchings capacities.

To continue our discussion, let us introduce the notion of Liouville domains with corners.

Definition 4.2. (i): Let W be a manifold with corners of dimension $2n$, and X be a C^∞-vector field on W. We say that X *points strictly outwards* at $p \in \partial W$ if the following property holds: Let $(x_1, \ldots, x_{2n})$ be a local chart defined near p which is modeled on $(\mathbb{R}_{\ge 0})^k \times \mathbb{R}^{2n-k}$ and p corresponds to $(0, \ldots, 0)$. Let $X = \sum_{j=1}^{2n} X_j \frac{\partial}{\partial x_j}$. Then $X_1(p), \ldots, X_k(p) < 0$.

(ii): A *Liouville domain with corners* is a pair (W, ω) such that W is a compact manifold with corners and ω is a symplectic form on W, such that there exists a vector field X on W satisfying $L_X\omega = \omega$ and X points strictly outwards at every point on ∂W.

Any Liouville domain with corners can be approximated by a sequence of Liouville domains. By (restricted) monotonicity and conformality (see [9] Theorem 1.24), one can define capacities of Liouville domains with corners by taking limits.

Let M be a compact and connected Riemannian manifold with corners. Let ω_M be the canonical symplectic form on T^*M. Namely, $\omega_M = \sum_{i=1}^n dp_i dq_i$ where $q_1,\dots,q_n$ are local charts on M, and $p_1,\dots,p_n$ are charts on fibers with respect to the local frame $dq_1,\dots,dq_n$. Also, we define the canonical fiberwise radial vector field $R_M \in \mathscr{X}(T^*M)$ by $R_M := \sum_{i=1}^n p_i \frac{\partial}{\partial p_i}$. Finally, let D^*M be the unit disk cotangent bundle of M, i.e.

$$D^*M := \{(q,p) \in T^*M \mid q \in M,\, p \in T^*_qM,\ \|p\| \le 1\}.$$

Lemma 4.3. *(D^*M, ω_M) is a Liouville domain with corners.*

Proof. When $\partial M = \emptyset$, then $L_{R_M}\omega_M = \omega_M$ and R_M points strictly outwards at every point on $\partial(D^*M)$.

When $\partial M \ne \emptyset$, take a vector field V on M which points strictly outwards at every point on ∂M. Define $H_V : T^*M \to \mathbb{R}$ by $H_V(q,p) := p(V(q))$, and for any $\varepsilon > 0$, let us define a vector field $X_{V,\varepsilon}$ on T^*M by $X_{V,\varepsilon} := R_M + \varepsilon X_{H_V}$. Then $L_{X_{V,\varepsilon}}\omega_M = \omega_M$ for any ε. Moreover, when ε is sufficiently close to 0, then $X_{V,\varepsilon}$ points strictly outwards at every point on $\partial(D^*M)$. $\square$

Now we can formulate the following conjecture.

Conjecture 4.4. *Let M be a compact, connected, oriented and spin Riemannian manifold with corners. Then*

$$c(M) = c(D^*M, \omega_M), \qquad c_k^{S^1}(M) = c_k^{\mathrm{GH}}(D^*M, \omega_M) \quad (\forall k \ge 1).$$

When $\partial M = \emptyset$, we have seen that $c(M) = \infty$. On the other hand, $c(D^*M, \omega_M) = \infty$ was already observed by Viterbo ([16] Example 2 on pp.1007), based on the well-known isomorphism between symplectic homology of D^*M and homology of ΩM, which was first discovered by Viterbo ([15], [17]). If one assumes an appropriate (quantitative and S^1-equivariant) version of this isomorphism, then one obtains $c_k^{S^1}(M) = c_k^{\mathrm{GH}}(D^*M, \Omega_M)\,(\forall k \ge 1)$ when $\partial M = \emptyset$. A qualitative version of this idea was already discussed in [16] Section 5.2, with applications to the Weinstein conjecture and the Lagrangian embedding problem.

5. Capacities of rectangles

For any positive real numbers $0 < a_1 \le \cdots \le a_n$, let

$$R_{a_1,\dots,a_n} := (0,a_1) \times \cdots \times (0,a_n), \qquad \bar{R}_{a_1,\dots,a_n} := [0,a_1] \times \cdots \times [0,a_n].$$

$\bar{R}_{a_1,\dots,a_n}$ equipped with the Euclidean metric is a compact Riemannian manifold with corners. The goal of this section is to prove the following theorem.

Theorem 5.1. *For any positive real numbers* $0 < a_1 \leq \cdots \leq a_n$ *and any* $k \in \mathbb{Z}_{\geq 1}$,

$$\begin{aligned}&\frac{c_k^{S^1}(\bar{R}_{a_1,\ldots,a_n})}{2}\\ &\quad = \min\{\sqrt{(k_1a_1)^2 + \cdots + (k_na_n)^2} \mid k_1,\ldots,k_n \in \mathbb{Z}_{\geq 0},\, k_1 + \cdots + k_n = k\}.\end{aligned} \tag{9}$$

In the rest of this paper, the RHS of (9) is denoted by $\mu_k(a_1,\ldots,a_n)$.
Let us show that Theorem 5.1 supports the latter assertion of Conjecture 4.4.
First we introduce a few notations. For any $\Omega \subset (\mathbb{R}_{\geq 0})^n$, let

$$X_\Omega := \{(q_1,\ldots,q_n,p_1,\ldots,p_n) \in T^*\mathbb{R}^n \mid (\pi(q_1^2+p_1^2),\ldots,\pi(q_n^2+p_n^2)) \in \Omega\}.$$

For any positive real numbers $0 < a_1 \leq \cdots \leq a_n$, let

$$E^+_{2a_1,\ldots,2a_n} := \{(x_1,\ldots,x_n) \in (\mathbb{R}_{\geq 0})^n \mid (x_1/2a_1)^2 + \cdots + (x_n/2a_n)^2 \leq 1\}.$$

Finally, let $\omega_n := \sum_{i=1}^n dp_i dq_i \in \Omega^2(T^*\mathbb{R}^n)$.

By Theorem 7 of [14], the open symplectic manifolds $(\operatorname{int} D^*\bar{R}_{a_1,\ldots,a_n}, \omega_n)$ and $(\operatorname{int} X_{E^+_{2a_1,\ldots,2a_n}}, \omega_n)$ are symplectomorphic. In particular, for any $\varepsilon \in (0,1)$ there are symplectic embeddings

$$(D^*\bar{R}_{a_1,\ldots,a_n}, (1-\varepsilon)\omega_n) \to (X_{E^+_{2a_1,\ldots,2a_n}}, \omega_n) \to (D^*\bar{R}_{a_1,\ldots,a_n}, (1+\varepsilon)\omega_n).$$

By the (restricted) monotonicity and conformality of Gutt–Hutchings capacities, we obtain

$$\begin{aligned}(1-\varepsilon)^2 c_k^{\mathrm{GH}}(D^*\bar{R}_{a_1,\ldots,a_n}, \omega_n) &\leq c_k^{\mathrm{GH}}(X_{E^+_{2a_1,\ldots,2a_n}}, \omega_n)\\ &\leq (1+\varepsilon)^2 c_k^{\mathrm{GH}}(D^*\bar{R}_{a_1,\ldots,a_n}, \omega_n)\end{aligned}$$

for any $\varepsilon \in (0,1)$. Hence, we obtain

$$c_k^{\mathrm{GH}}(D^*\bar{R}_{a_1,\ldots,a_n}, \omega_n) = c_k^{\mathrm{GH}}(X_{E^+_{2a_1,\ldots,2a_n}}, \omega_n) \quad (\forall k \geq 1). \tag{10}$$

Remark 5.2. The above argument which deduces (10) from Theorem 7 of [14] is suggested by the referee.

On the other hand, by Theorem 1.6 of [9],

$$c_k^{\mathrm{GH}}(X_{E^+_{2a_1,\ldots,2a_n}}, \omega_n) = 2\mu_k(a_1,\ldots,a_n) \quad (\forall k \geq 1). \tag{11}$$

Hence, we have checked that Theorem 5.1 implies the latter assertion of Conjecture 4.4 for rectangles.

Let us explain the plan of this section. In Sect. 5.1, we reduce Theorem 5.1 to key lemmas: Lemmas 5.3 and 5.4. In Sect. 5.2, we prove some preparatory results on Bott-Morse theory on free loop spaces, which we use to prove these lemmas. In Sect. 5.3, we prove these lemmas. In Sect. 5.4, we show an application of Theorem 5.1 to billiard dynamics.

5.1. Reduction to key lemmas

In this subsection, we reduce Theorem 5.1 to Lemmas 5.3 and 5.4, which we state below.

Lemma 5.3. *For any $a > 0$, $k \in \mathbb{Z}_{\geq 0}$ and $b > 2(ka)^2$, there holds*

$$H_{\leq 2k}(\Omega^b\mathbb{R} \cup \Omega_{\bar{R}_a}\mathbb{R}, \Omega_{\bar{R}_a}\mathbb{R}) = 0.$$

Lemma 5.4. *Let $0 < a_1 \leq \cdots \leq a_n$ be real numbers, and $k \in \mathbb{Z}_{\geq 1}$. For any real numbers b_1 and b_2 such that $0 < b_1 < b_2 < 2\mu_k(a_1, \ldots, a_n)^2$, there holds*

$$H^{S^1}_{n+2k-1}(\Omega^{b_2}\mathbb{R}^n \cup \Omega_{\bar{R}_{a_1,\ldots,a_n}}\mathbb{R}^n, \Omega^{b_1}\mathbb{R}^n \cup \Omega_{\bar{R}_{a_1,\ldots,a_n}}\mathbb{R}^n) = 0.$$

Proof of Theorem 5.1 modulo Lemmas 5.3 and 5.4. In this proof, we abbreviate $(a_1, \ldots, a_n)$ by a. By Lemma 3.5, there holds $c^{S^1}_k(\bar{R}_a) = c^{S^1}_k(\mathbb{R}^n, \bar{R}_a)$. Thus, it is sufficient to show that $c^{S^1}_k(\mathbb{R}^n, \bar{R}_a) = 2\mu_k(a)$.

First, we prove

$$c^{S^1}_k(\mathbb{R}^n, \bar{R}_a) \leq 2\mu_k(a). \tag{12}$$

If $b > 2\mu_k(a)$, there exist $k_1, \ldots, k_n \in \mathbb{Z}_{\geq 0}$ and $b_1, \ldots, b_n \in \mathbb{R}_{>0}$, such that $k_1 + \cdots + k_n = k$, $b_1 + \cdots + b_n < b^2/2$ and $b_i > 2(k_i a_i)^2$ $(1 \leq \forall i \leq n)$. Then, there holds

$$\prod_{i=1}^{n}(\Omega^{b_i}\mathbb{R} \cup \Omega_{\bar{R}_{a_i}}\mathbb{R}, \Omega_{\bar{R}_{a_i}}\mathbb{R}) \subset (\Omega^{b^2/2}\mathbb{R}^n \cup \Omega_{\bar{R}_a}\mathbb{R}^n, \Omega_{\bar{R}_a}\mathbb{R}^n).$$

Let us denote the product on the LHS by X. Since $H_{\leq 2k_i}(\Omega^{b_i}\mathbb{R} \cup \Omega_{\bar{R}_{a_i}}\mathbb{R}, \Omega_{\bar{R}_{a_i}}\mathbb{R}) = 0\,(1 \leq \forall i \leq n)$ by Lemma 5.3, the Künneth formula implies that $H_{<n+2k}(X) = 0$. By Remark 2.1, $H^{S^1}_{<n+2k}(X) = 0$. In particular,

$$\begin{aligned}\iota := i^{b_1}_{\bar{R}_{a_1}} \times \cdots \times i^{b_n}_{\bar{R}_{a_n}} &: (\mathbb{R}^n, \mathbb{R}^n \setminus \bar{R}_a) \\ &= \prod_{i=1}^{n}(\mathbb{R}, \mathbb{R} \setminus \bar{R}_{a_i}) \to \prod_{i=1}^{n}(\Omega^{b_i}\mathbb{R} \cup \Omega_{\bar{R}_{a_i}}\mathbb{R}, \Omega_{\bar{R}_{a_i}}\mathbb{R})\end{aligned}$$

satisfies $H^{S^1}_*(\iota)(\mu_{\mathbb{R}^n,\bar{R}_a} \otimes [\mathbb{C}P^{k-1}]) = 0$. On the other hand,

$$i^{b^2/2}_{R_a} : (\mathbb{R}^n, \mathbb{R}^n \setminus \bar{R}_a) \to (\Omega^{b^2/2}\mathbb{R}^n \cup \Omega_{\bar{R}_a}\mathbb{R}^n, \Omega_{\bar{R}_a}\mathbb{R}^n)$$

factors ι, thus $H^{S^1}_*(i^{b^2/2}_{\bar{R}_a})(\mu_{\mathbb{R}^n,\bar{R}_a} \otimes [\mathbb{C}P^{k-1}]) = 0$. Therefore, $c^{S^1}_k(\mathbb{R}^n, \bar{R}_a) \leq b$, thus (12) is proved.

Next we prove

$$c^{S^1}_k(\mathbb{R}^n, \bar{R}_a) \geq 2\mu_k(a). \tag{13}$$

If this does not hold, there exist b_1 and b_2 such that

$$0 < \sqrt{2b_1} < c^{S^1}_k(\mathbb{R}^n, \bar{R}_a) < \sqrt{2b_2} < 2\mu_k(a).$$

Then, the map

$$H^{S^1}_{n+2k-2}(\Omega^{b_1}\mathbb{R}^n \cup \Omega_{\bar{R}_a}\mathbb{R}^n, \Omega_{\bar{R}_a}\mathbb{R}^n) \to H^{S^1}_{n+2k-2}(\Omega^{b_2}\mathbb{R}^n \cup \Omega_{\bar{R}_a}\mathbb{R}^n, \Omega_{\bar{R}_1}\mathbb{R}^n)$$

is not injective. This contradicts Lemma 5.4, thus (13) is proved. $\square$

5.2. Bott–Morse theory on free loop spaces

5.2.1. Setting. Let X be an open subset of $\Omega\mathbb{R}^n = L^{1,2}(S^1, \mathbb{R}^n)$. For any $\gamma \in X$, the tangent space $T_\gamma X$ is naturally identified with $\Omega\mathbb{R}^n$, in particular it is equipped with the inner product and the norm

$$\langle \eta, \zeta \rangle := \int_{S^1} \langle \eta(t), \zeta(t) \rangle + \langle \eta'(t), \zeta'(t) \rangle \, dt, \qquad \|\eta\| := \sqrt{\langle \eta, \eta \rangle}.$$

Let $f : X \to \mathbb{R}$ be a C^∞-function. The vector field ∇f on X is defined so that $\langle \nabla f, \eta \rangle = df(\eta)$ for any $\eta \in TX$. Let $\mathrm{CrP}(f) \subset X$ denote the set of critical points of f, and $\mathrm{CrV}(f) \subset \mathbb{R}$ denote the set of critical values of f, namely

$$\mathrm{CrP}(f) := \{x \in X \mid \nabla f(x) = 0\}, \qquad \mathrm{CrV}(f) := \{f(x) \mid x \in \mathrm{CrP}(f)\}.$$

For any $x \in \mathrm{CrP}(f)$, let $\mathrm{Hess}(f : x)$ denote the Hessian of f at x, and let $\nu(f : x)$ and $\mathrm{ind}(f : x)$ denote the nullity and index of $\mathrm{Hess}(f : x)$.

f is called Bott–Morse if the following conditions hold:

- Every connected component of $\mathrm{CrP}(f)$ is a finite-dimensional compact submanifold of X. Let $\{R_i\}_{i \in I}$ denote the set of these connected components.
- For every $x \in R_i$, $\ker(\mathrm{Hess}(f : x)) = T_x R_i$. Also, $\mathrm{ind}(f : x)$ is constant on $x \in R_i$, which we denote by λ_{R_i}. We define the vector bundle $N(R_i) \to R_i$ so that $N(R_i)_x$ is the negative eigenspace of $\mathrm{Hess}(f : x)$ for every $x \in R_i$.

We also consider the following conditions:

- A sequence $(x_j)_{j \geq 1}$ on X is called a *PS (Palais–Smale)-sequence* if $(f(x_j))_{j \geq 1}$ is bounded and $\lim_{j \to \infty} \|\nabla f(x_j)\| = 0$. f satisfies the *PS-condition* if any PS-sequence contains a convergent subsequence.
- f is *forward-complete* if the following condition is satisfied: if $J \subset \mathbb{R}$ is a closed interval with $\min J > -\infty$, then $f^{-1}(J)$ is closed in $\Omega\mathbb{R}^n$. In particular, $f^{-1}(J)$ equipped with the $L^{1,2}$-metric is complete.

Lemma 5.5. *Let X be an open subset of $\Omega\mathbb{R}^n$, and $f : X \to \mathbb{R}$ be a Bott–Morse function which is forward-complete and satisfies the PS-condition. Let $a < c < b$ be real numbers such that $\mathrm{CrV}(f) \cap [a, b] = \{c\}$. Let $R_1, \ldots, R_m$ be connected components of $\mathrm{CrP}(f) \cap f^{-1}(c)$. Then the following holds.*

(i): $H_*(\{f \leq b\}, \{f \leq a\}) \cong \bigoplus_{i=1}^m H_*(N(R_i), N(R_i) \setminus R_i)$.

(ii): *Suppose that X admits a smooth S^1-action such that f is S^1-invariant. Then $N(R_i) \to R_i$ admits a natural S^1-action, and*

$$H_*^{S^1}(\{f \leq b\}, \{f \leq a\}) \cong \bigoplus_{i=1}^m H_*^{S^1}(N(R_i), N(R_i) \setminus R_i).$$

Remark 5.6. Lemma 5.5 follows from standard results in Bott–Morse theory in infinite-dimensional setting. See Lemma 3.2, Theorem 7.4 and Corollary 7.2 in [2] Chapter 1. Note that [2] works on *complete* Banach–Finsler manifolds (the completeness is essentially used in the proof of the deformation Lemma, that is Lemma 3.2 in [2]). On the other hand X is an open subset of $\Omega\mathbb{R}^n$, thus (in general) not complete. However, since we have assumed that f is forward-complete, the arguments in [2] work in our setting.

5.2.2. Variational method for Lagrangian action functional. Let $R = R_{a_1,\dots,a_n}$ be an open rectangle. For any $U \in C^\infty(R)$, we define $\mathscr{L}_U : \Omega R \to \mathbb{R}$ by

$$\mathscr{L}_U(\gamma) := \int_{S^1} \frac{|\gamma'|^2}{2} - U(\gamma).$$

Then, $\gamma \in \mathrm{CrP}(\mathscr{L}_U)$ if and only if $\gamma'' + \nabla U(\gamma) \equiv 0$.

Lemma 5.7. *There exists $u \in C^\infty(0,1)$ satisfying the following conditions:*

- $u(t) = u(1-t)$ *for any* $0 < t < 1$.
- $\min u = u(1/2) = 0$.
- $u''(t) > 0$ *for any* $0 < t < 1$.
- $u^{(3)}(t) < 0$ *and* $u^{(3)}(1-t) > 0$ *for any* $0 < t < 1/2$.
- *There exist* $\delta > 0$ *and* $C > 0$ *such that* $u(t) = C + t^{-2}$, $u(1-t) = C + t^{-2}$ *for any* $0 < t < \delta$.

If u satisfies these conditions, then

$$2u(t)u''(t) - u'(t)^2 > 0 \qquad (\forall t \in (0,1) \setminus \{1/2\}). \tag{14}$$

Proof. Take positive real numbers a, b, c, such that $a < 1/2$ and

$$-2a^{-3} = b\left(a - \frac{1}{2}\right) + c\left(a - \frac{1}{2}\right)^3, \qquad 6a^{-4} > b + 3c\left(a - \frac{1}{2}\right)^2.$$

Then, let us define a continuous function $w : (0,1) \to \mathbb{R}$ by

$$w(t) := \begin{cases} -2t^{-3} & (0 < t \le a) \\ b\left(t - \frac{1}{2}\right) + c\left(t - \frac{1}{2}\right)^3 & (a \le t \le 1/2) \\ -w(1-t) & (1/2 \le t \le 1). \end{cases}$$

By "mollifying" w, one can define a C^∞-function $v : (0,1) \to \mathbb{R}$ such that

- $v(t) = -v(1-t)$ and $v'(t) > 0$ for any $0 < t < 1$,
- $v''(t) < 0$ and $v''(1-t) > 0$ for any $0 < t < 1/2$,
- $v(t) = -2t^{-3}$ and $v(1-t) = 2t^{-3}$ for any t sufficiently close to 0.

Then, $u : (0,1) \to \mathbb{R}$ defined by $u(t) := \int_{1/2}^t v(s)\,ds$ satisfies required conditions.

To prove (14), let $m(t) := 2u(t)u''(t) - u'(t)^2$. It is easy to check that $m(1/2) = 0$ and $m'(t) < 0 < m'(1-t)$ for any $t \in (0,1/2)$. Then (14) follows immediately. $\square$

For any $a_1, \dots, a_n$, let us define $U_{a_1,\dots,a_n} \in C^\infty(R_{a_1,\dots,a_n})$ by

$$U_{a_1,\dots,a_n}(x_1, \dots, x_n) := u(x_1/a_1) + \cdots + u(x_n/a_n).$$

We are going to show that $\mathscr{L}_{\varepsilon U_{a_1,\dots,a_n}}$ is a Bott–Morse function on $\Omega R_{a_1,\dots,a_n}$.

First we consider the case $n = 1$. Let us fix $a > 0$ and $\varepsilon > 0$. For any $E > 0$, there exists a unique pair $(\gamma_E, T(E))$ such that $T(E) > 0$ and $\gamma_E : [0, T(E)] \to (0, a)$ is a C^∞-map satisfying

$$\gamma_E{}'' + \varepsilon \nabla U_a(\gamma_E) \equiv 0, \quad \gamma_E(0) = a/2, \quad \gamma_E{}'(0) = \sqrt{2E},$$
$$\gamma_E{}'(t) \geq 0 \, (0 \leq \forall t \leq T(E)), \quad \gamma_E{}'(T(E)) = 0.$$

Lemma 5.8. *$T'(E) < 0$ for any $E > 0$.*

Proof. Let us define a diffeomorphism $v : [0, \infty) \to [0, a/2)$ so that

$$\varepsilon U_a\left(\frac{a}{2} + v(y)\right) = y \quad (\forall y \in [0, \infty)). \tag{15}$$

Then

$$T(E) = \int_0^{v(E)} \frac{dx}{\gamma_E{}'(\gamma_E{}^{-1}(x))} = \int_0^E \frac{v'(y)}{\sqrt{2(E-y)}}\, dy,$$
$$T'(E) = \frac{1}{2\sqrt{2E}} \int_0^1 \frac{v'(Ez) + 2Ez \cdot v''(Ez)}{\sqrt{1-z}}\, dz.$$

On the other hand, by direct computations (15) implies that

$$v'(y) + 2yv''(y) = \frac{a}{\varepsilon} \cdot \frac{u'(A(y))^2 - 2u(A(y))u''(A(y))}{u'(A(y))^3} \quad \left(A(y) := \frac{1}{2} + \frac{v(y)}{a}\right)$$

for any $y > 0$. Then by (14), $v'(y) + 2yv''(y) < 0$ for any $y > 0$. Thus, $T'(E) < 0$ for any $E > 0$. $\square$

For each integer $k \geq 1$, we define E_k by $T(E_k) = 1/4k$. Also, there exists unique $\gamma : S^1 \to (0, a)$ satisfying $\gamma'' + \varepsilon \nabla U_a(\gamma) \equiv 0$ and $\gamma|_{[0,1/4k]} = \gamma_{E_k}$. Let us denote it as $\gamma_k^{\varepsilon,a}$.

Lemma 5.9. *For any $\varepsilon, a > 0$ and $k \in \mathbb{Z}_{\geq 1}$, there holds $\nu(\gamma_k^{\varepsilon,a}) = 1$, $\mathrm{ind}(\gamma_k^{\varepsilon,a}) = 2k$.*

Proof. Let $T^*R_a := \{(q, p) \mid q \in R_a, p \in T_q^*\mathbb{R}\}$, and define $h \in C^\infty(\mathbb{R}^a)$ by $h(q, p) := \varepsilon u_a(q) + p^2/2$. Let X_h denote its Hamiltonian vector field, namely

$$X_h(q, p) = \frac{\partial h}{\partial p}\frac{\partial}{\partial q} - \frac{\partial h}{\partial q}\frac{\partial}{\partial p} = p\frac{\partial}{\partial q} - \varepsilon u_a'(q)\frac{\partial}{\partial p}.$$

Let $(\varphi_h^t)_{t \in \mathbb{R}}$ denote the isotopy on T^*R_a generated by X_h. Then, elements in $\mathrm{CrP}(\mathscr{L}_{\varepsilon u_a})$ correspond to fixed points of φ_h^1. For any $k \in \mathbb{Z}_{\geq 1}$ and $t \in \mathbb{R}$, let $\Phi_k^t := d\varphi_h^t(a/2, \sqrt{2E_k})$. Then, $(\Phi_k^t)_{t \in \mathbb{R}}$ is a path of symplectic matrices. By [12] Theorem 1 on pp.168, we obtain

$$\nu(\gamma_k^{\varepsilon,a}) = \dim \ker(\Phi_k^1 - I_2), \qquad \mathrm{ind}(\gamma_k^{\varepsilon,a}) = \mathrm{ind}_{\mathrm{CZ}}(\Phi_k^t)_{0 \leq t \leq 1},$$

where I_2 denotes the unit matrix, and $\mathrm{ind}_{\mathrm{CZ}}$ denotes the Conley–Zehnder index.

For any $t \in \mathbb{R}$, let us define $\rho^q(t)$, $\rho^p(t)$, $\theta^q(t)$, $\theta^p(t)$ by

$$\Phi_k^t(\partial/\partial q) = \rho^q(t)(\cos\theta^q(t), -\sin\theta^q(t)), \quad \Phi_k^t(\partial/\partial p) = \rho^p(t)(\sin\theta^p(t), \cos\theta^p(t)),$$

$\rho^q(t), \rho^p(t) > 0\,(\forall t \in \mathbb{R})$ and $\theta^q(0) = \theta^p(0) = 0$. Then $\theta^q(1) = 2k\pi$ and $\theta^p(1) < \frac{4k+1}{2}\pi$. Moreover, Lemma 5.8 ($T'(E) < 0$ for any $E > 0$) implies that $\theta^p(1) > 2k\pi$. Then, it follows that $\dim\ker(\Phi_k^1 - I_2) = 1$, and $\mathrm{ind}_{\mathrm{CZ}}(\Phi_k^t)_{0\le t\le 1} = 2k$. □

Lemma 5.10. *Let a be a positive real number, let $(\varepsilon_j)_{j\ge1}$ be a sequence of positive real numbers satisfying $\lim_{j\to\infty}\varepsilon_j = 0$, and let $(k_j)_{j\ge1}$ be a non-decreasing sequence of positive integers. Then $\lim_{j\to\infty}\frac{\mathscr{L}_{\varepsilon_j U_a}(\gamma_{k_j}^{\varepsilon_j,a})}{k_j^2} = 2a^2$.*

Proof. For simplicity, we only consider the case $a = 1$. For every integer $j \ge 1$, let us define $\Gamma_j : [0,1] \to R_1$ by $\Gamma_j(t) := \gamma_{k_j}^{\varepsilon_j,1}\left(\frac{4k_j-1+t}{4k_j}\right)$. Also, let $c_j := \frac{\varepsilon_j}{16k_j^2}$. Then, direct computations show

$$\Gamma_j{}'(0) = 0, \quad \Gamma_j{}'(t) \ge 0\,(\forall t \ge 0), \quad \Gamma_j(1) = 1/2, \quad \Gamma_j{}'' + c_j\nabla u(\Gamma_j) \equiv 0,$$
$$\frac{\mathscr{L}_{\varepsilon_j U_1}(\gamma_{k_j}^{\varepsilon_j,1})}{k_j^2} = 16\left(\int_0^1 \frac{(\Gamma_j{}')^2}{2} - c_j u(\Gamma_j)\,dt\right) = 16\left(\|\Gamma_j'\|_{L^2}^2 - \frac{(\Gamma_j{}'(1))^2}{2}\right).$$

Hence, it is sufficient to prove that

$$\lim_{j\to\infty}\|\Gamma_j'\|_{L^2}^2 - \frac{(\Gamma_j{}'(1))^2}{2} = \frac{1}{8}. \tag{16}$$

For every $j \ge 1$, let $h_j := \Gamma_j(0)$. By $\lim_{j\to\infty} c_j = 0$ we obtain $\lim_{j\to\infty} h_j = 0$. Take positive real numbers C and δ so that $u(t) = C + t^{-2}$ for any $t \in (0,\delta]$. For any $h \in [h_j,\delta]$, let us define $T_j(h)$ so that $\Gamma_j(T_j(h)) = h$. Then, direct computations show

$$T_j(h) = \int_{h_j}^h \frac{xh_j}{\sqrt{2c_j(x^2-h_j^2)}}\,dx = h_j\sqrt{\frac{h^2-h_j^2}{2c_j}}.$$

If j is sufficiently large so that $2h_j \le \delta$, then

$$1 \ge T_j(\delta) = h_j\sqrt{\frac{\delta^2-h_j^2}{2c_j}} \ge \frac{h_j\delta}{2\sqrt{2c_j}}.$$

Hence, $S := \sup_j \frac{h_j}{\sqrt{2c_j}} < \infty$. For any $h \in (0,\delta]$ and sufficiently large j, we obtain $T_j(h) \le Sh$. On the other hand $\lim_{j\to\infty}\frac{1/2-h}{(1-T_j(h))\cdot\Gamma_j'(1)} = 1$ for any $h \in (0,\delta]$, thus $\lim_{j\to\infty}\Gamma_j'(1) = 1/2$. Moreover, for any $h \in (0,\delta]$ and sufficiently large j,

$$\sqrt{1-Sh}\cdot\Gamma_j'(T_j(h)) \le \sqrt{1-T_j(h)}\cdot\Gamma_j'(T_j(h)) \le \|\Gamma_j'\|_{L^2} \le \Gamma_j'(1).$$

Thus, $\lim_{j\to\infty}\|\Gamma_j'\|_{L^2} = 1/2$. Hence, we obtain (16). □

For any $k \ge 1$, let $\Gamma_k^{\varepsilon,a}$ denote the set of reparametrizations of $\gamma_k^{\varepsilon,a}$. Then $\Gamma_k^{\varepsilon,a}$ is a submanifold of ΩR_a which is diffeomorphic to S^1. Also, let $\gamma_0^{\varepsilon,a}$ be the constant loop at $a/2$, and $\Gamma_0^{\varepsilon,a} := \{\gamma_0^{\varepsilon,a}\}$. Then $\mathrm{CrP}(\mathscr{L}_{\varepsilon U_a}) = \bigsqcup_{k\geq 0} \Gamma_k^{\varepsilon,a}$. Lemma 5.9 shows that $\mathscr{L}_{\varepsilon U_a}$ is a Bott–Morse function, and

$$\text{rank of } N(\Gamma_k^{\varepsilon,a}) = \begin{cases} 2k & (k \geq 1), \\ 1 & (k = 0). \end{cases}$$

$\mathscr{L}_{\varepsilon U_a}$ is invariant by the natural S^1-action on ΩR_a. Therefore, for each k, $\Gamma_k^{\varepsilon,a}$ and $N(\Gamma_k^{\varepsilon,a})$ admit the natural S^1-action. When $k \geq 1$, the S^1-action on $\Gamma_k^{\varepsilon,a} \cong S^1$ is k-fold.

Now let n be a positive integer, and $0 < a_1 \leq \cdots \leq a_n$ be positive real numbers. $\mathscr{L}_{\varepsilon U_{a_1,\ldots,a_n}}$ is a Bott–Morse function on $\Omega R_{a_1,\ldots,a_n}$, and

$$\mathrm{CrP}(\mathscr{L}_{\varepsilon U_{a_1,\ldots,a_n}}) = \bigsqcup_{k_1,\ldots,k_n \geq 0} \Gamma_{k_1}^{\varepsilon,a_1} \times \cdots \times \Gamma_{k_n}^{\varepsilon,a_n}.$$

Let us set $\Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n} := \Gamma_{k_1}^{\varepsilon,a_1} \times \cdots \times \Gamma_{k_n}^{\varepsilon,a_n}$. The functional $\mathscr{L}_{\varepsilon U_{a_1,\ldots,a_n}}$ is invariant by the natural S^1-action on $\Omega R_{a_1,\ldots,a_n}$, thus S^1 naturally acts on $\Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n}$ and $N(\Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n})$.

Lemma 5.11. *For any integers $k_1, \ldots, k_n \geq 0$, let*

$$\nu(k_1, \ldots, k_n) := \sharp\{i \mid k_i = 0\}, \qquad c_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n} := \mathscr{L}_{\varepsilon U_{a_1,\ldots,a_n}}(\Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n}).$$

The following claims hold.

(i): $\Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n}$ *is diffeomorphic to* $T^{n-\nu(k_1,\ldots,k_n)}$.

(ii): *Any* $\gamma \in \Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n}$ *satisfies*

$$\nu(\gamma) = n - \nu(k_1, \ldots, k_n), \qquad \mathrm{ind}(\gamma) = 2(k_1 + \cdots + k_n) + \nu(k_1, \ldots, k_n).$$

(iii): $H_{<2(k_1+\cdots+k_n)+\nu(k_1,\ldots,k_n)}(N(\Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n}), N(\Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n}) \setminus \Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n}) = 0.$

(iv): *If* $(k_1, \ldots, k_n) \neq (0, \ldots, 0)$, *then*

$$H^{S^1}_{\geq n+2(k_1+\cdots+k_n)}(N(\Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n}), N(\Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n}) \setminus \Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n}) = 0.$$

(v): $\lim_{\varepsilon\to 0} c_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n} = 2\{(k_1a_1)^2 + \cdots + (k_na_n)^2\}$.

(vi): *Let* $b \in (0, \infty) \setminus \{2\{(k_1a_1)^2 + \cdots + (k_na_n)^2\} \mid k_1, \ldots, k_n \in \mathbb{Z}_{\geq 0}\}$. *Then, for sufficiently small* $\varepsilon > 0$, *there holds*

$$c_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n} < b \iff (k_1a_1)^2 + \cdots + (k_na_n)^2 < b/2$$

for any $k_1, \ldots, k_n \in \mathbb{Z}_{\geq 0}$.

Proof. (i) and (ii) are straightforward. (iii) follows from (ii). (iv) holds since

$$\begin{aligned} &H^{S^1}_{*+\nu(k_1,\ldots,k_n)}(N(\Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n}), N(\Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n}) \setminus \Gamma_{k_1,\ldots,k_n}^{\varepsilon,a_1,\ldots,a_n}) \\ &\cong H^{S^1}_{*}\Bigl(\prod_{k_i\neq 0} (N(\Gamma_{k_i}^{\varepsilon,a_i}), N(\Gamma_{k_i}^{\varepsilon,a_i}) \setminus \Gamma_{k_i}^{\varepsilon,a_i})\Bigr), \end{aligned}$$

and Lemma 5.12 (see below) shows that the RHS is zero if $* \geq n - \nu(k_1, \ldots, k_n) + \sum_{i=1}^n 2k_i$. (v) and (vi) follow from Lemma 5.10. □

In the next lemma, we essentially use that we are working on $\mathbb{Q}$-coefficients.

Lemma 5.12. *Let l be a positive integer, $V^1, \dots, V^l$ be real vector bundles on S^1, and $m_1, \dots m_l$ be positive integers. For each $i \in \{1, \dots, l\}$, let $r_i := \operatorname{rk} V_i$. Suppose that, for each i, V^i admits an S^1-action of the form*

$$\theta \cdot (t, v) = (t + m_i\theta, v') \qquad (\theta \in S^1,\, t \in S^1,\, v \in V^i_t,\, v' \in V^i_{t+m_i\theta}).$$

Let T^l denote the lth product of S^1, and consider the vector bundle $V := V^1 \times \cdots \times V^l$ on T^l. Then, with respect to the diagonal S^1-action on V and T^l,

$$H^{S^1}_*(V, V \setminus T^l) \cong \begin{cases} H_{*-(r_1+\cdots+r_l)}(T^{l-1}) & (V^i \textit{is orientable for every} 1 \leq i \leq l), \\ 0 & (\textit{otherwise}). \end{cases}$$

In particular, $H^{S^1}_{\geq l+r_1+\cdots+r_l}(V, V \setminus T^l) = 0$.

Proof. First, we consider the case $l = 1$. By the homeomorphism

$$((V^1, V^1 \setminus S^1) \times ES^1)/S^1 \cong ((V^1_0, V^1_0 \setminus \{0\}) \times ES^1)/\mathbb{Z}_{m_1},$$

we obtain

$$H^{S^1}_*(V^1, V^1 \setminus S^1) \cong H^{\mathbb{Z}_{m_1}}_*(V^1_0, V^1_0 \setminus \{0\}) \cong \begin{cases} \mathbb{Q} & (V^1 \text{is orientable and} * = r_1), \\ 0 & (\text{otherwise}). \end{cases}$$

Next we consider the case $l \geq 2$. Consider the following homeomorphism:

$$\begin{aligned} ((V, V \setminus T^l) \times ES^1)/S^1 \cong{}& ((V^1_0, V^1_0 \setminus \{0\}) \\ &\times (V^2 \times \cdots \times V^l, V^2 \times \cdots \times V^l \setminus T^{l-1}) \times ES^1)/\mathbb{Z}_{m_1}. \end{aligned}$$

The RHS is a fiber bundle over $ES^1/\mathbb{Z}_{m_1}$, where the fiber is homeomorphic to

$$(V^1_0, V^1_0 \setminus \{0\}) \times (V^2 \times \cdots \times V^l, V^2 \times \cdots \times V^l \setminus T^{l-1}).$$

If V^i is orientable for every $i \in \{1, \dots, l\}$, then the homology of the fiber is isomorphic to $H_{*-(r_1+\cdots+r_l)}(T^{l-1})$. The E^2-term of the Leray–Serre spectral sequence of this fiber bundle is isomorphic to the homology of the fiber, and $d_2 = 0$. Hence, we obtain $H^{S^1}_*(V, V \setminus T^l) \cong H_{*-(r_1+\cdots+r_l)}(T^{l-1})$.

If there exists $i \in \{1, \dots, l\}$ such that V^i is not orientable, without loss of generality we may assume that $i \geq 2$. Then $H_*(V^2 \times \cdots \times V^l, V^2 \times \cdots \times V^l \setminus T^{l-1}) = 0$, which implies $H^{S^1}_*(V, V \setminus T^l) = 0$. □

To apply Lemma 5.5, we need the following lemma:

Lemma 5.13. *Let $\varepsilon > 0$ and $0 < a_1 \leq \cdots \leq a_n$ be real numbers.*

(i): *If a sequence $(\gamma_j)_{j\geq 1}$ on ΩR_a is $L^{1,2}$-bounded and satisfies $\lim_{j\to\infty} \operatorname{dist}(\gamma_j, \partial R_{a_1,\dots,a_n}) = 0$, then $\lim_{j\to\infty} \mathscr{L}_{\varepsilon U_{a_1,\dots,a_n}}(\gamma_j) = -\infty$.*

(ii): *$\mathscr{L}_{\varepsilon U_{a_1,\dots,a_n}} : \Omega R_{a_1,\dots,a_n} \to \mathbb{R}$ is forward-complete, and satisfies the PS-condition.*

Proof. These are direct consequences of results in [11] Sect. 2. Although [11] Sect. 2 works on *path spaces* and manifolds *without* corners, the results there extend to the situation in the present lemma. Specifically, (i) follows from Lemma 2.3 of [11] (which is equal to Lemma 3.6 of [1]), the forward-completeness follows from (i), and the PS-condition follows from Lemma 2.8 of [11]. □

Lemma 5.14. *For any $a > 0$, $k \in \mathbb{Z}_{\geq 0}$ and $b > 2(ka)^2$, there holds*

$$H_{\leq 2k+1}(\Omega R_a, \{\mathscr{L}_{\varepsilon U_a} < b\}) = 0$$

for sufficiently small $\varepsilon > 0$.

Proof. By Lemma 5.11 (vi), if $\varepsilon > 0$ is sufficiently small,

$$c_l^{\varepsilon,a} \geq b \implies (la)^2 \geq b/2 \implies l \geq k+1.$$

Then, the present lemma follows from Lemma 5.5 (i) and Lemma 5.11 (iii). □

Lemma 5.15. *Let $0 < a_1 \leq \cdots \leq a_n$ be real numbers and $k \in \mathbb{Z}_{\geq 1}$. Let b_1, b_2 be real numbers such that $0 < b_1 < b_2 < 2\mu_k(a_1, \ldots, a_n)^2$. Then*

$$H^{S^1}_{n+2k-1}(\{\mathscr{L}_{\varepsilon U_{a_1,\ldots,a_n}} < b_2\}, \{\mathscr{L}_{\varepsilon U_{a_1,\ldots,a_n}} < b_1\}) = 0$$

for sufficiently small $\varepsilon > 0$.

Proof. By Lemma 5.11 (vi), if ε is sufficiently small, then

$$c^{\varepsilon,a_1,\ldots,a_n}_{l_1,\ldots,l_n} < b_2 \implies (l_1a_1)^2 + \cdots + (l_na_n)^2 < b_2/2 \implies l_1 + \cdots + l_n \leq k-1.$$

Then, the present lemma follows from Lemma 5.11 (iv) and Lemma 5.5 (ii). □

5.3. Proofs of Lemmas 5.3 and 5.4

Now we prove Lemmas 5.3 and 5.4, thus complete our proof of Theorem 5.1.

Proof of Lemma 5.3. We may assume that $a = 1$. Our goal is to prove that if $b > 2k^2$ then $H_{\leq 2k}(\Omega^b\mathbb{R} \cup \Omega_{\bar{R}_1}\mathbb{R}, \Omega_{\bar{R}_1}\mathbb{R}) = 0$.

Step 1. We prove that, for any $b' > b > 2k^2$, $T > 0$ and $\delta \in (0, 1/2)$, the following map (induced by inclusions) is zero if $\delta' \in (0, \delta)$ is sufficiently small:

$$H_{\leq 2k+1}(\Omega^T R_1, \Omega^b R_1 \cup \Omega^T_{[\delta',1-\delta']} R_1) \to H_{\leq 2k+1}(\Omega R_1, \Omega^{b'} R_1 \cup \Omega_{[\delta,1-\delta]} R_1). \tag{17}$$

To prove this, first notice that for sufficiently small $\varepsilon > 0$

$$H_{\leq 2k+1}(\Omega R_1, \{\mathscr{L}_{\varepsilon U_1} < b\}) = 0, \qquad \{\mathscr{L}_{\varepsilon U_1} < b\} \subset \Omega^{b'} R_1 \cup \Omega_{[\delta,1-\delta]} R_1.$$

Indeed, the first condition follows from Lemma 5.14, and the second condition is easy to check. By Lemma 5.13 (i), $\Omega^T_{[\delta',1-\delta']} R_1 \subset \{\mathscr{L}_{\varepsilon U_1} < b\}$ for sufficiently small $\delta' \in (0, \delta)$. Then (17) is zero since it factors $H_{\leq 2k+1}(\Omega R_1, \{\mathscr{L}_{\varepsilon U_1} < b\})$.

Step 2. We prove that, for any $b' > b > 2k^2$, $T \geq b$ and $\delta \in (0, 1/2)$, the map

$$H_{\leq 2k+1}(\Omega^T\mathbb{R}, \Omega^b\mathbb{R} \cup \Omega^T_{\bar{R}_1}\mathbb{R}) \to H_{\leq 2k+1}(\Omega\mathbb{R}, \Omega^{b'}\mathbb{R} \cup \Omega_{[\delta,1-\delta]}\mathbb{R}) \tag{18}$$

is zero. To prove this, take $\delta' \in (0, \delta)$ so that (17) is zero, and consider the following commutative diagram:

$$\begin{tikzcd} H_{\le 2k+1}(\Omega^T R_1, \Omega^b R_1 \cup \Omega^T_{[\delta',1-\delta']} R_1) \arrow[r] \arrow[d] & H_{\le 2k+1}(\Omega R_1, \Omega^{b'} R_1 \cup \Omega_{[\delta,1-\delta]} R_1) \arrow[d] \\ H_{\le 2k+1}(\Omega^T \mathbb{R}, \Omega^b \mathbb{R} \cup \Omega^T_{[\delta',1-\delta']} \mathbb{R}) \arrow[r] & H_{\le 2k+1}(\Omega \mathbb{R}, \Omega^{b'} \mathbb{R} \cup \Omega_{[\delta,1-\delta]} \mathbb{R}). \end{tikzcd}$$

Then the top horizontal map is zero, and the left and right vertical maps are excision isomorphisms. Thus the bottom horizontal map is zero. Then (18) is also zero since it factors the bottom horizontal map.

Step 3. We prove that, for any $b' > b > 2k^2$ and $T \ge b$, the map

$$H_{\le 2k+1}(\Omega^T \mathbb{R}, \Omega^b \mathbb{R} \cup \Omega^T_{\bar{R}_1} \mathbb{R}) \to H_{\le 2k+1}(\Omega R, \Omega^{b'} \mathbb{R} \cup \Omega_{\bar{R}_1} \mathbb{R}) \tag{19}$$

is zero. To prove this, take b'' and δ so that $\delta > 0$ and $b < b'' \le (1-2\delta)^2 b'$. Define an affine map $A : \mathbb{R} \to \mathbb{R}$ by $A(t) := (t-\delta)/(1-2\delta)$. Then, any $\gamma \in \Omega^{b''} \mathbb{R} \cup \Omega_{[\delta,1-\delta]} \mathbb{R}$ satisfies $A \circ \gamma \in \Omega^{b'} \mathbb{R} \cup \Omega_{\bar{R}_1} \mathbb{R}$. Then we obtain the following commutative diagram:

$$\begin{tikzcd} H_{\le 2k+1}(\Omega^T \mathbb{R}, \Omega^b \mathbb{R} \cup \Omega^T_{\bar{R}_1} \mathbb{R}) \arrow[r, "(18)"] \arrow[dr, "(19)"'] & H_{\le 2k+1}(\Omega \mathbb{R}, \Omega^{b''} \mathbb{R} \cup \Omega_{[\delta,1-\delta]} \mathbb{R}) \arrow[d] \\ & H_{\le 2k+1}(\Omega \mathbb{R}, \Omega^{b'} \mathbb{R} \cup \Omega_{\bar{R}_1} \mathbb{R}). \end{tikzcd}$$

Since the horizontal map (18) is zero by Step 2, the diagonal map (19) is also zero.

Step 4. Since homology commutes with direct limits, for any $b > 0$

$$H_*(\Omega \mathbb{R}, \Omega^b \mathbb{R} \cup \Omega_{\bar{R}_1} \mathbb{R}) \cong \varinjlim_{\substack{T \to \infty \\ \varepsilon \to +0}} H_*(\Omega^T \mathbb{R}, \Omega^{b-\varepsilon} \mathbb{R} \cup \Omega^T_{\bar{R}_1} \mathbb{R}).$$

By Step 3, $H_{\le 2k+1}(\Omega \mathbb{R}, \Omega^b \mathbb{R} \cup \Omega_{\bar{R}_1} \mathbb{R}) = 0$ if $b > 2k^2$. Since $H_*(\Omega \mathbb{R}, \Omega_{\bar{R}_1} \mathbb{R}) = 0$, we obtain $H_{\le 2k}(\Omega^b \mathbb{R} \cup \Omega_{\bar{R}_1} \mathbb{R}, \Omega_{\bar{R}_1} \mathbb{R}) = 0$ if $b > 2k^2$. This completes the proof. □

The proof of Lemma 5.4 is quite similar to the proof of Lemma 5.3.

Proof of Lemma 5.4. We fix $a_1, \dots, a_n$, and abbreviate $R_{a_1,\dots,a_n}$ by R, $U_{a_1,\dots,a_n}$ by U, and $\mu_k(a_1, \dots, a_n)$ by μ. Our goal is to prove

$$H^{S^1}_{n+2k-1}(\Omega^{b_2} \mathbb{R}^n \cup \Omega_{\bar{R}} \mathbb{R}^n, \Omega^{b_1} \mathbb{R}^n \cup \Omega_{\bar{R}} \mathbb{R}^n) = 0$$

for any $0 < b_1 < b_2 < 2\mu^2$. For any $\delta \in (0, a_1/2)$, let $K(\delta) := [\delta, a_1 - \delta] \times \cdots \times [\delta, a_n - \delta]$.

Step 1. For any $0 < b_1 < b_1' < b_2 < b_2' < 2\mu^2$, $T > 0$ and $\delta \in (0, a_1/2)$, the following map (induced by inclusions) is zero if $\delta' \in (0, \delta)$ is sufficiently small:

$$\begin{aligned} &H^{S^1}_{n+2k-1}(\Omega^{b_2}R \cup \Omega^T_{K(\delta')}R, \Omega^{b_1}R \cup \Omega^T_{K(\delta')}R) \\ &\quad \to H^{S^1}_{n+2k-1}(\Omega^{b_2'}R \cup \Omega_{K(\delta)}R, \Omega^{b_1'}R \cup \Omega_{K(\delta)}R). \end{aligned} \tag{20}$$

To prove this, first notice that for sufficiently small $\varepsilon > 0$

$$\begin{aligned} &H^{S^1}_{n+2k-1}(\{\mathscr{L}_{\varepsilon U} < b_2\}, \{\mathscr{L}_{\varepsilon U} < b_1\}) = 0, \\ &\{\mathscr{L}_{\varepsilon U} < b_i\} \subset \Omega^{b_i'}R \cup \Omega_{K(\delta)}R \quad (i \in \{1,2\}). \end{aligned}$$

The first condition follows from Lemma 5.15, and the second condition is easy to check. If δ' is sufficiently small, then $\Omega^{b_i}R \cup \Omega^T_{K(\delta')}R \subset \{\mathscr{L}_{\varepsilon U} < b_i\}$ for $i \in \{1,2\}$. Then (20) is zero since it factors $H^{S^1}_{n+2k-1}(\{\mathscr{L}_{\varepsilon U} < b_2\}, \{\mathscr{L}_{\varepsilon U} < b_1\})$.

Step 2. For any $0 < b_1 < b_1' < b_2 < b_2' < 2\mu^2$, $T \geq b_2$ and $\delta \in (0, a_1/2)$, the map (induced by inclusions)

$$\begin{aligned} &H^{S^1}_{n+2k-1}(\Omega^{b_2}\mathbb{R}^n \cup \Omega^T_{\bar{R}}\mathbb{R}^n, \Omega^{b_1}\mathbb{R}^n \cup \Omega^T_{\bar{R}}\mathbb{R}^n) \\ &\quad \to H^{S^1}_{n+2k-1}(\Omega^{b_2'}\mathbb{R}^n \cup \Omega_{K(\delta)}\mathbb{R}^n, \Omega^{b_1'}\mathbb{R}^n \cup \Omega_{K(\delta)}\mathbb{R}^n) \end{aligned}$$

is zero. The proof is similar to Step 2 in the proof of Lemma 5.3.

Step 3. For any $0 < b_1 < b_1' < b_2 < b_2' < 2\mu^2$ and $T \geq b_2$, the map

$$\begin{aligned} &H^{S^1}_{n+2k-1}(\Omega^{b_2}\mathbb{R}^n \cup \Omega^T_{\bar{R}}\mathbb{R}^n, \Omega^{b_1}\mathbb{R}^n \cup \Omega^T_{\bar{R}}\mathbb{R}^n) \\ &\quad \to H^{S^1}_{n+2k-1}(\Omega^{b_2'}\mathbb{R}^n \cup \Omega_{\bar{R}}\mathbb{R}^n, \Omega^{b_1'}\mathbb{R}^n \cup \Omega_{\bar{R}}\mathbb{R}^n) \end{aligned}$$

is zero. The proof is similar to Step 3 in the proof of Lemma 5.3.

Step 4. Since homology commutes with direct limits,

$$\begin{aligned} &H^{S^1}_{n+2k-1}(\Omega^{b_2}\mathbb{R}^n \cup \Omega_{\bar{R}}\mathbb{R}^n, \Omega^{b_1}\mathbb{R}^n \cup \Omega_{\bar{R}}\mathbb{R}^n) \\ &\quad \cong \lim_{\substack{T \to \infty \\ \varepsilon \to +0}} H^{S^1}_{n+2k-1}(\Omega^{b_2-\varepsilon}\mathbb{R}^n \cup \Omega^T_{\bar{R}}\mathbb{R}^n, \Omega^{b_1-\varepsilon}\mathbb{R}^n \cup \Omega^T_{\bar{R}}\mathbb{R}^n). \end{aligned}$$

By Step 3, this limit is zero if $0 < b_1 < b_2 < 2\mu^2$. This completes the proof. □

5.4. Application to periodic billiard trajectories

Let us say that a periodic billiard trajectory is *prime* if it is not a multiple cover of a shorter trajectory. As an application of Theorem 5.1, we prove the existence of "many" prime periodic billiard trajectories on a billiard table which is C^0-close to a "generic" rectangle.

Proposition 5.16. *Suppose that $0 < a_1 < \cdots < a_n$ are real numbers such that $a_1^2, \ldots, a_n^2$ are linearly independent over $\mathbb{Q}$. For any integer $m \geq 1$, there exists $\varepsilon > 0$ such that, for any open set $U \subset \mathbb{R}^n$ with C^∞-boundary satisfying*

$$\bar{R}_{a_1,\ldots,a_n} \subset U \subset [-\varepsilon, a_1 + \varepsilon] \times \cdots \times [-\varepsilon, a_n + \varepsilon],$$

there exist at least m distinct prime periodic billiard trajectories on U.

Remark 5.17. For convex domains, much better lower bounds of the number of periodic billiard trajectories are known, e.g. [5]. On the other hand, in Proposition 5.16 we do not assume convexity of U. Moreover, we assume nothing about topological type of U.

Proof. The proof consists of four steps.

Step 1. Let $\alpha := (a_1^{-2} + \cdots + a_n^{-2})^{-1/2}$, and abbreviate $R_{a_1,\ldots,a_n}$ by R. Then

$$\frac{c_k^{S^1}}{2k} = \min \left\{ \frac{\sqrt{(k_1 a_1)^2 + \cdots + (k_n a_n)^2}}{k} \,\middle|\, k_1 + \cdots + k_n = k \right\} \geq \alpha,$$

and the equality holds if and only if there exists $(k_1, \ldots, k_n)$ such that $k_1 + \cdots + k_n = k$ and $(k_1 : \cdots : k_n) = (a_1^{-2} : \cdots : a_n^{-2})$. However, this cannot happen since $a_1^2, \ldots, a_n^2$ are linearly independent over $\mathbb{Q}$. On the other hand, $\lim_{k\to\infty} c_k^{S^1}(R)/k = 2\alpha$, thus there exists an increasing sequence of integers $(k_i)_{i\geq 1}$ such that $(c_{k_i}^{S^1}(R)/k_i)_i$ is strictly decreasing.

Step 2. For any $i \geq 1$, let us take $p_{i,1}, \ldots, p_{i,n} \in \mathbb{Z}_{\geq 0}$ such that

$$p_{i,1} + \cdots + p_{i,n} = k_i, \quad c_{k_i}^{S^1}(R) = 2\sqrt{(p_{i,1}a_1)^2 + \cdots + (p_{i,n}a_n)^2}.$$

For any $i < j$ there holds $c_{k_i}^{S^1}(R)/k_i \neq c_{k_j}^{S^1}(R)/k_j$, thus $(p_{i,1} : \cdots : p_{i,n}) \neq (p_{j,1} : \cdots : p_{j,n})$. Since $a_1^2, \ldots, a_n^2$ are linearly independent over $\mathbb{Q}$,

$$\frac{c_{k_j}^{S^1}(R)^2}{c_{k_i}^{S^1}(R)^2} = \frac{(p_{j,1}a_1)^2 + \cdots + (p_{j,n}a_n)^2}{(p_{i,1}a_1)^2 + \cdots + (p_{i,n}a_n)^2} \notin \mathbb{Q}.$$

In particular, $c_{k_j}^{S^1}(R)/c_{k_i}^{S^1}(R) \notin \mathbb{Q}$.

Step 3. For any $\varepsilon > 0$, let us abbreviate $[-\varepsilon, a_1+\varepsilon] \times \cdots \times [-\varepsilon, a_n+\varepsilon]$ by $R(\varepsilon)$. Let us prove that, for any positive integers K and m, there exists $\varepsilon(K,m) > 0$ with the following property: for any open set U with C^∞-boundary,

$$R \subset U \subset R(\varepsilon(K,m)) \implies \left\{ \frac{c_{k_j}^{S^1}(U)}{c_{k_i}^{S^1}(U)} \,\middle|\, 1 \leq i < j \leq m \right\} \cap \bigcup_{k=1}^{K} \frac{1}{k}\mathbb{Z} = \emptyset.$$

If this is not the case, there exist $1 \leq i < j \leq m$, $(U_p)_{p\geq 1}$ and $(\varepsilon_p)_{p\geq 1}$ such that

$$\lim_{p\to\infty} \varepsilon_p = 0, \quad R \subset U_p \subset R(\varepsilon_p)\ (\forall p \geq 1), \quad \frac{c_{k_j}^{S^1}(U_p)}{c_{k_i}^{S^1}(U_p)} \in \bigcup_{k=1}^{K} \frac{1}{k}\mathbb{Z}.$$

Then we obtain

$$\frac{c_{k_j}^{S^1}(R)}{c_{k_i}^{S^1}(R)} = \lim_{p\to\infty} \frac{c_{k_j}^{S^1}(U_p)}{c_{k_i}^{S^1}(U_p)} \in \bigcup_{k=1}^{K} \frac{1}{k}\mathbb{Z},$$

which contradicts Step 2.

Step 4. Let m be a positive integer, and let $K_m := k_m + [(n+1)/2]$. We prove that, for any open set U with C^∞-boundary satisfying $R \subset U \subset$

$R(\varepsilon(K_m, m))$, there exist at least m distinct prime periodic billiard trajectories on U. By Step 3,

$$\left\{ \frac{c^{S^1}_{k_j}(U)}{c^{S^1}_{k_i}(U)} \middle| 1 \le i < j \le m \right\} \cap \bigcup_{k=1}^{K_m} \frac{1}{k}\mathbb{Z} = \emptyset.$$

By Proposition 3.9, for each $1 \le i \le m$, there exists a periodic billiard trajectory γ_i on U with length $c^{S^1}_{k_i}(U)$ and $\sharp\mathscr{B}_{\gamma_i} \le n + 2k_i - 1$. Let γ'_i be a prime periodic billiard trajectory and m_i be an integer, such that γ_i is an m_i-fold multiple of γ'_i. Then $m_i \le k_i + (n-1)/2 \le K_m$, since $2m_i \le \sharp\mathscr{B}_{\gamma'_i} \cdot m_i = \sharp\mathscr{B}_{\gamma_i} \le n + 2k_i - 1$. If $\gamma'_i = \gamma'_j$ for some $1 \le i < j \le m$, then

$$\frac{c^{S^1}_{k_j}(U)}{c^{S^1}_{k_i}(U)} = \frac{\mathrm{len}(\gamma_j)}{\mathrm{len}(\gamma_i)} = \frac{m_j}{m_i} \in \bigcup_{k=1}^{K_m} \frac{1}{k}\mathbb{Z}.$$

This is a contradiction, thus $\gamma'_1, \ldots, \gamma'_m$ are distinct prime periodic billiard trajectories on U. This completes the proof of Proposition 5.16. □

6. Questions

Finally, let us discuss a few questions. In this section, all Riemannian manifolds are assumed to be compact, connected, orientable and with corners.

Question 6.1. Does $\lim_{k\to\infty} \frac{c^{S^1}_k(M)}{k}$ exist for any (or, a generic) Riemannian manifold M? If it exists, does the limit have any geometric/dynamical meaning?

Comment. Theorem 5.1 implies that the limit exists for any rectangle. On the other hand, Gutt–Hutchings [9] proved that, if X is a star-shaped domain in $\mathbb{C}^n$ satisfying a certain condition (see [9] Remark 1.22), then $\lim_{k\to\infty} \frac{c^{\mathrm{GH}}_k(X)}{k}$ exists and is equal to the maximal size of a symplectic cube which admits a symplectic embedding into X.

Question 6.2. Does the following property hold for any (or, a generic) Riemannian manifold M? There exists a sequence $(\gamma_k)_{k\ge1}$ of periodic billiard trajectories on M such that $\mathrm{len}(\gamma_k) = c^{S^1}_k(M)\,(\forall k \ge 1)$ and $\bigcup_{k=1}^\infty \mathrm{Im}(\gamma_k)$ is dense in M.

Comment. Theorem 5.1 implies that this property holds for a generic rectangle.

Let us state our next question as a conjecture.

Conjecture 6.3. *For any integer $k \ge 1$,*

$$c^{S^1}_k(B^2(1)) = 2(k+1) \max_{1\le j\le k} \sin\left(\frac{\pi j}{k+1}\right). \tag{21}$$

Comment. Ramos [13] considers a domain $\Omega \subset (\mathbb{R}_{\geq 0})^2$ which is prescribed by the coordinate axes and the curve

$$C : (2\sin(t/2) - t\cos(t/2), 2\sin(t/2) + (2\pi - t)\cos(t/2)) \quad (0 \leq t \leq 2\pi),$$

and proves ([13] Theorem 3) that $(\mathrm{int}\, D^*B^2(1), \omega_2)$ is symplectomorphic to $(\mathrm{int}\, X_\Omega, \omega_2)$. Since X_Ω is a concave toric domain, one can compute its Gutt–Hutchings capacities by [9] Theorem 1.14. Namely,

$$c_k^{\mathrm{GH}}(X_\Omega) = \max\{\min_{w \in C} v \cdot w \mid v = (v_1, v_2) \in (\mathbb{Z}_{>0})^2,\ v_1 + v_2 = k + 1\},$$

and little computations show that it is equal to the RHS of (21). Thus Conjecture 6.3 follows from Conjecture 4.4 and the argument similar to the proof of (10).

Conjecture 6.3 and its generalizations to (higher dimensional) ellipsoids will be interesting problems. It will be also interesting to compute capacities of $T^n \setminus \mathrm{int}\,(K)$, where T^n is an n-dimensional flat torus and K is a compact and convex subset of T^n such that ∂K is of C^∞ and strictly convex. Billiards on such manifolds are *dispersive billiards*, whose dynamical properties are very different from those of integrable billiards.

Acknowledgements

This paper is inspired and strongly influenced by papers of Claude Viterbo, in particular [16], [17] and [18]. The author appreciates the editors of this special volume for the opportunity to dedicate this paper to Professor Viterbo on the occasion of his 60th birthday. He also thanks the referee for helpful comments, and for suggesting a proof of (10). This research is supported by JSPS KAKENHI Grant Number 18K13407 and 19H00636.

References

[1] Benci, V., Giannoni, F.: Periodic bounce trajectories with a low number of bounce points. Ann. Inst. H. Poincaré Anal. Non Linéaire. **6**, 73–93 (1989)

[2] Chang, K.-C.: Infinite dimensional Morse theory and multiple solution problems, Progress in Nonlinear Differential Equations and their Applications, 6. Birkhäuser Boston Inc, Boston (1993)

[3] Ekeland, I., Hofer, H.: Symplectic topology and Hamiltonian dynamics. Math. Z. **200**, 355–378 (1989)

[4] Ekeland, I., Hofer, H.: Symplectic topology and Hamiltonian dynamics II. Math. Z. **203**, 553–567 (1990)

[5] Farber, M., Tabachinikov, S.: Topology of cyclic configuration spaces and periodic trajectories of multi-dimensional billiards. Topology. **41**, 553–589 (2002)

[6] Floer, A., Hofer, H.: Symplectic homology I: open sets in $\mathbb{C}^n$. Math. Z. **215**, 37–88 (1994)

[7] Floer, A., Hofer, H., Wysocki, K.: Applications of symplectic homology I. Math. Z. **217**, 577–606 (1994)

[8] Ginzburg, V., Shon, J.: On the filtered symplectic homology of prequantization bundles, Internat. J. Math. 29 1850071, 35 pp (2018)

[9] Gutt, J., Hutchings, M.: Symplectic capacities from positive S^1-equivariant symplectic homology, Algebr. Geom. Topol **18**, 3537–3600 (2018)

[10] Irie, K.: Symplectic homology of disc cotangent bundles of domains in Euclidean space. J. Symplectic Geom. **12**, 511–552 (2014)

[11] Irie, K.: Periodic billiard trajectories and Morse theory on loop spaces. Comment. Math. Helv. **90**, 225–254 (2015)

[12] Long, Y.: Index theory for symplectic paths with applications, Progress in Mathematics, 207. Birkhäuser Verlag, Basel (2002)

[13] Ramos, V.: Symplectic embeddings and the Lagrangian bidisk. Duke Math. J. **166**, 1703–1738 (2017)

[14] Ramos, V., Sepe, D.: On the rigidity of Lagrangian products. J. Symplectic Geom. **17**, 1447–1478 (2019)

[15] Viterbo, C.: Generating functions, symplectic geometry, and applications, Proceedings of the International Congress of Mathematicians (Zürich, 1994), Birkahäuser, Basel, pp. 537–547, (1995)

[16] Viterbo, C.: Functors and computations in Floer homology with applications I. Geom. Funct. Anal. **9**, 985–1033 (1999)

[17] Viterbo, C.: Functors and computations in Floer homology with applications II, arXiv:1805.01316

[18] Viterbo, C.: Metric and isoperimetric problems in symplectic geometry. J. Amer. Math. Soc. **13**, 411–431 (2000)

Kei Irie
Research Institute for Mathematical Sciences
Kyoto University
Kyoto 606-8502
Japan
e-mail: iriek@kurims.kyoto-u.ac.jp

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:28
https://doi.org/10.1007/s11784-022-00953-w
Published online April 7, 2022

Journal of Fixed Point Theory and Applications

Remarks on the systoles of symmetric convex hypersurfaces and symplectic capacities

Joontae Kim, Seongchan Kim and Myeonggi Kwon

Abstract. In this note we study the systoles of convex hypersurfaces in $\mathbb{R}^{2n}$ invariant under an anti-symplectic involution. We investigate a uniform upper bound of the ratio between the systole and the symmetric systole of the hypersurfaces using symplectic capacities from Floer theory. We discuss various concrete examples in which the ratio can be understood explicitly.

Mathematics Subject Classification. Primary: 37J45; Secondary: 37C80, 53D40.

Keywords. Systoles, symmetric periodic orbits, real symplectic capacities.

1. Introduction

Let Σ be a smooth manifold of dimension $2n-1$ equipped with a global 1-form α such that $\alpha \wedge (d\alpha)^{n-1}$ is nowhere vanishing. Such a pair (Σ, α) is called a (co-oriented) contact manifold. We assume throughout that Σ is closed and connected. There exists a unique vector field $R = R_\alpha$ on Σ characterized by the conditions $d\alpha(R, \cdot) = 0$ and $\alpha(R) = 1$. The vector field R is called the Reeb vector field associated with α. A periodic (Reeb) orbit is a smooth curve $\gamma\colon \mathbb{R}/\tau\mathbb{Z} \to \Sigma$ solving the differential equation $\dot\gamma = R \circ \gamma$. The *systole* of (Σ, α) is defined as

$$\ell_{\min}(\Sigma, \alpha) = \inf\{\tau > 0 \mid \tau \text{ is the period of a periodic orbit on } (\Sigma, \alpha)\} > 0.$$

By convention, the infimum of the empty set is infinity.

Suppose that the contact manifold (Σ, α) is equipped with an anti-contact involution ρ, meaning that $\rho^2 = \mathrm{Id}$ and $\rho^*\alpha = -\alpha$. The triple

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

(Σ, α, ρ) is called a *real contact manifold*. A periodic orbit γ on (Σ, α, ρ) is said to be *symmetric* if $\rho(\mathrm{Im}(\gamma)) = \mathrm{Im}(\gamma)$. By definition, the *symmetric systole* $\ell_{\min}^{\mathrm{sym}}(\Sigma, \alpha, \rho)$ is the infimum over the periods of symmetric periodic orbits on Σ. We then define the *symmetric ratio* $\mathfrak{R}(\Sigma, \alpha, \rho)$ as

$$\mathfrak{R}(\Sigma, \alpha, \rho) := \frac{\ell_{\min}^{\mathrm{sym}}(\Sigma, \alpha, \rho)}{\ell_{\min}(\Sigma, \alpha)} \in [1, \infty], \tag{1.1}$$

provided that there exists a periodic orbit on (Σ, α).

Example 1.1. Let $\Sigma \subset \mathbb{R}^{2n}$ be a smooth, compact, and starshaped hypersurface with respect to the origin. Assume that Σ is invariant under the complex conjugation

$$\rho_0(x_1, y_1, \ldots, x_n, y_n) = (x_1, -y_1, \ldots, x_n, -y_n).$$

The triple (Σ, α, ρ_0) is a real contact manifold, where $\alpha = \lambda_0|_\Sigma$ is the restriction of the Liouville form $\lambda_0 = \frac{1}{2}\sum_{j=1}^n (x_j \mathrm{d}y_j - y_j \mathrm{d}x_j)$ to Σ. The Reeb orbits on Σ are reparametrizations of the Hamiltonian orbits on Σ of any Hamiltonian $H\colon \mathbb{R}^{2n} \to \mathbb{R}$ having Σ as a regular level set. The existence of a symmetric periodic orbit was established by Rabinowitz [27], implying that $\mathfrak{R}(\Sigma, \alpha, \rho_0)$ is finite.

As of a prominent class of real contact manifolds, we are mainly interested in symmetric convex hypersurfaces in $\mathbb{R}^{2n}$. In this paper convex domains in $\mathbb{R}^{2n}$ are assumed to contain the origin in the interior, and starshaped domains are starshaped with respect to the origin. Let $K \subset \mathbb{R}^{2n}$ be a compact convex domain with smooth boundary which is invariant under an anti-symplectic involution ρ of $\mathbb{R}^{2n}$, i.e. $\rho^2 = \mathrm{Id}$ and $\rho^*\mathrm{d}\lambda_0 = -\mathrm{d}\lambda_0$. We call the boundary of K a *symmetric convex hypersurface*. Assume that the fixed point set $\mathrm{Fix}(\rho)$ intersects the boundary ∂K. This condition necessarily holds if $(\partial K, \rho)$ admits a symmetric periodic orbit. We can find a Liouville form λ on the symplectic manifold $(K, \mathrm{d}\lambda_0)$ such that its Liouville vector field is transverse along the boundary ∂K and ρ is exact with respect to λ i.e. $\rho^*\lambda = -\lambda$. For example one takes the average $\lambda := \frac{1}{2}(\lambda_0 - \rho^*\lambda_0)$, see Lemma 4.7. We define the *(symmetric) systoles* of the symmetric convex hypersurface $(\partial K, \rho)$ by the ones of the real contact manifold $(\partial K, \alpha := \lambda|_{\partial K}, \rho)$:

$$\ell_{\min}(\partial K) := \ell_{\min}(\partial K, \alpha) \quad \text{and} \quad \ell_{\min}^{\mathrm{sym}}(\partial K, \rho) := \ell_{\min}^{\mathrm{sym}}(\partial K, \alpha, \rho). \tag{1.2}$$

They are independent of the choice of the Liouville form λ because the (symmetric) systoles coincide with the minimal actions of (symmetric) closed characteristics on $(\partial K, \rho)$.

Remark 1.2. More precisely, recall that the symplectic form $\mathrm{d}\lambda_0 = \sum_{j=1}^n \mathrm{d}x_j \wedge \mathrm{d}y_j$ induces the *characteristic line bundle* $\mathcal{L}_{\partial K} := \ker(\mathrm{d}\lambda_0|_{\partial K})$ over ∂K. This defines a 1-dimensional foliation of ∂K whose closed leaf γ (i.e. an embedded circle whose tangent spaces lie in $\mathcal{L}_{\partial K}$) is called a *closed characteristic of* ∂K. Its *action* is defined by $\mathcal{A}(\gamma) := \int_{S^1} \gamma^*\lambda$ where λ is a Liouville form on K such that $\mathrm{d}\lambda = \mathrm{d}\lambda_0$. By Stokes theorem $\mathcal{A}(\gamma)$ is independent of the choice of λ. A closed characteristic γ on ∂K is called *symmetric* if γ is invariant under ρ. Now if λ is chosen as above so that $(\partial K, \lambda|_{\partial K}, \rho)$ is a real contact manifold, then there is the correspondence between (symmetric) closed characteristics of ∂K and (symmetric) periodic orbits on $(\partial K, \lambda|_{\partial K})$. Indeed, any closed characteristic of ∂K is parametrized by a periodic Reeb orbit, and vice versa. The action of a closed characteristic and the period of the corresponding periodic orbit coincide. Therefore, (1.2) is independent of the choice of λ, but does depend on the hypersurface ∂K and the symplectic form $\mathrm{d}\lambda_0$.

The main result of this note is the following estimate on the symmetric ratio for symmetric convex hypersurfaces.

Theorem 1.3. *Let $K \subset \mathbb{R}^{2n}$ be a compact and convex domain with smooth boundary which is invariant under an anti-symplectic involution ρ of $\mathbb{R}^{2n}$. Assume that $\mathrm{Fix}(\rho) \cap \partial K \neq \emptyset$. Then the symmetric ratio of the symmetric convex hypersurface $(\partial K, \rho)$ satisfies*

$$1 \leq \mathfrak{R}(\partial K, \rho) \leq 2. \tag{1.3}$$

In particular, on the boundary ∂K, there exists a symmetric periodic orbit of period less than or equal to $2\ell_{\min}(\partial K)$.

It is particularly interesting to ask when the symmetric ratio is exactly equal to one. This means that the smallest period among all periodic orbits can be realized by symmetric one. In Sect. 2 we examine this question with explicit examples of symmetric hypersurfaces. Smooth starshaped toric domains, for instance, admit a family of anti-symplectic involutions including complex conjugation, and we can explicitly understand their Reeb flows. We observe in Sect. 2.3 that the symmetric ratio in this case is always equal to one even without convexity. On the other hand, there are symmetric starshaped domains whose boundary has the symmetric ratio bigger than 1. In Sect. 2.4, we construct such examples by perturbing the standard contact form on the unit sphere following the Bourgeois' perturbation scheme for Morse–Bott contact forms [8, Section 2.2].

Remark 1.4. Even if the symmetric ratio is equal to one, there can exist a non-symmetric periodic orbit of the smallest period. Moreover, a symmetric periodic orbit of the smallest period might not be unique. For example, consider the unit round sphere $S^{2n-1} \subset \mathbb{R}^{2n} \equiv \mathbb{C}^n$ for $n \geq 2$ with complex conjugation ρ_0. The contact form is given by the restriction of the Liouville form as in Example 1.1. The associated Reeb flow is periodic, and the periodic orbit γ through $z \in S^{2n-1}$ can be parametrized as $\gamma(t) = e^{2it}z$, $t \in \mathbb{R}$. Then γ is symmetric with respect to ρ_0 if and only if $\gamma(t_0) \in \mathbb{R}^n$ for some $t_0 \in \mathbb{R}$.

Another interesting aspect of the estimate (1.3) is that it gives a uniform upper bound of the symmetric ratio for convex hypersurfaces in $\mathbb{R}^{2n}$. Such an upper bound does not necessarily exist for a larger class of hypersurfaces. For example, in Sect. 2.5, we exhibit symmetric starshaped hypersurfaces, which are Bordeaux-bottle-shaped, whose symmetric ratio is arbitrary large. In Sect. 2.6, we provide examples of restricted contact type, not starshaped, hypersurfaces in Hamiltonian systems whose symmetric ratio is also arbitrary large.

Our discussions up to now suggest the following questions:

- Is the symmetric ratio for symmetric convex hypersurfaces in $\mathbb{R}^{2n}$ equal to exactly one?
- Under what conditions on real contact manifolds can we find a uniform upper bound of its symmetric ratio? For example, one can consider *dynamical convexity* for contact manifolds as a substitute of geometric convexity.

Remark 1.5. A convex body $K \subset \mathbb{R}^{2n}$, i.e. a compact convex subset in $\mathbb{R}^{2n}$ with non-empty interior, is called *centrally symmetric* if it is invariant under the antipodal map on $\mathbb{R}^{2n}$. Note that the antipodal map is not anti-symplectic but symplectic. It is shown in Akopyan–Karasev [4, Corollary 2.2] that any closed characteristic of minimal action on the boundary of K is itself centrally symmetric, cf. Remark 1.4.

In Sect. 4, we present an approach to obtain the upper bound in Theorem 1.3 employing symplectic capacities from Floer theory. We first bound the symmetric ratio from above in terms of the symplectic homology capacity (the SH capacity) c_{SH} and the wrapped Floer homology capacity (the HW capacity) c_{HW}. An essential ingredient is the recent result of Abbondandolo–Kang [1] and Irie [19] showing for *convex* domains that the systole $\ell_{\min}(\partial K)$ coincides with the SH capacity $c_{\mathrm{SH}}(K)$. Together with the spectral property of the HW capacity in Proposition 4.4, we deduce that

$$\frac{\ell_{\min}^{\mathrm{sym}}(\partial K, \rho)}{\ell_{\min}(\partial K)} \leq \frac{2c_{\mathrm{HW}}(K, \rho)}{c_{\mathrm{SH}}(K)}.$$

We can then bound the ratio of the capacities from above using Floer theory. In Sect. 3.3 we recall a construction of well-known comparison homomorphisms in Floer homology, called *closed-open maps*. They are defined by counting certain Floer disks with one interior puncture (asymptotic to a Hamiltonian 1-orbit) and one boundary puncture (asymptotic to a Hamiltonian 1-chord) with Lagrangian boundary condition. See Fig. 4. We call them *Floer chimneys* as in [5, Figure 11]. Closed-open maps are compatible with the action filtrations on the Floer homologies in the sense of Theorem 3.9. As also observed in [7], it is rather straightforward to obtain the desired upper bound from the existence of filtered closed-open maps.

At the heuristic level the underlying geometric idea is the following. By the spectral properties, the SH capacity $c_{\mathrm{SH}}(K)$ is the action $\mathcal{A}(\gamma)$ of a periodic orbit γ on ∂K and the HW capacity $c_{\mathrm{HW}}(K, \rho)$ is the action $\mathcal{A}(x)$ of a chord x on $(\partial K, \partial\, \mathrm{Fix}(\rho))$. Closed-open maps in principle tell us that there exists a J-holomorphic chimney asymptotic to γ at the interior puncture and asymptotic to x at the boundary puncture. Since the energy of J-holomorphic chimneys is necessarily non-negative, one has $c_{\mathrm{HW}}(K, \rho) = \mathcal{A}(x) \leq \mathcal{A}(\gamma) = c_{\mathrm{SH}}(K)$ by Stokes' theorem.

The wrapped Floer homology capacity for symmetric domains can be seen as a symplectic capacity for symplectic manifolds with symmetries, which we call a *real symplectic capacity*. The upper bound in Theorem 1.3 hinges on relationships between real and non-real symplectic capacities. In

Sect. 5 we discuss further examples of real symplectic capacities which might be of independent interest. See also [11] for more information on symplectic capacities.

Remark 1.6. One finds a motivation to study the symmetric systole in the context of the planar circular restricted three-body problem (PCR3BP). This problem studies the motion of a massless body influenced by two bodies of positive mass according to Newton's law of gravitation, where the two massive bodies move in circles about their common center of mass, and the massless body is confined to the plane determined by the two bodies. Denote by c_* the energy value of the Hamiltonian H of the PCR3BP such that for every $c < c_*$, the level set $H^{-1}(c)$ contains two bounded components near either massive body. In what follows, we concentrate on one of the two bounded components, denoted by Σ_c. It is invariant under the anti-symplectic involution ρ whose fixed point set projects into the configuration space $\mathbb{R}^2$ as a subset of the horizontal axis. In [6] Birkhoff found a Reeb chord on Σ_c via shooting argument and closed it up using ρ to obtain a symmetric periodic orbit, called a *retrograde periodic orbit*. In a real-world situation, a direct periodic orbit is more important since most orbits of moons in the solar system are direct. However, Birkhoff did not give an analytic proof of the existence of a direct periodic orbit. Instead, he conjectured that for each $c < c_*$, the retrograde periodic orbit on Σ_c bounds a disk-like global surface of section. Birkhoff believed that a fixed point of the associated first return map, whose existence is assured by Brouwer's translation theorem, corresponds to a direct periodic orbit. One way to prove this conjecture is to look at the period of the retrograde periodic orbit. Indeed, the SFT-compactness theorem says that if the retrograde periodic orbit has the smallest period, then this would imply Birkhoff's conjecture. For details, we refer to a beautiful exposition [15].

2. Examples

In this section we discuss examples for the symmetric ratio (1.1) on various symmetric hypersurfaces.

2.1. In dimension two

Let W be a subset of $\mathbb{R}^2$ that is diffeomorphic to a closed disc and invariant under an anti-symplectic involution ρ. There exists a unique simply covered periodic orbit γ, which is a parametrization of the ρ-invariant circle ∂W. Moreover, γ is ρ-symmetric. It follows that $\ell_{\min}(\partial W) = \ell_{\min}^{\mathrm{sym}}(\partial W)$ and hence $\mathfrak{R}(\partial W, \rho) = 1$.

2.2. Ellipsoids

Given $a_j \in \mathbb{R}_{>0}$, $j = 1, \ldots, n$, the associated ellipsoid is given by

$$E(a_1, \ldots, a_n) := \left\{ z \in \mathbb{C}^n \;\middle|\; \sum_{j=1}^{n} \frac{\pi |z_j|^2}{a_j} \leq 1 \right\}.$$

With respect to the standard contact form, i.e. the restriction of the Liouville form from Example 1.1, complex conjugation ρ_0 provides an anti-contact involution on the boundary $\partial E(a_1, \ldots, a_n)$. The Reeb flow can explicitly be written by coordinate-wise rotations on $\mathbb{C}^n$. Periodic orbits are of the form

$$\gamma(t) = \left(e^{\frac{2\pi it}{a_1}} z_1, e^{\frac{2\pi it}{a_2}} z_2, \ldots, e^{\frac{2\pi it}{a_n}} z_n\right) \tag{2.1}$$

for some $(z_1, z_2, \ldots, z_n) \in \partial E(a_1, \ldots, a_n)$. Assuming $a_1 \leq a_2 \leq \cdots \leq a_n$ without loss of generality, the periodic orbit $\gamma_1(t) = (e^{\frac{2\pi it}{a_1}} z_1, 0, \ldots, 0)$ with $z_1 \neq 0$ attains the minimal period and is symmetric with respect to ρ_0. Hence the symmetric ratio is equal to one. In general, a periodic orbit of the form (2.1) is ρ_0-symmetric if and only if $\gamma(t_0) \in \mathbb{R}^n$ for some $t_0 \in \mathbb{R}$.

2.3. Smooth starshaped toric domains

Define the moment map $\mu\colon \mathbb{C}^n \to \mathbb{R}^n_{\geq 0}$ as

$$\mu(z_1, \ldots, z_n) = \pi(|z_1|^2, \ldots, |z_n|^2).$$

It is invariant under the exact anti-symplectic involution

$$\rho_\theta(z) = \left(e^{i\theta_1}\overline{z}_1, \ldots, e^{i\theta_n}\overline{z}_n\right) \tag{2.2}$$

for each $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n$. For a domain $\Omega \subset \mathbb{R}^n_{\geq 0}$, the preimage $X_\Omega := \mu^{-1}(\Omega) \subset \mathbb{C}^n$ is called a *toric domain*. Note that any toric domain is ρ_θ-invariant. For example, the ellipsoid $E(a_1, \ldots, a_n)$ is a smooth toric domain associated to the simplex

$$\Omega = \left\{ x \in \mathbb{R}^n_{\geq 0} \;\middle|\; \sum_{j=1}^{n} \frac{x_j}{a_j} \leq 1 \right\}.$$

A toric domain is not necessarily smooth, but, in this note, we only consider *smooth* ones.

In what follows we assume that a domain $\Omega \subset \mathbb{R}^n_{\geq 0}$ is smooth, compact, and starshaped (with respect to the origin). Then the associated toric domain $X_\Omega \subset \mathbb{C}^n$ is a smooth toric domain that is compact and starshaped.

We shall show that $\mathfrak{R}(\partial X_\Omega, \rho_\theta) = 1$ for every $\theta \in \mathbb{R}^n$.

Note that X_Ω is invariant under the $\mathbb{T}^n$-family of the exact symplectomorphisms

$$\sigma_\phi(z) = (e^{i\phi_1} z_1, \ldots, e^{i\phi_n} z_n), \quad \phi = (\phi_1, \ldots, \phi_n) \in \mathbb{R}^n.$$

If γ is a periodic orbit on ∂X_Ω, then so is $\sigma_\phi(\gamma)$. Each fiber torus $\mu^{-1}(w)$, $w \in \Omega$, is foliated by periodic orbits, see e.g. [17, Section 2.2], and hence any periodic orbit on ∂X_Ω is contained in a fiber torus.

For a fixed $\theta \in \mathbb{R}^n$, each fiber torus contains a ρ_θ-symmetric periodic orbit. Indeed, in view of the fact that $\rho_\theta^* R = -R$, where R is the Reeb vector field on ∂X_Ω, a periodic orbit γ is ρ_θ-symmetric if and only if $\gamma(\mathbb{R}) \cap \mathrm{Fix}(\rho_\theta) \neq \emptyset$. For a periodic orbit γ in a fiber torus $\mathcal{T}$, it is always possible to find $\phi \in \mathbb{R}^n$ such that $\sigma_\phi(\gamma)$ intersects $\mathrm{Fix}(\rho_\theta)$. Then $\sigma_\phi(\gamma)$ is a ρ_θ-symmetric periodic orbit in $\mathcal{T}$.

As all periodic orbits belonging to the same fiber torus have the same period, this implies that $\mathfrak{R}(\partial X_\Omega, \rho_\theta) = 1$. Actually, for every periodic orbit γ on ∂X_Ω, there exists $\theta = \theta(\gamma) \in \mathbb{R}^n$ such that γ is a ρ_θ-periodic orbit.

Recall that a toric domain X_Ω is said to be *convex* if

$$\widehat{\Omega} = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid (|x_1|, \dots, |x_n|) \in \Omega\} \subset \mathbb{R}^n$$

is convex. In this case, we know which (symmetric) periodic orbit attains the smallest period. From the convexity of $\widehat{\Omega}$, we can show by computing the Reeb vector field that the fiber orbit at a point of $\partial\Omega$ along a coordinate axis, i.e. an intersection point of $\partial\Omega$ with a coordinate axis, attains the smallest period. Moreover, it is also obvious from the Reeb flow that such a periodic orbit is ρ_θ-symmetric for every $\theta \in \mathbb{R}^n$.

2.4. Starshaped domains with the symmetric ratio slightly bigger than one

For every $\theta \in \mathbb{R}^n$ we can construct a ρ_θ-symmetric starshaped domain K in $\mathbb{R}^{2n}$ with $\mathfrak{R}(\partial K, \rho_\theta) > 1$, where ρ_θ is defined as in (2.2), by perturbing the round sphere. Without loss of generality, we only consider the case of complex conjugation $\rho = \rho_0$.

Let $B \subset (\mathbb{R}^{2n}, \lambda_0)$ denote the closed unit ball. For $h \in C^\infty(\partial B, \mathbb{R})$ with $h \geq 1$, we define the starshaped domain in $\mathbb{R}^{2n}$

$$K_h{:=}B \cup_{\partial B} \{(r, x) \in [1, \infty) \times \partial B \mid x \in \partial B,\ r \leq h(x)\}$$

by attaching the graph of h along the boundary ∂B via the Liouville flow of λ_0. Note that ∂K_h is contactomorphic to the unit sphere ∂B equipped with the contact form $h\alpha_0$. Since the Reeb flow ϕ^t on $(\partial B, \alpha_0)$ satisfies $\rho \circ \phi^{-t} \circ \rho = \phi^t$, the involution ρ of ∂B descends to the involution $\bar{\rho}$ of $\partial B/S^1 \cong \mathbb{C}P^{n-1}$, where the S^1-action on ∂B is given by the Reeb flow.

Take a $\bar{\rho}$-invariant Morse function $\bar{f} \geq 0$ on $\partial B/S^1$ which attains the minimum precisely at a pair of two critical points away from the fixed point set of $\bar{\rho}$. We write $f \in C^\infty(\partial B, \mathbb{R})$ for the lifting of $\bar{f}$. Set $h_\epsilon{:=}1+\epsilon f$ for $\epsilon > 0$. Since h_ϵ is ρ-invariant, the starshaped domain K_{h_ϵ} is symmetric. We claim that for $\epsilon >$ sufficiently small we have $\mathfrak{R}(\partial K_{h_\epsilon}, \rho) > 1$, but this will be close to 1. We denote by $T_{\min}$ the minimal period of the Reeb flow of the standard contact sphere $(\partial B, \alpha_0)$. Recall from Bourgeois [8, Section 2.2] that for any $T > T_{\min}$ there exists $\epsilon > 0$ such that the periodic orbits of $(\partial B, \alpha_\epsilon{:=}h_\epsilon\alpha)$ of period less than T are non-degenerate and correspond to the critical points of $\bar{f}$. For a critical point $\bar{x}$ of $\bar{f}$ the corresponding periodic orbit is the S^1-fiber $\gamma_{\bar{x}}$ of the fibration $\partial B \to \partial B/S^1$ at $\bar{x}$, and its period is given by $T_{\min}h_\epsilon(x)$ for any lift $x \in \partial B$ of $\bar{x}$. The periodic orbit $\gamma_{\bar{x}}$ is symmetric if and only if $\bar{x}$ is a fixed point of $\bar{\rho}$. Now we take $T > 0$ slightly bigger than $T_{\min}$. For $\epsilon > 0$ small enough, the minimum period of non-symmetric periodic orbits is strictly smaller than the minimum period of symmetric periodic orbits. This shows that $\mathfrak{R}(\partial K_{h_\epsilon}, \rho) > 1$. It is worth noting that $\mathfrak{R}(\partial K_{h_\epsilon}, \rho)$ can be arbitrarily close to one.

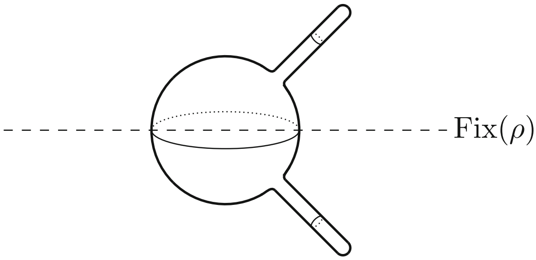


FIGURE 1. A symmetric Bordeaux-bottle-shaped domain having two necks

2.5. Bordeaux-bottle-shaped hypersurfaces of arbitrarily large symmetric ratio

Recall that the classical Bordeaux-bottle $K \subset \mathbb{R}^{2n}$ is a smooth starshaped domain obtained by gluing a thin neck, modeled on the symplectic 2-subspace $\mathbb{R}^2 \times \{0\}$, along the boundary of the unit ball B, see [18, Section 3.5]. Let ρ denote complex conjugation on $\mathbb{R}^{2n}$. For a given symplectic 2-subspace $V \subset \mathbb{R}^{2n}$ with $\rho(V) \cap V = \{0\}$, we glue the two thin necks, associated to V and $\rho(V)$ respectively, along the boundary ∂B to form a ρ-symmetric Bordeaux-bottle-shaped domain K_V with two necks. See Fig. 1. Since ρ sends one neck to the other, any periodic orbits on the thin necks are not symmetric. Moreover, symmetric periodic orbits on ∂K_V exist on ∂B, and the period of any symmetric periodic orbits of ∂K_V is uniformly bounded from below. Making the necks narrow, we obtain a symmetric starshaped domain K_V with arbitrarily large symmetric ratio. Below, we provide a detailed account of this construction.

Consider the symplectic 2-subspace V in $(\mathbb{R}^{2n}, \omega_0 = \sum_{j=1}^n \mathrm{d}x_j \wedge \mathrm{d}y_j)$ spanned by

$$\begin{aligned} v_1 &= (1, 0, 0, 1, 0, \dots, 0), \\ Jv_1 &= (0, 1, -1, 0, 0, \dots, 0), \end{aligned}$$

where J denotes the standard complex structure on $\mathbb{C}^n \cong \mathbb{R}^{2n}$. A simple computation shows that

$$\rho(V) \cap V = \{0\}. \tag{2.3}$$

Since ρ is anti-symplectic, $\rho(V)$ is a symplectic 2-subspace. We then obtain a symplectic orthogonal decomposition

$$\mathbb{R}^{2n} = V \oplus \rho(V) \oplus W,$$

where $W = (V \oplus \rho(V))^{\perp}$ denotes the symplectic complement of $V \oplus \rho(V)$. Since $V \oplus \rho(V)$ is ρ-invariant, so is W. Using the Gram–Schmidt process [24, Lemma 2.6.6], we can construct a unitary basis on $\mathbb{R}^{2n}$,

$$\{v_1, Jv_1, \dots, v_n, Jv_n\}, \tag{2.4}$$

such that

- $v_2 {:=} \rho(v_1)$ and $Jv_2 = J\rho(v_1) = -\rho(Jv_1)$,
- $\{v_3, Jv_3, \dots, v_n, Jv_n\}$ is a unitary basis for W.

We denote by

$$(\mathbf{x}', \mathbf{y}') = (x_1', y_1', \dots, x_n', y_n') \tag{2.5}$$

the symplectic coordinates on $\mathbb{R}^{2n}$ with respect to the basis (2.4). Then K_V is defined to be a ρ-symmetric smooth starshaped domain in $\mathbb{R}^{2n}$ consisting of

- a bounded piece of the neck $N_V = \{(\mathbf{x}', \mathbf{y}') \mid (x_1')^2 + (y_1')^2 \le \epsilon\}$ of V;
- a bounded piece of the neck $N_{\rho(V)} = \{(\mathbf{x}', \mathbf{y}') \mid (x_2')^2 + (y_2')^2 \le \epsilon\}$ of $\rho(V)$;
- the unit ball $B = \{(\mathbf{x}', \mathbf{y}') \mid \sum_{j=1}^n (x_j')^2 + (y_j')^2 \le 1\}$.

Here, a smoothing procedure is required as in the well-known case of a Bordeaux-bottle having one neck. We emphasize that the smoothing procedure in the standard Bordeaux-bottle is still enough for our case, since the gluing regions of ∂N_V and $\partial N_{\rho(V)}$ along ∂B are disjoint due to (2.3). Moreover, the unit ball in the coordinates (2.5) coincides with the unit ball in the standard coordinates.

We claim that the symmetric ratio $\mathfrak{R}(\partial K_V, \rho)$ can be arbitrarily large by choosing $\epsilon > 0$ small enough. Since V is chosen to be symplectic, every periodic orbit on the boundary of the neck N_V is of the form

$$\gamma(t) = \left(w_1 e^{\frac{2it}{\epsilon}}, w_2, \dots, w_n\right),$$

where the identifications (2.5) and $w_j = x_j' + iy_j'$ are used. They have small periods depending on $\epsilon > 0$. The similar holds for periodic orbits on $\partial N_{\rho(V)}$. Thanks to (2.3), any periodic orbits on the necks ∂N_V and $\partial N_{\rho(V)}$ are not symmetric under ρ. As mentioned before, symmetric periodic orbits of ∂K_V exist on the boundary ∂B, and the period of any symmetric periodic orbits of ∂K_V is uniformly bounded from below. Therefore, the claim follows.

2.6. Hypersurfaces of arbitrarily large symmetric ratio in Hamiltonian systems

Recall that a hypersurface Σ in $\mathbb{R}^{2n}$ is called of *restricted contact type* if there exists a Liouville vector field X which is defined in a neighborhood of the hypersurface and which is transverse to Σ. If Σ is of restricted contact type with the radial vector field $X = \frac{r}{2}\partial_r$, then it is starshaped. Here we provide a restricted contact type, but not starshaped hypersurface of arbitrarily large symmetric ratio.

Consider a mechanical Hamiltonian $H(q,p) = \frac{1}{2}|p|^2 + V(q)$, $(q,p) \in \mathbb{R}^2 \times \mathbb{R}^2$, where the potential V is invariant under the involution $(q_1, q_2) \mapsto (-q_1, q_2)$. It follows that H is invariant under the anti-symplectic involution $\rho(q_1, q_2, p_1, p_2) = (-q_1, q_2, p_1, -p_2)$, and hence for every $E \in \mathbb{R}$, the energy level set $H^{-1}(E)$ is ρ-invariant. We assume the following.

- There exist exactly two saddle points $(\pm a, 0)$ of V such that $V(\pm a, 0) = 0$.

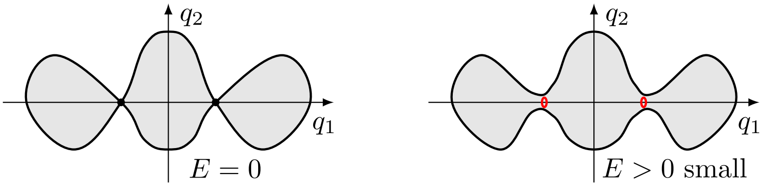


FIGURE 2. The projections of the energy levels $H^{-1}(E)$ into the position space $\mathbb{R}^2$

- For $E > 0$ small enough, $H^{-1}(-E)$ consists of three 3-spheres.

The first condition implies that the equilibriums $(\pm a, 0, 0, 0)$ of H are of saddle-center type, and the second condition implies that the energy level $H^{-1}(E)$ for $E = 0$ and for $E > 0$ small enough project into the position space $\mathbb{R}^2$ as in Fig. 2. For every $E > 0$ sufficiently small, $H^{-1}(E)$ is not starshaped, but of restricted contact type as H is of mechanical type. Since $(\pm a, 0, 0, 0)$ are of saddle-center type, in view of a well-known theorem by Lyapunov, if $E > 0$ is small enough, $H^{-1}(E)$ carries periodic orbits $\gamma_1, \gamma_2 = \rho(\gamma_1)$, called the Lyapunov orbits (red curves in Fig. 2). As $E \to 0^+$, they converge to equilibriums. Moreover, in a sufficiently small neighborhood of equilibriums, there exists no periodic orbit other than the associated Lyapunov orbit, and periodic orbits that pass this neighborhood have sufficiently large periods. This in particular implies that if $E > 0$ is small enough, then the periods of the Lyapunov orbits are extremely small, but the periods of other periodic orbits are bounded from blow by some positive constant. As the Lyapunov orbits are not ρ-symmetric, we conclude that the symmetric ratio can be chosen arbitrarily large.

3. Closed-open maps

3.1. Symplectic homology

We briefly recall the construction of symplectic homology without technical details. We refer the reader to [9, Section 2] for a detailed description. We work with Liouville domains, and prominent examples are starshaped domains in $\mathbb{R}^{2n}$ including smooth convex bodies. In this paper, we always use $\mathbb{Z}_2$-coefficients.

Let (W, λ) be a Liouville domain with a Liouville form λ. This means W is a compact smooth manifold with boundary and λ is a 1-form on W such that $d\lambda$ is symplectic and its Liouville vector field is positively transverse along the boundary. The restriction $\alpha := \lambda|_{\partial W}$ of the Liouville form defines a contact form on the boundary, and we denote the contact boundary by $(\Sigma, \alpha) := (\partial W, \lambda|_{\partial W})$. The completion $(\widehat{W}, \widehat{\lambda})$ of the Liouville domain (W, λ) is an open symplectic manifold defined by attaching (a positive part of) the symplectization $([1, \infty) \times \Sigma, r\alpha)$ to the domain (W, λ) along the boundary via the Liouville flow. Here $r \in [1, \infty)$ denotes the Liouville coordinate.

Example 3.1. The closed unit ball $B^{2n} \subset \mathbb{R}^{2n}$ with the standard symplectic form $\omega_0 = \sum_{j=1}^n \mathrm{d}x_j \wedge \mathrm{d}y_j$ is a Liouville domain with a Liouville form $\lambda_0 = \frac{1}{2}\sum_{j=1}^n (x_j \mathrm{d}y_j - y_j \mathrm{d}x_j)$. The contact type boundary is the standard contact sphere (S^{2n-1}, α_0) with $\alpha_0 = \lambda_0|_{\partial B^{2n}}$. The completion of B^{2n} recovers $\mathbb{R}^{2n}$. More generally, any starshaped domains in $\mathbb{R}^{2n}$, including smooth convex ones, fit into our setup for Floer theory.

3.1.1. Admissible Hamiltonians. We take an *admissible* time-dependent Hamiltonian $H_{S^1} : S^1 \times \widehat{W} \to \mathbb{R}$, meaning that all 1-periodic orbits of the Hamiltonian vector field $X_{H_{S^1}}$ are non-degenerate, H_{S^1} is negative and C^2-small (and Morse) in the interior of $W \subset \widehat{W}$, and H_{S^1} is linear at the end with respect to the Liouville coordinate r, independent of the time parameter $t \in S^1$. The derivative $H'_{S^1}(r)$ at the end is called the *slope* of the Hamiltonian H_{S^1}. We assume that the slope is positive and not equal to the period of a periodic Reeb orbit in the contact boundary (Σ, α). See Remark 3.4.

Remark 3.2. Our convention for Hamiltonian vector fields is that $\omega(X_H, \cdot) = \mathrm{d}H$.

Denote the set of contractible 1-periodic orbits of H_{S^1} by $\mathcal{P}(H_{S^1})$. To each 1-periodic orbit $\gamma \in \mathcal{P}(H_{S^1})$ we can associate an integer called the Conley–Zehnder index $\mathrm{CZ}(\gamma)$ by taking a capping disk of γ. We assume that $c_1(TW)$ vanishes on $\pi_2(W)$ for well-definedness of the index $\mathrm{CZ}(\gamma)$. See [9] for details on the index.

3.1.2. Chain complex. Let $J_{S^1} = \{J_t\}_{t \in S^1}$ be a time-dependent family of compatible almost complex structures on $(\widehat{W}, \widehat{\lambda})$ which is admissible in the sense of [9]. The *Floer chain group* $\mathrm{CF}_*(H_{S^1}, J_{S^1})$ for the pair (H_{S^1}, J_{S^1}) is a $\mathbb{Z}$-graded vector space over $\mathbb{Z}_2$, generated by the 1-periodic orbits of $\mathcal{P}(H_{S^1})$ and graded by the negative Conley–Zehnder index $|\gamma| = -\mathrm{CZ}(\gamma)$:

$$\mathrm{CF}_k(H_{S^1}, J_{S^1}) = \bigoplus_{\substack{\gamma \in \mathcal{P}(H_{S^1}) \\ |\gamma| = k}} \mathbb{Z}_2 \langle \gamma \rangle.$$

For two distinct 1-periodic orbits $\gamma_\pm \in \mathcal{P}(H_{S^1})$, define the moduli space of *Floer cylinders* $\mathcal{M}(\gamma_-, \gamma_+, H_{S^1}, J_{S^1})$ from γ_- to γ_+, modulo the natural $\mathbb{R}$-action, by

$$\begin{aligned} \mathcal{M}(\gamma_-, \gamma_+, H_{S^1}, J_{S^1}) = \{u : \mathbb{R} \times S^1 \to \widehat{W} \mid \lim_{s \to \pm\infty} u(s,t) = \gamma_\pm(t), \\ (\mathrm{d}u - X_{H_{S^1}} \otimes \mathrm{d}t)^{0,1} = 0\}/\mathbb{R}. \end{aligned} \tag{3.1}$$

See the left in Fig. 3.

Proposition 3.3. *Let $\gamma_- \neq \gamma_+$. For generic J_{S^1}, the moduli space $\mathcal{M}(\gamma_-, \gamma_+, H_{S^1}, J_{S^1})$ is a smooth manifold of dimension $|\gamma_-| - |\gamma_+| - 1$.*

Remark 3.4. Since H_{S^1} and J_{S^1} are admissible, Floer trajectories must lie in a compact region in $\widehat{W}$ by a maximum principle.

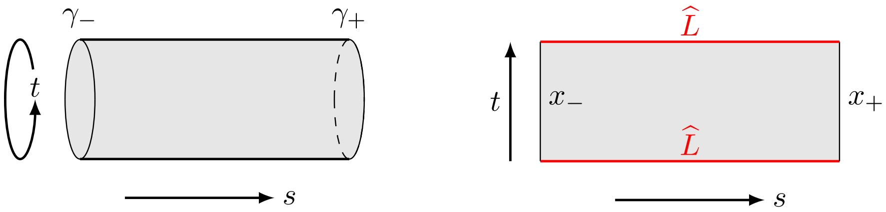


FIGURE 3. A Floer cylinder (left) and strip (right)

The differential $\partial : \mathrm{CF}_k(H_{S^1}, J_{S^1}) \to \mathrm{CF}_{k-1}(H_{S^1}, J_{S^1})$ is defined by counting rigid Floer trajectories between 1-periodic orbits as follows:

$$\partial(\gamma_-) = \sum_{\substack{\gamma_+ \in \mathcal{P}(H_{S^1}) \\ |\gamma_+| = k-1}} \#_2 \mathcal{M}(\gamma_-, \gamma_+, H_{S^1}, J_{S^1})\gamma_+. \tag{3.2}$$

The Floer–Gromov compactness and the gluing construction in Floer theory show that $\partial^2 = 0$, and hence we obtain the Floer chain complex $(\mathrm{CF}_*(H_{S^1}, J_{S^1}), \partial)$. The Floer homology $\mathrm{HF}_*(H_{S^1}, J_{S^1})$ of the pair (H_{S^1}, J_{S^1}) is defined by

$$\mathrm{HF}_*(H_{S^1}, J_{S^1}) = \mathrm{H}_*(\mathrm{CF}_*(H_{S^1}, J_{S^1}), \partial).$$

3.1.3. Symplectic homology. Standard continuation maps in Hamiltonian Floer homology define a direct system of Floer homology groups $\mathrm{HF}_*(H_{S^1}, J_{S^1})$ directed by increasing the slope τ of Hamiltonains. See e.g. [12, Section 4.4]. The *symplectic homology* of the Liouville domain (W, λ) is defined to be the direct limit

$$\mathrm{SH}_*(W, \lambda) = \varinjlim_{\tau \to \infty} \mathrm{HF}_*(H_{S^1}, J_{S^1}).$$

3.1.4. Action filtration. For an admissible Hamiltonian H_{S^1} we have the associated action functional $\mathcal{A}_{H_{S^1}} : \mathcal{L}\widehat{W} \to \mathbb{R}$ on the free loop space $\mathcal{L}\widehat{W}$ of the completion $\widehat{W}$ given by

$$\mathcal{A}_{H_{S^1}}(\gamma) = -\int_{S^1} \gamma^*\lambda - \int_0^1 H_{S^1}(t, \gamma(t))dt.$$

We call the value $\mathcal{A}_{H_{S^1}}(\gamma)$ the *action* of γ. Since Floer trajectories decrease action values, we obtain an action filtration on Floer chain complexes by collecting generators of action less than $a \in \mathbb{R}$

$$\mathrm{CF}_k^a(H_{S^1}, J_{S^1}) = \bigoplus_{\substack{\gamma \in \mathcal{P}(H_{S^1}) \\ |\gamma| = k \\ \mathcal{A}_{H_{S^1}}(\gamma) < a}} \mathbb{Z}_2\langle\gamma\rangle.$$

The corresponding filtered Floer homology is denoted by $\mathrm{HF}_*^a(H_{S^1}, J_{S^1})$, and taking the direct limit we define the *filtered symplectic homology*

$$\mathrm{SH}_*^a(W, \lambda) = \varinjlim_{\tau \to \infty} \mathrm{HF}_*^a(H_{S^1}, J_{S^1}).$$

3.1.5. Tautological exact sequences. Let $a < b$. The action filtration on the chain complex $\mathrm{CF}_*(H_{S^1}, J_{S^1})$ induces the following natural short exact sequence of chain complexes:

$$0 \to \mathrm{CF}^a_*(H_{S^1}, J_{S^1}) \to \mathrm{CF}^b_*(H_{S^1}, J_{S^1}) \to \mathrm{CF}^{[a,b)}_*(H_{S^1}, J_{S^1}) \to 0$$

where $\mathrm{CF}^{[a,b)}_*(H_{S^1}, J_{S^1})$ is the chain complex defined to be the quotient

$$\mathrm{CF}^{[a,b)}_*(H_{S^1}, J_{S^1}) = \mathrm{CF}^b_*(H_{S^1}, J_{S^1}) / \mathrm{CF}^a_*(H_{S^1}, J_{S^1})$$

with the induced differential. We obtain, passing to the direct limit, an associated long exact sequence in symplectic homology

$$\to \mathrm{SH}^a_k(W) \to \mathrm{SH}^b_k(W) \to \mathrm{SH}^{[a,b)}_k(W) \to \mathrm{SH}^a_{k-1}(W) \to . \tag{3.3}$$

In particular, due to the assumption that H_{S^1} is C^2-small and Morse on W, if $\epsilon > 0$ sufficiently small, we have a canonical identification

$$\mathrm{SH}^\epsilon_k(W) \cong \mathrm{H}_{k+n}(W, \partial W). \tag{3.4}$$

We then have the *(filtered) tautological exact sequence* in symplectic homology

$$\to \mathrm{H}_{k+n}(W, \partial W) \to \mathrm{SH}^a_k(W) \to \mathrm{SH}^{[\epsilon,a)}_k(W) \to \mathrm{H}_{k+n-1}(W, \partial W) \to .$$

For each $a > 0$, we shall denote the map from $\mathrm{H}_{k+n}(W, \partial W)$ to $\mathrm{SH}^a_k(W)$ in the sequence by

$$j^a : \mathrm{H}_{k+n}(W, \partial W) \to \mathrm{SH}^a_k(W).$$

3.2. Wrapped Floer homology

We shortly review a construction of wrapped Floer homology which is an open string analogue of symplectic homology. We refer to [3, 21] for details.

3.2.1. Chain complex. Let L be an *admissible* Lagrangian in a Liouville domain (W, λ), meaning that L is a connected and exact Lagrangian which intersects the contact boundary (Σ, α) in a Legendrian $\mathcal{L} := \partial L = L \cap \Sigma$ and the Liouville vector field is tangent to TL near the boundary. By attaching $[1, \infty) \times \mathcal{L}$ to L along the Legendrian boundary $\mathcal{L}$ we have a completed exact Lagrangian $\widehat{L}$ in the completion $(\widehat{W}, \widehat{\lambda})$. Roughly speaking, the wrapped Floer homology $\mathrm{HW}_*(L)$ is a version of Lagrangian Floer homology of $\widehat{L}$ in $\widehat{W}$.

A time-independent Hamiltonian $H : \widehat{W} \to \mathbb{R}$ is called *admissible* if every Hamiltonian 1-chord relative to $\widehat{L}$ is non-degenerate, H is negative and C^2-small in the interior of $W \subset \widehat{W}$, and H is linear at the end with respect to $r \in [1, \infty)$. We assume that the slope τ of H is positive and is not equal to the length of a Reeb chord in $(\Sigma, \alpha, \mathcal{L})$. Recall that a *Reeb chord* in $(\Sigma, \alpha, \mathcal{L})$ is an orbit $x : [0, T] \to \Sigma$ of the Reeb flow on (Σ, α) with $x(0), x(T) \in \mathcal{L}$. We call T the *length* of the Reeb chord x.

Denote the set of contractible, as an element of $\pi_1(\widehat{W}, \widehat{L})$, Hamiltonian 1-chords by $\mathcal{P}_L(H)$. We associate the index $|x| = -\mu(x) - \frac{n}{2} \in \mathbb{Z}$ for each nondegenerate contractible 1-chord in $\mathcal{P}_L(H)$, where $\mu(x)$ is the Maslov index defined in [21, Definition 2.3]. Assume that $c_1(TW) = 0$ and $\pi_1(L) = 0$ for well-definedness of $\mu(x)$.

Example 3.5. Consider complex conjugation ρ_0 on the closed ball B^{2n}. Its fixed point set $L = \mathrm{Fix}(\rho_0) = B^{2n} \cap \mathbb{R}^n$, called a real Lagrangian, defines an admissible Lagrangian in (B^{2n}, λ_0). More generally, let W be a starshaped domain in $\mathbb{R}^{2n}$ invariant under an exact anti-symplectic involution ρ of $(\mathbb{R}^{2n}, \lambda_0)$ i.e. $\rho^*\lambda_0 = -\lambda_0$. Then the fixed point set $L := \mathrm{Fix}(\rho|_W)$ defines an admissible Lagrangian:

- From the classical Smith theory, we have $\dim \mathrm{H}_*(W;\mathbb{Z}_2) \geq \dim \mathrm{H}_*(L;\mathbb{Z}_2)$ and $\chi(W) = \chi(L) \mod 2$. It follows that L is nonempty and connected.
- Since the Liouville flow on $(\mathbb{R}^{2n}, \lambda_0)$ commutes with ρ and flows radially from the origin, the real Lagrangian L intersects the boundary ∂W.
- As in [22, Lemma 3.1], L is an exact Lagrangian, the intersection $L \cap \partial W$ is a Legendrian, and the Liouville vector field is tangent to TL near the boundary.

Note also that $c_1(TW) = 0$ for any starshaped domain W whereas $\pi_1(L)$ is not necessarily a trivial group. If the anti-symplectic involution ρ is *linear* e.g. complex conjugation, then L is diffeomorphic to the closed ball B^n and hence $\pi_1(L) = 0$ in this case.

Let $J = \{J_t\}_{t\in[0,1]}$ be an admissible time-dependent family of compatible almost complex structures. For two distinct 1-chords $x_\pm \in \mathcal{P}_L(H)$, the moduli space $\mathcal{M}(x_-, x_+, H, J)$ of *Floer strips* from x_- to x_+, modulo the natural $\mathbb{R}$-action, is defined by

$$\begin{aligned}\mathcal{M}(x_-, x_+, H, J) = \{u : \mathbb{R}\times[0,1] \to \widehat{W} \mid \lim_{s\to\pm\infty} u(s,t) &= x_\pm(t),\\ (\mathrm{d}u - X_H \otimes \mathrm{d}t)^{0,1} &= 0, \\ u(s,0), u(s,1) &\in \widehat{L}\}/\mathbb{R}.\end{aligned} \tag{3.5}$$

See the right in Fig. 3. For generic J, the moduli space $\mathcal{M}(x_-, x_+, H, J)$ is a smooth manifold of dimension $|x_-| - |x_+| - 1$.

The Floer chain complex for the pair (H, J) is defined by

$$\mathrm{CF}_k(H,J) = \bigoplus_{\substack{x\in\mathcal{P}_L(H)\\ |x|=k}} \mathbb{Z}_2\langle x\rangle$$

equipped with the differential $\partial : \mathrm{CF}_k(H,J) \to \mathrm{CF}_{k-1}(H,J)$ given by

$$\partial(x_-) = \sum_{\substack{x_+\in\mathcal{P}_L(H)\\ |x_+|=k-1}} \#_2\mathcal{M}(x_-, x_+, H, J)x_+. \tag{3.6}$$

We obtain the Floer homology group $\mathrm{HF}_*(H,J)$ as the homology of the chain complex $(\mathrm{CF}_*(H,J), \partial)$, and by taking the direct limit as in the symplectic homology, we define the *wrapped Floer homology* of the Lagrangian L in (W, λ) by

$$\mathrm{HW}_*(L) = \varinjlim_{\tau\to\infty} \mathrm{HF}_*(H,J).$$

3.2.2. Filtered wrapped Floer homology. Wrapped Floer homology shares many analogous properties with symplectic homology. In particular, we have a natural action filtration and tautological exact sequences. The action filtration on $\mathrm{HW}_*(L)$ is given by the action functional $\mathcal{A}_H : \mathcal{L}_L\widehat{W} \to \mathbb{R}$ on the free path space $\mathcal{L}_L\widehat{W}$ of the completion $\widehat{W}$ relative to $\widehat{L}$, defined by

$$\mathcal{A}_H(x) = -\int_{[0,1]} x^*\lambda - \int_0^1 H(x(t))\mathrm{d}t + f_L(x(1)) - f_L(x(0)).$$

Here $f_L \in C^\infty(\widehat{L}, \mathbb{R})$ is a primitive of the form $\widehat{\lambda}|_{\widehat{L}}$. For $a \in \mathbb{R}$, we denote the filtered chain complex by $\mathrm{CF}^a_*(H, J)$ and the filtered wrapped Floer homology by $\mathrm{HW}^a_*(L)$. For $a < b$ a long exact sequence analogous to (3.3) is written as

$$\to \mathrm{HW}^a_k(L) \to \mathrm{HW}^b_k(L) \to \mathrm{HW}^{[a,b)}_k(L) \to \mathrm{HW}^a_{k-1}(L) \to . \tag{3.7}$$

In particular, for $\epsilon > 0$ sufficiently small so that

$$\mathrm{HW}^\epsilon_k(L) \cong \mathrm{H}_{k+n}(L, \partial L) \tag{3.8}$$

we have the tautological long exact sequence in wrapped Floer homology

$$\to \mathrm{H}_{k+n}(L, \partial L) \to \mathrm{HW}^a_k(L) \to \mathrm{HW}^{[\epsilon,a)}_k(L) \to \mathrm{H}_{k+n-1}(L, \partial L) \to .$$

For each $a > 0$, as in symplectic homology, we denote the map from H_{k+n} $(L, \partial L)$ to $\mathrm{HW}^a_k(L)$ in the sequence by

$$j^a : \mathrm{H}_{k+n}(L, \partial L) \to \mathrm{HW}^a_k(L).$$

3.3. Closed-open maps

Closed-open maps are natural homomorphisms from symplectic homology to wrapped Floer homology. In Sect. 4 we use them to relate symplectic capacities from the two Floer homologies. In this section we shall briefly outline a construction of closed-open maps based on [2,16,29]. See also [5].

3.3.1. Floer data. Closed-open maps are defined by counting curves in $\widehat{W}$ which we call *Floer chimneys*. The domain $\mathcal{T}$ of Floer chimneys is given by the closed unit disk $\mathbb{D}$ with an interior puncture and a boundary puncture,

$$\mathcal{T} = (\mathbb{D} \setminus \{0, 1\}, i)$$

where i is the standard complex structure. See the left in Fig. 4. We equip $\mathcal{T}$ a negative cylindrical end $\varepsilon_0 : (-\infty, 0] \times S^1 \to \mathcal{T}$ near 0 and a positive strip-like end $\varepsilon_1 : [0, \infty) \times [0, 1] \to \mathcal{T}$ near 1.

A Floer data $(H_\mathcal{T}, J_\mathcal{T})$ for chimneys is given as follows. Let H_{S^1} : $S^1 \times \widehat{W} \to \mathbb{R}$ and $H : \widehat{W} \to \mathbb{R}$ be admissible Hamiltonians for symplectic homology and wrapped Floer homology, respectively, of the same slope τ. A Hamiltonian $H_\mathcal{T} : \mathcal{T} \times \widehat{W} \to \mathbb{R}$ is called *admissible* if

- $H_\mathcal{T}(\varepsilon_0(s, t), w) = H_{S^1}(t, w)$;
- $H_\mathcal{T}(\varepsilon_1(s, t), w) = H(w)$;
- for each $z \in \mathcal{T}$, the Hamiltonian $H_\mathcal{T}(z, \cdot) : \widehat{W} \to \mathbb{R}$ is admissible with slope τ and is independent of z at the end. We call τ the *slope* of $H_\mathcal{T}$.

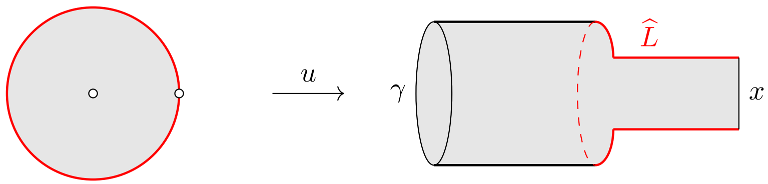


FIGURE 4. A Floer chimney

For admissible almost complex structures J_{S_1} and J as in Sect. 3.1.2 and 3.2.1, we take an admissible $\mathcal{T}$-family of compatible almost complex structures $J_{\mathcal{T}}$ given in an analogous way to the Hamiltonian case so that $J_{\mathcal{T}} = J_{S^1}$ and $J_{\mathcal{T}} = J$ near the punctures.

3.3.2. Floer chimneys. To write the Floer equation for chimneys, we fix a 1-form β on $\mathcal{T}$ with the following properties.

- $\mathrm{d}\beta \leq 0$ with respect to the fixed volume form on $\mathcal{T}$.
- $\beta|_{\partial\mathcal{T}} = 0$, and $\beta|_{\nu(\partial\mathcal{T})} = 0$ where $\nu(\partial\mathcal{T})$ is a neighborhood of $\partial\mathcal{T}$.
- With respect to the coordinate charts ε_0 and ε_1, we set $\beta = \mathrm{d}t$.

Remark 3.6. The conditions on β guarantee that Floer chimneys stay in a compact region in $\widehat{W}$. One can show this using a convexity argument in [3, Lemma 7.2], which replaces the maximum principle.

Take $\gamma \in \mathcal{P}(H_{S^1})$ and $x \in \mathcal{P}_L(H)$. A *Floer chimney from γ to x* is a map $u : \mathcal{T} \to \widehat{W}$ satisfying the following conditions, see Fig. 4.

- (Floer equation) u is a solution of the equation
$$(\mathrm{d}u - X_{H_{\mathcal{T}}} \otimes \beta)^{0,1} = 0.$$
- (Asymptotic condition)
$$\lim_{s\to-\infty} u(\varepsilon_0(s,t)) = \gamma(t), \quad \lim_{s\to\infty} u(\varepsilon_1(s,t)) = x(t).$$
- (Lagrangian boundary) $u(\partial\mathcal{T}) \subset \widehat{L}$.

We denote the moduli space of Floer chimneys from γ to x by
$$\mathcal{M}(\gamma, x, H_{\mathcal{T}}, J_{\mathcal{T}}) = \{u : \mathcal{T} \to \widehat{W} \mid u \text{ is a Floer chimney from } \gamma \text{ to } x\}. \quad (3.9)$$

Proposition 3.7. (See [2, Lemma 5.3]) *For generic $J_{\mathcal{T}}$, the moduli space $\mathcal{M}(\gamma, x, H_{\mathcal{T}}, J_{\mathcal{T}})$ is a smooth manifold of dimension*
$$\dim \mathcal{M}(\gamma, x, H_{\mathcal{T}}, J_{\mathcal{T}}) = |\gamma| - |x| - n.$$

If $|\gamma| = |x| + n$, the moduli space is compact and zero dimensional. This allows us to define the map
$$\mathcal{CO} : \mathrm{CF}_k(H_{S^1}, J_{S^1}) \to \mathrm{CF}_{k-n}(H, J)$$
by counting rigid Floer chimneys
$$\mathcal{CO}(\gamma) = \sum_{\substack{x\in\mathcal{P}_L(H)\\ |x|=k-n}} \#_2\mathcal{M}(\gamma, x, H_{\mathcal{T}}, J_{\mathcal{T}})x. \quad (3.10)$$

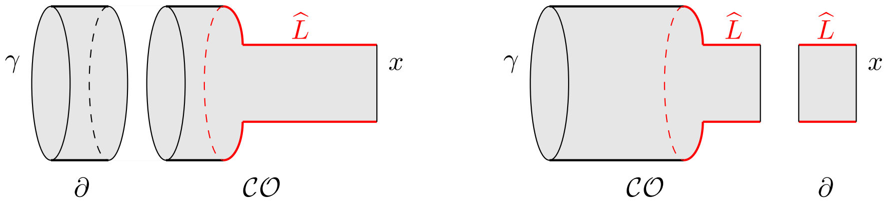


FIGURE 5. Broken Floer chimneys showing $\partial \circ \mathcal{CO} = \mathcal{CO} \circ \partial$

The codimension 1 boundary strata of the moduli space of Floer chimneys, described in [2, Lemma 5.3], shows that the map $\mathcal{CO} : \mathrm{CF}_k(H_{S^1}, J_{S^1}) \to \mathrm{CF}_{k-n}(H, J)$ is a chain map; see Fig. 5. We have the induced homomorphism on homology groups

$$\mathcal{CO} : \mathrm{HF}_k(H_{S^1}, J_{S^1}) \to \mathrm{HF}_{k-n}(H, J).$$

Taking homotopies of admissible Hamiltonians $H_{\mathcal{T}}$, a standard argument in Floer theory in [12, Section 4.4] allows us to pass it to the direct limit via continuation maps

$$\mathcal{CO} : \mathrm{SH}_k(W) \to \mathrm{HW}_{k-n}(L).$$

We call this map the *closed-open map* from symplectic homology to wrapped Floer homology.

Remark 3.8. In [29, Theorem 8.2], it is shown that $\mathcal{CO} : \mathrm{SH}_*(W) \to \mathrm{HW}_*(L)$ is a unital ring homomorphism with respect to the standard ring structures described e.g. in [28].

3.3.3. Filtered closed-open maps. Closed-open maps respect the action filtrations. To see this, one introduces the *topological energy* of Floer chimneys as follows:

$$E(u) := \int_{\mathcal{T}} u^*\mathrm{d}\lambda - u^*\mathrm{d}H_{\mathcal{T}} \wedge \beta - u^* H_{\mathcal{T}} \mathrm{d}\beta$$

where $u \in \mathcal{M}(\gamma, x, H_{\mathcal{T}}, J_{\mathcal{T}})$ as in Sect. 3.3.2. It is observed in [2, Appendix B] that $E(u) \geq 0$, and a direct computation shows that

$$E(u) = \mathcal{A}_{H_{S^1}}(\gamma) - \mathcal{A}_H(x).$$

In particular, Floer chimneys decrease action values. For each $a \in \mathbb{R}$, we have *filtered closed-open maps*

$$\mathcal{CO}^a : \mathrm{SH}_k^a(W) \to \mathrm{HW}_{k-n}^a(L).$$

The filtered closed-open maps are compatible with the tautological exact sequences (3.3) and (3.7) in the following sense. This is also observed in [7, Section 5.2.1] and essentially follows from [5, Theorem 1.5].

Theorem 3.9. *The closed-open map*

$$\mathcal{CO}^a : \mathrm{SH}_k^a(W) \to \mathrm{HW}_{k-n}^a(L)$$

for $a > 0$ fits into the following commutative diagram:

$$\begin{tikzcd} \mathrm{H}_{k+n}(W, \partial W) \arrow[r, "j^a"] \arrow[d, "i_!"] & \mathrm{SH}_k^a(W) \arrow[d, "\mathcal{CO}^a"] \\ \mathrm{H}_k(L, \partial L) \arrow[r, "j^a"] & \mathrm{HW}_{k-n}^a(L) \end{tikzcd} \tag{3.11}$$

where the left vertical $i_! : \mathrm{H}_{k+n}(W, \partial W) \to \mathrm{H}_k(L, \partial L)$ is the natural map induced by the inclusion $i : L \to W$.

Proof. Let $\epsilon > 0$ be a sufficiently small number such that (3.4) and (3.8) hold. The commutative diagram (on the right) in [5, Theorem 1.5] together with the compatibility of closed-open maps with the action filtrations yields the following commutative diagram.

$$\begin{tikzcd} \mathrm{H}_{k+n}(W, \partial W) \arrow[r, "\psi"] \arrow[d, "i_!"] & \mathrm{SH}_k^\epsilon(W) \arrow[r, "\iota_a"] \arrow[d, "\mathcal{CO}^\epsilon"] & \mathrm{SH}_k^a(W) \arrow[d, "\mathcal{CO}^a"] \\ \mathrm{H}_k(L, \partial L) \arrow[r, "\psi"] & \mathrm{HW}_{k-n}^\epsilon(L) \arrow[r, "\iota_a"] & \mathrm{HW}_{k-n}^a(L) \end{tikzcd} \tag{3.12}$$

Here, the maps ψ, constructed in [28, Section 15], are the analogues of the (relative) Piunikhin–Salamon–Schwarz isomorphism [26], and the maps ι_a are the natural inclusions from the respective tautological exact sequences (3.3) and (3.7). Indeed, the exactness of L replaces the monotonicity assumption in [5, Theorem 1.5] from which we obtain the commutativity of the left square in (3.12) modulo conventional differences between Floer homology and cohomology. The commutativity of the right square in (3.12) is an immediate consequence of the fact that the closed-open map $\mathcal{CO}$ respects the action filtration on $\mathrm{SH}_*(W)$ and $\mathrm{HW}_*(L)$. As in [28, Lemma 15.1], we know that $j^a = \iota_a \circ \psi$, and the commutative diagram (3.11) in the assertion therefore follows from (3.12). □

3.3.4. Without absolute grading. Even though we have worked with the absolute $\mathbb{Z}$-grading on $\mathrm{SH}_*(W)$ and $\mathrm{HW}_*(L)$ for the sake of completeness, the Floer homologies and the filtered closed-open maps with Theorem 3.9 readily work regardless of the grading. The discussions in Sect. 4 do not require the Floer homologies and the closed-open maps to be graded, and the topological assumptions $c_1(TW) = 0$ and $\pi_1(L) = 0$ are therefore not necessary for our applications; see Example 3.5.

In this case, as a fairly standard way in Floer theory, we take the zero-dimensional component of the moduli spaces (3.1), (3.5), (3.9) of Floer solutions instead of fixing the difference of the absolute gradings of asymptotes; the Fredholm index of underlying Fredholm problems determines the local dimension of the corresponding moduli spaces. Then we define the differentials (3.2), (3.6) and the chain map (3.10) on ungraded Floer chain groups by counting elements of the zero-dimensional component of the respective moduli spaces. The analysis on Floer solutions and the action filtrations on Floer chain complexes are independent of the absolute grading. We therefore obtain the ungraded filtered closed-open map $\mathcal{CO}^a : \mathrm{SH}_*^a(W) \to \mathrm{HW}_*^a(L)$ with the commutative diagram:

$$\begin{tikzcd} \mathrm{H}_*(W, \partial W) \arrow[r, "j^a"] \arrow[d, "i_!"] & \mathrm{SH}_*^a(W) \arrow[d, "\mathcal{CO}^a"] \\ \mathrm{H}_*(L, \partial L) \arrow[r, "j^a"] & \mathrm{HW}_*^a(L) \end{tikzcd}$$

4. Floer homology capacities

4.1. SH capacity

Let (W, λ) be a Liouville domain as in Sect. 3.1. We define the *symplectic homology capacity* or shortly the SH *capacity* $c_{\mathrm{SH}}(W, \lambda)$ of the domain (W, λ) by

$$c_{\mathrm{SH}}(W) = c_{\mathrm{SH}}(W, \lambda) = \inf\{a > 0 \mid j^a[W, \partial W] = 0\} \in [0, \infty]$$

where the map $j^a : \mathrm{H}_*(W, \partial W) \to \mathrm{SH}_*^a(W)$ is constructed in Sect. 3.1.5. If $j^a[W, \partial W] \neq 0$ for all $a > 0$, then we conventionally put $c_{\mathrm{SH}}(W) = \infty$.

Proposition 4.1. *The* SH *capacity satisfies the following properties.*

(1) (Conformality) For a positive real number r, we have

$$c_{\mathrm{SH}}(W, r\lambda) = rc_{\mathrm{SH}}(W, \lambda).$$

(2) (Monotonicity) For a generalized Liouville embedding $(W_1, \lambda_1) \hookrightarrow (W_2, \lambda_2)$, we have

$$c_{\mathrm{SH}}(W_1, \lambda_1) \leq c_{\mathrm{SH}}(W_2, \lambda_2).$$

(3) (Spectrality) If $c_{\mathrm{SH}}(W) < \infty$, there exists a periodic Reeb orbit γ on the contact boundary (Σ, α) such that

$$c_{\mathrm{SH}}(W) = \ell(\gamma)$$

where $\ell(\gamma)$ denotes the period of γ.

Remark 4.2. A symplectic embedding $\varphi : (W_1, \lambda_1) \hookrightarrow (W_2, \lambda_2)$ is called a *generalized Liouville embedding* if $(\varphi^*\lambda_2 - \lambda_1)|_{\partial W_1} = 0$ in $\mathrm{H}^1(\partial W_1)$. In particular, if W_1 and W_2 are both starshaped domains in $\mathbb{R}^{2n}$, every symplectic embedding is a generalized Liouville embedding since $\mathrm{H}^1(S^{2n-1}) = 0$ for $n \geq 2$.

Remark 4.3. The SH capacity $c_{\mathrm{SH}}(W)$ is finite if and only if $\mathrm{SH}_*(W) = 0$.

Proof of Proposition 4.1. For smooth convex domains in $\mathbb{R}^{2n}$, the above properties are presented e.g. in [19, Section 2.4]. For general Liouville domains, the monotonicity comes from the existence of a natural homomorphism $\mathrm{SH}_*^a(W_2) \to \mathrm{SH}_*^a(W_1)$, called a transfer map, in symplectic homology for generalized Liouville embeddings as in [17, Theorem 1.24]. The spectrality follows from essentially the same argument as in [17, Lemma 4.2], using the relationship between action values of Hamiltonian 1-orbits of admissible Hamiltonians and Reeb orbits on the contact boundary; see [17, Remark 5.6].

4.2. HW capacity

We can define an open string analogue of the SH capacity using wrapped Floer homology. Let L be an admissible Lagrangian in a Liouville domain (W, λ). Recall that L is assumed to be connected; Sect. 3.2.1. The *wrapped Floer homology capacity* or shortly HW *capacity* is defined as

$$c_{\mathrm{HW}}(W) = c_{\mathrm{HW}}(W, \lambda, L) = \inf\{a > 0 \mid j^a[L, \partial L] = 0\} \in [0, \infty]$$

where the map $j^a : \mathrm{H}_*(L, \partial L) \to \mathrm{HW}^a_*(L)$ is defined in Sect. 3.2.2. We set $c_{\mathrm{HW}}(W) = \infty$ if $j^a[L, \partial L] \neq 0$ for all $a > 0$. The following is completely analogous to that for the SH capacity in Proposition 4.1; we omit its proof.

Proposition 4.4. *The* HW *capacity satisfies the following properties.*

(1) (Conformality) For a positive real number r, we have

$$c_{\mathrm{HW}}(W, r\lambda, L) = r c_{\mathrm{HW}}(W, \lambda, L).$$

(2) (Monotonicity) For a generalized Liouville embedding $\varphi : (W_1, \lambda_1) \to (W_2, \lambda_2)$ with $\varphi(L_1) \subset L_2$, we have

$$c_{\mathrm{HW}}(W_1) \leq c_{\mathrm{HW}}(W_2).$$

(3) (Spectrality) If $c_{\mathrm{HW}}(W) < \infty$, there exists a Reeb chord x on the contact boundary $(\Sigma, \alpha, \mathcal{L})$ such that

$$c_{\mathrm{HW}}(W) = \ell(x)$$

where $\ell(x)$ denotes the length of x.

Remark 4.5. The SH capacity is also known as the *Floer–Hofer–Wysocki capacity* defined in [14], and the HW capacity is referred to as *Lagrangian Floer–Hofer–Wysocki capacity* in [7].

Remark 4.6. The HW capacity $c_{\mathrm{HW}}(W, \lambda, L)$ is finite if and only if $\mathrm{HW}_*(L) = 0$, which is in particular the case when $\mathrm{SH}_*(W) = 0$, see [28, Theorem 10.6].

4.3. Proof of Theorem 1.3

In this section, we give a proof of the estimate (1.3).

Let K be a smooth compact convex domain in $\mathbb{R}^{2n}$ which is invariant under an anti-symplectic involution ρ of $(\mathbb{R}^{2n}, \mathrm{d}\lambda_0)$, and the real Lagrangian $\mathrm{Fix}(\rho)$ intersects the boundary ∂K. To apply our Floer setup, we choose a Liouville form λ on K with $\mathrm{d}\lambda = \mathrm{d}\lambda_0$ such that ρ is an *exact* anti-symplectic involution with respect to λ and the associated Liouville vector field is positively transverse along the boundary. For example one takes the average $\lambda := \frac{1}{2}(\lambda_0 - \rho^*\lambda_0)$:

Lemma 4.7. *Let (W, λ_0) be a Liouville domain with an anti-symplectic involution $\rho : W \to W$. Then $\lambda := \frac{1}{2}(\lambda_0 - \rho^*\lambda_0)$ satisfies that $\mathrm{d}\lambda = \mathrm{d}\lambda_0$, $\rho^*\lambda = -\lambda$, and the corresponding Liouville vector field X_λ of λ is positively transverse along the boundary ∂W.*

Proof. It is immediate to see that $\mathrm{d}\lambda = \mathrm{d}\lambda_0$ and $\rho^*\lambda = -\lambda$. We claim that the Liouville vector field X_λ of λ, defined by $\mathrm{d}\lambda_0(X_\lambda, \cdot) = \lambda$, is positively transverse along the boundary ∂W. Observe that $X_{\rho^*\lambda_0} = -\rho^* X_{\lambda_0}$. Indeed, for any vector $Y \in TW$,

$$\begin{aligned}\mathrm{d}\lambda_0(\rho^* X_{\lambda_0}, Y) &= \rho^*\mathrm{d}\lambda_0(X_{\lambda_0}\circ\rho, \rho^*(Y\circ\rho))\\ &= -\mathrm{d}\lambda_0(X_{\lambda_0}\circ\rho, \rho^*(Y\circ\rho))\\ &= -\lambda_0(\rho^*(Y\circ\rho))\\ &= -\rho^*\lambda_0(Y).\end{aligned}$$

Since the diffeomorphism $\rho : W \to W$ preserves the boundary ∂W and the interior of W, respectively, the push-forward of any outward normal vector along the boundary under ρ is again an outward normal vector. From this fact, we deduce that the pull-back $\rho^* X_{\lambda_0}$ is positively transverse along the boundary. Therefore the convex sum

$$X_\lambda = \frac{1}{2}(X_{\lambda_0} - X_{\rho^*\lambda_0}) = \frac{1}{2}(X_{\lambda_0} + \rho^* X_{\lambda_0})$$

is positively transverse along ∂W as well. □

Note that the systoles $\ell_{\min}(\partial K)$ and $\ell^{\mathrm{sym}}_{\min}(\partial K, \rho)$ defined in (1.2) do not depend on the choice of the Liouville form λ as explained in Remark 1.2, and we can work with λ instead of λ_0. The triple (K, λ, ρ) now fits our Floer setup as in Example 3.5. Note that we do not assume $\pi_1(L) = 0$; see Sect. 3.3.4.

Abbreviate $\alpha = \lambda|_{\partial K}$ and denote the restriction of ρ to ∂K again by the same letter. First we relate the above Floer homology capacities with (symmetric) systoles. The following is a non-trivial fact relating the systole $\ell_{\min}(\partial K)$ with the SH capacity, which is recently proved in [1] and [19].

Theorem 4.8. (Abbondandolo–Kang, Irie) *Let K be a smooth convex body in $\mathbb{R}^{2n}$. Then the* SH *capacity of K coincides with the systole of the contact boundary $(\partial K, \alpha)$*

$$c_{\mathrm{SH}}(K) = \ell_{\min}(\partial K).$$

Remark 4.9. The inequality $c_{\mathrm{SH}}(K) \geq \ell_{\min}(\partial K)$ is obvious from the spectrality of $c_{\mathrm{SH}}(K)$ in Proposition 4.1. There is a starshaped and non-convex K for which the strict inequality $c_{\mathrm{SH}}(K) > \ell_{\min}(\partial K)$ holds. See for example [18, Section 3.5].

Since $\mathrm{SH}_*(K) = 0$, it follows from Remark 4.6 that $\mathrm{HW}_*(L) = 0$ as well. By the spectrality of c_{HW} in Proposition 4.4, there exists a symmetric periodic orbit on $(\partial K, \rho)$. In view of the one-to-one correspondence between symmetric periodic orbits and pairs of Reeb chords on the symmetric convex hypersurface ∂K, the spectrality of c_{HW} yields the following comparison.

Proposition 4.10. *The* HW *capacity of (K, ρ) and the symmetric systole of $(\partial K, \rho)$ satisfy*

$$\ell^{\mathrm{sym}}_{\min}(\partial K, \rho) \leq 2c_{\mathrm{HW}}(K, \rho).$$

Remark 4.11. It should be possible to establish a real analogue of Theorem 4.8 for symmetric convex hypersurfaces, asserting that $2c_{\mathrm{HW}}(K,\rho) = \ell^{\mathrm{sym}}_{\min}(\partial K,\rho)$.

Closed-open maps give the following relationship between the SH capacity and the HW capacity, which was also observed in [7, (i) in Theorem 1.5]. We state it for general Liouville domains:

Proposition 4.12. *For an admissible Lagrangian L in a Liouville domain (W,λ),*

$$c_{\mathrm{HW}}(W) \leq c_{\mathrm{SH}}(W).$$

Proof. This is a direct consequence of Theorem 3.9. If $j^a[W,\partial W] = 0$ in $\mathrm{SH}^a_*(W)$, it follows from the commutative diagram (3.11) that

$$0 = (\mathcal{CO}^a \circ j^a)[W,\partial W] = (j^a \circ i_!)[W,\partial W] = j^a[L,\partial L]$$

where the last equality holds because the natural map $i_! : \mathrm{H}_*(W,\partial W) \to \mathrm{H}_*(L,\partial L)$ sends the fundamental class $[W,\partial W]$ to the fundamental class $[L,\partial L]$. Therefore we have $a \geq c_{\mathrm{HW}}(W)$, and consequently we conclude $c_{\mathrm{HW}}(W) \leq c_{\mathrm{SH}}(W)$. $\square$

We now obtain the desired estimate.

Proof of Theorem 1.3. Theorem 4.8 and Proposition 4.10 tell us that

$$1 \leq \mathfrak{R}(\partial K,\rho) = \frac{\ell^{\mathrm{sym}}_{\min}(\partial K,\rho)}{\ell_{\min}(\partial K)} \leq \frac{2c_{\mathrm{HW}}(K,\rho)}{c_{\mathrm{SH}}(K)}.$$

An application of Proposition 4.12 to (K,α,ρ) provides

$$\frac{2c_{\mathrm{HW}}(K,\rho)}{c_{\mathrm{SH}}(K)} \leq 2,$$

finishing the proof. $\square$

5. Real symplectic capacities

Let (M,ω,ρ) be a real symplectic manifold, meaning that a symplectic manifold (M,ω) is equipped with an anti-symplectic involution ρ, i.e. $\rho^*\omega = -\omega$. We always assume that $\mathrm{Fix}(\rho) \neq \emptyset$ so that it is a Lagrangian submanifold of M. A *real symplectic embedding* $\Psi\colon (M_1,\omega_1,\rho_1) \to (M_2,\omega_2,\rho_2)$ between two real symplectic manifolds is an embedding of M_1 into M_2 such that $\Psi^*\omega_2 = \omega_1$ and $\Psi^*\rho_2 = \rho_1$.

Definition 5.1. A *real symplectic capacity* is a function c which assigns to a real symplectic manifold (M,ω,ρ) a number $c(M,\omega,\rho) \in [0,+\infty]$ having the following properties:

- (Monotonicity) If real symplectic manifolds (M_1,ω_1,ρ_1) and (M_2,ω_2,ρ_2) have the same dimension, and if there exists a real symplectic embedding $\Psi\colon (M_1,\omega_1,\rho_1) \to (M_2,\omega_2,\rho_2)$, then we have $c(M_1,\omega_1,\rho_1) \leq c(M_2,\omega_2,\rho_2)$;
- (Conformality) $c(M,r\omega,\rho) = rc(M,\omega,\rho)$ for all $r > 0$;

- (Nontriviality) $0 < c(B^{2n}(1), \omega_0, \rho_0)$ and $c(Z^{2n}(1), \omega_0, \rho_0) < \infty$, where $B^{2n}(1) = \{z \in \mathbb{C}^n \mid ||z||^2 < 1\}$ and $Z^{2n}(1) = \{z \in \mathbb{C}^n \mid |z_1|^2 < 1\}$. Here $\omega_0 = \mathrm{d}\lambda_0$ denotes the standard symplectic form on $\mathbb{R}^{2n}$, and ρ_0 denotes complex conjugation.

A real symplectic capacity c is said to be *normalized* if

$$c(B^{2n}(1), \omega_0, \rho_0) = c(Z^{2n}(1), \omega_0, \rho_0) = 1.$$

Remark 5.2. The notion of real symplectic capacities was first introduced by Liu and Wang [23], where the authors referred to it as "symmetrical" symplectic capacities.

Example 5.3. We provide several examples of real symplectic capacities.

(i) The *real Gromov width* $c_B^{\mathrm{real}}(M, \omega, \rho)$ is defined as the supremum over all $r > 0$ such that $(B^{2n}(r), \omega_0, \rho_0)$ real symplectically embeds into (M, ω, ρ). It is normalized and the smallest in the sense that if c is a real symplectic capacity, then $c_B^{\mathrm{real}}(M, \omega, \rho) \leq c(M, \omega, \rho)$ for all real symplectic manifolds (M, ω, ρ).

(ii) In [23], Liu and Wang constructed the real Hofer–Zehnder capacity $c_{\mathrm{HZ}}^{\mathrm{real}}$, which is normalized, by imitating the construction of Hofer–Zehnder capacity [18]. Let $K \subset \mathbb{R}^{2n}$ be a compact convex domain invariant under a linear anti-symplectic involution ρ. It was shown in [20, Theorem 1.3] that the real Hofer–Zehnder capacity of (K, ρ) agrees with the symmetric systole, i.e. $c_{\mathrm{HZ}}^{\mathrm{real}}(K, \rho) = \ell_{\min}^{\mathrm{sym}}(\partial K, \rho)$.

(iii) Following the construction of the (first) Ekeland–Hofer capacity [13], Jin and Lu defined the real Ekeland–Hofer capacity $c_{\mathrm{EH}}^{\mathrm{real}}(\cdot, \rho)$ for compact domains $K \subset \mathbb{R}^{2n}$ invariant under a fixed linear anti-symplectic involution ρ, see [20]. It is normalized. Strictly speaking, it is not a real symplectic capacity as it is defined only for domain in $\mathbb{R}^{2n}$ and satisfies only restricted monotonicity: if $K_1 \subset K_2$ are compact domains in $\mathbb{R}^{2n}$ that are invariant under *a fixed* linear anti-symplectic involution ρ, then we have $c_{\mathrm{EH}}^{\mathrm{real}}(K_1, \rho) \leq c_{\mathrm{EH}}^{\mathrm{real}}(K_2, \rho)$. Nonetheless, we call it a real symplectic capacity. For a compact convex domain $K \subset \mathbb{R}^{2n}$ invariant under a linear anti-symplectic involution ρ, it agrees with the symmetric systole of $(\partial K, \rho)$. Consequently, for every symmetric convex domain (K, ρ) with ρ being linear, the real Hofer–Zehnder capacity and the real Ekeland–Hofer capacity agree, see [20, Theorem 1.10].

(iv) Let (W, λ, ρ) be a real Liouville domain, i.e. (W, λ) is a Liouville domain equipped with an exact anti-symplectic involution ρ. The wrapped Floer homology capacity $c_{\mathrm{HW}}(W, \lambda, \rho)$, constructed using wrapped Floer homology, satisfies restricted monotonicity, meaning that if there exists a generalized real Liouville embedding from (W_1, λ_1, ρ_1) into (W_2, λ_2, ρ_2), then $c_{\mathrm{HW}}(W_1, \lambda_1, \rho_1) \leq c_{\mathrm{HW}}(W_2, \lambda_2, \rho_2)$. See Sect. 4.2 for the construction. Recall that a generalized real symplectic embedding is a real symplectic embedding $\varphi\colon (W_1, \lambda_1, \rho_1) \to (W_2, \lambda_2, \rho_2)$ such that $(\varphi^*\lambda_2 - \lambda_1)|_{\partial W_1} = 0$ in $\mathrm{H}^1(\partial W_1)$. In particular, if W_1 and W_2 are both star-shaped domains in $\mathbb{R}^{2n}$, every real symplectic embedding is a generalized real Liouville embedding since $\mathrm{H}^1(S^{2n-1}) = 0$ for $n \geq 2$. We expect that the argument of [19] applies to all compact convex domains $K \subset \mathbb{R}^{2n}$ invariant under a linear anti-symplectic involution ρ, implying that $c_{\mathrm{HW}}(K,\rho) = \ell^{\mathrm{sym}}_{\min}(\partial K, \rho)$.

(v) Analogously to Gutt–Hutchings [17], we can construct, using positive equivariant wrapped Floer homology defined in [21], a sequence of real symplectic capacities $c_1 \le c_2 \le \cdots \le \infty$ for real Liouville domains. They satisfy all the conditions for real symplectic capacities, but the monotonicity. Instead, they satisfy the restricted monotonicity as the wrapped Floer homology capacity. Using a Gysin-type exact sequence in wrapped Floer theory (see [21, Proposition 3.27] and [10, Proposition 2.9]), it is not hard to see that $c_1(W,\lambda,\rho) \le c_{\mathrm{HW}}(W,\lambda,\rho)$ for every real Liouville domain (W,λ,ρ).

There is an old question about symplectic capacities asking if all normalized symplectic capacities agree on compact convex domains in $\mathbb{R}^{2n}$, see [24, Section 14.9, Problem 53] and [25, Section 5]. We finish this article with the following related conjecture.

Conjecture 5.4. *For convex domains in $\mathbb{R}^{2n}$ invariant under a fixed linear anti-symplectic involution, all normalized symplectic capacities and normalized real symplectic capacities are the same.*

Acknowledgements

The authors cordially thank Urs Frauenfelder, Jungsoo Kang, and Felix Schlenk for fruitful discussions, Yaron Ostrover for helpful comments. They also thank the anonymous referee for valuable suggestions. A part of this work was done while MK visited Korea Institute for Advanced Study. The authors are grateful for its warm hospitality. JK was partially supported by the POSCO Science Fellowship of POSCO TJ Park Foundation SK was supported by the grant 200021-181980/1 of the Swiss National Foundation. MK was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2021R1F1A1060118) and by Sunchon National University Research Fund in 2021 (Grant number: 2021-0238).

References

[1] Abbondandolo, A., Kang, J.: Symplectic homology of convex domains and Clarke's duality. to appear in Duke Math. J

[2] Abouzaid, M.: A geometric criterion for generating the Fukaya category. Publ. Math. Inst. Hautes Études Sci. **112**, 191–240 (2010)

[3] Abouzaid, M., Seidel, P.: An open string analogue of Viterbo functoriality. Geom. Topol. **14**(2), 627–718 (2010)

[4] Akopyan, A., Karasev, R.: Estimating symplectic capacities from lengths of closed curves on the unit spheres. arXiv preprint (2017)

[5] Albers, P.: A Lagrangian Piunikhin-Salamon-Schwarz morphism and two comparison homomorphisms in Floer homology. Int. Math. Res. Not. IMRN (4):Art. ID rnm134, 56 (2008)

[6] Birkhoff, G.: The restricted problem of three bodies. Rend. Circ. Matem. Palermo **39**, 265–334 (1915)

[7] Borman, M.S., McLean, M.: Bounding Lagrangian widths via geodesic paths. Compos. Math. **150**(12), 2143–2183 (2014)

[8] Bourgeois, F.: A Morse–Bott approach to contact homology. Thesis (Ph.D.)–Stanford Univeristy (2002)

[9] Bourgeois, F., Oancea, A.: Symplectic homology, autonomous Hamiltonians, and Morse-Bott moduli spaces. Duke Math. J. **146**(1), 71–174 (2009)

[10] Bourgeois, F., Oancea, A.: S^1-equivariant symplectic homology and linearized contact homology. Int. Math. Res. Not. IMRN **13**, 3849–3937 (2017)

[11] Cieliebak, K., Hofer, H., Latschev, J., Schlenk F.: Quantitative symplectic geometry. In Dynamics, ergodic theory, and geometry. volume 54 of Math. Sci. Res. Inst. Publ., page 1–44. Cambridge Univ. Press, Cambridge (2007)

[12] Cieliebak, K., Oancea, A.: Symplectic homology and the Eilenberg–Steenrod axioms. Algebr. Geom. Topol. **18**(4), 1953–2130 (2018) Appendix written jointly with Peter Albers

[13] Ekeland, I., Hofer, H.: Symplectic topology and Hamiltonian dynamics. Math. Z. **200**(3), 355–378 (1989)

[14] Floer, A., Hofer, H., Wysocki, K.: Applications of symplectic homology. I. Math. Z. **217**(4), 577–606 (1994)

[15] Frauenfelder, U., van Koert, O.: The restricted three body problem and holomorphic curves. Pathways in Mathematics. Birkhäuser Basel, 1 edition (2018)

[16] Ganatra, S.: Symplectic Cohomology and Duality for the Wrapped Fukaya Category. Thesis (Ph.D.)–Massachusetts Institute of Technology (2012)

[17] Gutt, J., Hutchings, M.: Symplectic capacities from positive S^1-equivariant symplectic homology. Algebr. Geom. Topol. **18**(6), 3537–3600 (2018)

[18] Hofer, H., Zehnder, E.: Symplectic invariants and Hamiltonian dynamics. Modern Birkhäuser Classics. Birkhäuser, Basel (2011). Reprint of the 1994 edition

[19] Irie, K.: Symplectic homology of fiberwise convex sets and homology of loop spaces. to appear in J. Symplectic Geom

[20] Jin, R., Lu, G.: Representation formula for symmetrical symplectic capacity and applications. Discrete Contin. Dyn. Syst. **40**(8), 4705–4765 (2020)

[21] Kim, J., Kim, S., Kwon, M.: Equivariant wrapped Floer homology and symmetric periodic Reeb orbits. online ready in Ergodic Theory Dynam. Systems (2021)

[22] Kim, J., Kwon, M., Lee, J.: Volume growth in the component of fibered twists. Commun. Contemp. Math. **20**(8), 1850014, 43 (2018)

[23] Liu, C., Wang, Q.: Symmetrical symplectic capacity with applications. Discrete Contin. Dyn. Syst. **32**(6), 2253–2270 (2012)

[24] McDuff, D., Salamon, D.: Introduction to Symplectic Topology, 3rd edn. Oxford Graduate Texts in Mathematics, Oxford (2017)

[25] Ostrover, Y.: When symplectic topology meets Banach space geometry. In Proceedings of the International Congress of Mathematicians—Seoul 2014. Vol. II. pages 959–981. Kyung Moon Sa, Seoul (2014)

[26] Piunikhin, S., Salamon, D., Schwarz, M.: Symplectic Floer-Donaldson theory and quantum cohomology. In Contact and symplectic geometry (Cambridge, 1994). volume 8 of Publ. Newton Inst. pages 171–200. Cambridge Univ. Press, Cambridge (1996)

[27] Rabinowitz, P.H.: On the existence of periodic solutions for a class of symmetric Hamiltonian systems. Nonlinear Anal. **11**(5), 599–611 (1987)

[28] Ritter, A.F.: Topological quantum field theory structure on symplectic cohomology. J. Topol. **6**(2), 391–489 (2013)

[29] Ritter, A.F., Smith, I.: The monotone wrapped Fukaya category and the open-closed string map. Select. Math. (N.S.) **23**(1), 533–642 (2017)

Joontae Kim
Department of Mathematics
Sogang University
35 Baekbeom-ro, Mapo-gu
Seoul 04107
Republic of Korea
e-mail: joontae@sogang.ac.kr

Seongchan Kim
Department of Mathematics Education
Kongju National University
Kongju 32588
Republic of Korea
e-mail: seongchankim@kongju.ac.kr

Myeonggi Kwon
Department of Mathematics Education
Sunchon National University
Suncheon 57922
Republic of Korea
e-mail: mkwon@scnu.ac.kr

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:21
https://doi.org/10.1007/s11784-022-00954-9
Published online April 1, 2022

Journal of Fixed Point Theory and Applications

Diffeomorphism type via aperiodicity in Reeb dynamics

Myeonggi Kwon, Kevin Wiegand and Kai Zehmisch

Abstract. We characterise boundary-shaped disc-like neighbourhoods of certain isotropic submanifolds in terms of aperiodicity of Reeb flows. We prove uniqueness of homotopy and diffeomorphism type of such contact manifolds assuming non-existence of short periodic Reeb orbits.

Mathematics Subject Classification. 53D35, 37C27, 37J55, 57R17.

1. Introduction

In their seminal work, Gromov [1] and Eliashberg [2] observed that foliations by holomorphic curves can be used to prove uniqueness of the diffeomorphism (in fact symplectomorphism) type of minimal symplectic fillings of the standard contact 3-sphere, i.e., all such fillings are diffeomorphic to the 4-ball D^4. The method they used, the so-called *filling by holomorphic curves* method, is obstructed by bubbling off of holomorphic spheres. Related classification results in dimension 4 can be found in [3–8].

On the other hand, Hofer [9] discovered a fundamental property of holomorphic curves in symplectisations; non-compactness properties of holomorphic curves of finite energy are strongly related to the existence of periodic Reeb orbits. Combining the method of filling by holomorphic curves with the theory of finite energy planes Eliashberg–Hofer [10] determined the diffeomorphism (in fact contactomorphism) type of certain contact manifolds with boundary S^2: any compact contact manifold with boundary $S^2 = \partial D^3$ is diffeomorphic to D^3 provided there exists a contact form that is equal to

This work is supported by the Deutsche Forschungsgemeinschaft as part of a project in the SFB/TRR 191 *Symplectic Structures in Geometry, Algebra and Dynamics* (project number 281071066—TRR 191), the Research Training Group RTG 2229 *Asymptotic Invariants and Limits of Groups and Spaces* (281869850) and under Germany's Excellence Strategy EXC2181/1-390900948 (the Heidelberg STRUCTURES Excellence Cluster)

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday".

the standard contact form on D^3 near the boundary S^2 such that the corresponding Reeb vector field does not admit a periodic orbit with period less than π. A similar characterisation of $D^2 \times S^1$ in terms of Reeb dynamics was obtained by Kegel–Schneider–Zehmisch [11].

In higher dimensions, the diffeomorphism type of symplectically aspherical fillings of the standard contact sphere was determined by Eliashberg–Floer–McDuff [12, Theorem 1.5]: any such filling is diffeomorphic to the ball D^{2n}. The proof they used was refined to the so-called *degree method* (see Sect. 3.2 for an explanation) by Barth–Geiges–Zehmisch [13] allowing a much wider class of contact type boundaries, see also [14–16].

The contact theoretic counterpart in higher dimensions was not clear for a while. It was conjectured by Bramham–Hofer [17] that the existence of trapped Reeb orbits on a compact contact manifold, whose boundary neighbourhoods look like neighbourhoods of $S^{2n} = \partial D^{2n+1}$ in D^{2n+1}, implies the existence of periodic Reeb orbits. A counterexample to that conjecture was given by Geiges–Röttgen–Zehmisch [18]. It suggests that the diffeomorphism type in higher dimensional contact geometry should be determined via a method not based on non-existence of trapped orbits as done in Eliashberg–Hofer [10].

In fact, using the degree method, Geiges–Zehmisch [19] proved that any compact strict contact manifold that has an aperiodic Reeb flow is diffeomorphic to D^{2n+1} provided that the following condition is satisfied: A neighbourhood of the boundary admits a strict contact embedding into the standard D^{2n+1} mapping the boundary to $S^{2n} = \partial D^{2n+1}$. This was generalised by Barth–Schneider–Zehmisch [20] to situations in which D^{2n+1} is replaced by the disc bundle of $\mathbb{R} \times T^*T^d \times \mathbb{C} \times \mathbb{C}^{n-1-d}$ whenever $n-1 \geq d$.

The aim of this work is to replace the torus T^d by more general d-dimensional manifolds, see Theorem 2.1 below. Again the argument will be based on the construction of a proper degree 1 evaluation map on the moduli space of 1-marked holomorphic discs with varying Lagrangian boundary conditions. The restriction to T^d in [20] was caused by the choice of the boundary conditions set up for the holomorphic discs. This led to trivialising the cotangent bundle of T^d in a Stein holomorphic fashion. To replace D^{2n+1} by the disc bundle of $\mathbb{R} \times T^*Q \times \mathbb{C} \times \mathbb{C}^{n-1-d}$ for a wider class of manifolds Q we choose different boundary conditions for the holomorphic discs. Instead of taking a foliation of T^*Q by sections we consider the foliation T^*Q given by the cotangent fibres. This will result in a more advanced analysis for the holomorphic discs. The essential point here will be a target rescaling argument in Sect. 7, which was invented by Bae–Wiegand–Zehmisch [21] in the context of virtually contact structures, to ensure C^0-bounds on holomorphic discs in the situation of general manifolds Q. Furthermore, to obtain C^0-bounds of holomorphic discs along their boundaries in T^*Q-direction, we develop an integrated maximum principle in Sects. 5 and 6.5 .

2. Aperiodicity and boundary shape

Strict contact manifolds (M, α) are naturally equipped with a nowhere vanishing vector field, namely the Reeb vector field of α. Assuming α to be **aperiodic**, i.e., assuming that the Reeb vector field does not admit any periodic solution, the diffeomorphism type of M can be determined in many situations. Here we are interested in comparing compact manifolds with boundary M with neighbourhoods of isotropic submanifolds of the sort $D\big(T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}\big)$. This requires boundary conditions for the Reeb vector field as we will explain in the following:

2.1. A model

Let Q be a closed, connected Riemannian manifold of dimension d and let $n \in \mathbb{N}$ such that $n-1 \geq d$. Define a strict contact manifold (C, α_0) by setting

$$C := \mathbb{R} \times T^*Q \times \mathbb{C} \times \mathbb{C}^{n-1-d}$$

and

$$\alpha_0 := \mathrm{d}b + \lambda + \frac{1}{2}\big(x_0 \mathrm{d}y_0 - y_0 \mathrm{d}x_0\big) - \sum_{j=1}^{n-1-d} y_j \mathrm{d}x_j,$$

where $b \in \mathbb{R}$, λ is the Liouville 1-form of T^*Q, $x_0 + \mathrm{i}y_0$ and $x_j + \mathrm{i}y_j$ are coordinates on $\mathbb{C}$ and $\mathbb{C}^{n-1-d}$, resp. Throughout the text, we will use vector notation $\mathbf{x}$ and $\mathbf{y}$ for the coordinate tuples $(x_1, \ldots, x_{n-1-d})$ and $(y_1, \ldots, y_{n-1-d})$, resp., so that we can abbreviate

$$-\mathbf{y}\mathrm{d}\mathbf{x} = -\sum_{j=1}^{n-1-d} y_j \mathrm{d}x_j.$$

The Reeb vector field of α is given by ∂_b, which is tangent to the real lines $\mathbb{R} \times \{*\}$.

By [22, Theorem 6.2.2] (C, α_0) is the model neighbourhood of an isotropic submanifold Q of a strict contact manifold provided that Q has trivial symplectic normal bundle and the dimension d of Q is smaller than n. Observe, that (C, α_0) is the contactisation of the Liouville manifold

$$\left(T^*Q \times \mathbb{C} \times \mathbb{C}^{n-1-d},\, \lambda + \frac{1}{2}\big(x_0 \mathrm{d}y_0 - y_0 \mathrm{d}x_0\big) - \mathbf{y}\mathrm{d}\mathbf{x}\right).$$

The statements about the model neighbourhood situation and contactisation of course hold in the critical case $d = n$ also. Simply ignore the Euclidean factors in the formulations.

2.2. Fibrewise shaped

The space C itself is the total space of the stabilised cotangent bundle $T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}$. Let $S \subset C$ be a hypersurface diffeomorphic to the unit sphere bundle $S\big(T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}\big)$ such that

(1) S intersects each fibre transversely in a sphere

$$S_q := S \cap \big(T_q^*Q \oplus \mathbb{R}^{2n+1-2d}\big), \quad q \in Q,$$

of dimension $2n-d$, and

(2) each S_q intersects the flow lines of ∂_b in at most two points. We require transverse intersections if such a flow line intersects S_q in two points. Points of tangency, i.e. points that correspond to single intersections, form a submanifold diffeomorphic to a $(2n-d-1)$-sphere.

In view of condition (1), we remark that the hypersurface S bounds a bounded domain D inside C, whose closure is diffeomorphic to the closed unit disc bundle $D\big(T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}\big)$. Condition (2) will play an important role in Sect. 3.1. We call S a **shape**.

2.3. Standard near the boundary

Let (M,α) be a strict contact manifold of dimension $2n+1$ that is **standard near the boundary**, i.e.

(1) connected, compact with boundary ∂M diffeomorphic to

$$\partial M \cong S\big(T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}\big),$$

(2) such that there exist an open collar neighbourhood $U \subset M$ of ∂M and an embedding $\varphi : (U, \partial U = \partial M) \to (D, S)$ such that $\varphi^*\alpha_0 = \alpha$ on U.

If φ is given we will call S the **shape** of M.

To quantify aperiodicity of (M,α) we denote by $\inf_0(\alpha) > 0$ the minimal action of all *contractible* closed Reeb orbits w.r.t. α. By Darboux's theorem, $\inf_0(\alpha)$ is indeed positive. For aperiodic α, we set $\inf_0(\alpha)$ to be ∞.

A second ingredient for quantisation comes with the subset

$$Z := \mathbb{R} \times T^*Q \times \mathbb{D} \times \mathbb{C}^{n-1-d}$$

of C denoting the closed unit disc in $\mathbb{C}$ by $\mathbb{D}$. We may assume that $S \subset \operatorname{Int} Z$ by scaling radially via $\big(t^2 b, t^2 \mathbf{w}, t z_0, t\mathbf{z}\big)$, $t \in (0,1)$, if necessary. The contact form α on M will be replaced by $t^2\alpha$ accordingly.

2.4. Main theorem

We compare the homology, homotopy and diffeomorphism type of M with the one of $D\big(T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}\big)$. This will be done in terms of embeddings

$$D\big(T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}\big) \longrightarrow M$$

determined by a small neighbourhood of a section $Q \to S$ as constructed, e.g. at the beginning of Sect. 9. We denote the image of such an embedding by

$$M_0 := D\big(T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}\big).$$

Theorem 2.1. *Let Q be an oriented, closed, connected Riemannian manifold of dimension d. Let $n \in \mathbb{N}$ such that $n-1 \geq d$. Let (M,α) be a strict contact manifold that is standard near the boundary as described in Sect. 2.3. Assume that the shape $S \cong \partial M$ of M is contained in the interior of (Z,α_0). If $\inf_0(\alpha) \geq \pi$, then the following is true:*

(i) *Any embedding $Q \to M$ given by a section $Q \to S$ induces isomorphisms of homology and surjections of fundamental groups. If in addition $\pi_1 Q$ is abelian, then the surjections are injective.*

(ii) *Assume that $\pi_1 Q$ is abelian and that at least one of the following conditions is satisfied:*
   (a) *$\pi_1 Q$ is finite.*
   (b) *Q is aspherical.*
   (c) *Q is simple and $S \to Q$ a trivial sphere bundle, or, more generally, S is a simple space.*

   Then, M is homotopy equivalent to M_0.

(iii) *If in addition to the assumptions in* (ii) (*including choices of one of the conditions* (a)–(c)) *we have that $2n+1 \geq 7$ and that the Whitehead group of $\pi_1 Q$ is trivial, then M is diffeomorphic to M_0.*

2.5. Comments on Theorem 2.1

In view of the contact connected sum, the bound π in the theorem is optimal, cf. [19, Remark 1.3.(1)]. The shape boundary condition can be isotoped to a round shape through shaped hypersurfaces. Hence, we recover the results from [19,20] and obtain independence of the choice of metric.

The orientation of Q will not be used in the compactness argument below. But will be needed for an orientation of the moduli space. Without orientation we only can talk about the mod-2 degree of the evaluation map. Hence, if Q is not orientable, only part (i) of the theorem remains true replacing homology by homology with $\mathbb{Z}_2$-coefficients.

Similarly, the boundary of M is necessarily connected, cf. [19, Remark 1.3.(4)]. Indeed, suppose ∂M has several components that have individually a shape embedding into potentially different stabilised cotangent bundles. Here, different Qs with varying dimensions are allowed. M itself satisfies the remaining stated properties from Theorem 2.1. In this situation one can set up the moduli space of holomorphic discs with respect to one distinguished boundary component; the other components will come with the maximum principle for holomorphic curves. In other words, the holomorphic disc analysis will be uneffected and the evaluation map on the moduli space will be of degree one. This contradicts the fact that no holomorphic disc can exceed one of the additional boundary components due to the maximum principle.

Example 2.2. In view of the Hadamard–Cartan and the Farrell–Jones theorems, the assumptions of Theorem 2.1 part (b) in (ii) and (iii) are satisfied for all Riemannian manifolds Q with abelian fundamental group and non-positive sectional curvature. Hence, we recover $Q = T^d$ from [20].

Example 2.3. A particular class of manifolds Q that satisfy the assumptions of Theorem 2.1 part (c) in (ii) and (iii) are products of unitary groups and spheres of any dimensions. Indeed, such Q always have stably trivial tangent bundle, are simple with fundamental group free abelian so that in particular the Whitehead group of those is trivial.

Remark 2.4. If we know more about the handle body structure of M conditions on the topology of Q can be relaxed. For example if M has the homotopy type of a CW complex of codimension 2 so that the inclusion $\partial M \subset M$ is π_1-injective the assumption $\pi_1 Q$ abelian in Theorem 2.1 can be dropped everywhere.

If M admits a handle body structure with all handles of index at most ℓ and if $d+\max(d,\ell) \leq 2n-1$, then M and M_0 are homotopy equivalent without any further conditions. This follows with the argument from [13, Theorem 7.2] using the diagram in Sect. 9.2. In particular, the CW-dimension of M must be equal to d. In fact, one can conclude with the diffeomorphism type as in [13, Theorem 9.4], cf. [13, Example 9.5].

3. The degree method

We will explain the main idea of the proof of Theorem 2.1, which will be given in Sects. 4–9.

3.1. Completion via gluing

Assuming $S \subset \operatorname{Int} Z$, we define smooth manifolds

$$\hat{C} := (C \setminus \operatorname{Int} D) \cup_\varphi M, \quad \hat{Z} := (Z \setminus \operatorname{Int} D) \cup_\varphi M$$

by gluing via φ and equip both with the contact form

$$\hat{\alpha} := \alpha_0 \cup_\varphi \alpha$$

that coincides with α on M and with α_0 on $C \setminus \operatorname{Int} D$. Because of the contact embedding φ of $U \supset \partial M$ into (Z, α_0) this is well defined. According to the second shape condition in Sect. 2.2, the gluing does not create additional periodic Reeb orbits inside $(\hat{C}, \hat{\alpha})$ so that $\inf_0(\alpha)$ and $\inf_0(\hat{\alpha})$ coincide.

3.2. Filling by holomorphic discs

To prove Theorem 2.1, we will argue as in [19,20]: The Liouville manifold

$$\left(T^*Q \times \mathbb{D} \times \mathbb{C}^{n-1-d},\, \lambda + \frac{1}{2}\big(x_0 \mathrm{d}y_0 - y_0 \mathrm{d}x_0\big) - \mathbf{y}\mathrm{d}\mathbf{x}\right)$$

is foliated by holomorphic discs $\{\mathbf{w}\} \times \mathbb{D} \times \{\mathbf{s} + \mathrm{i}\mathbf{t}\}$. Using the Niederkrüger transformation from Sect. 4.3 these discs can be lifted to holomorphic discs in the symplectisation of the contactisation (Z, α_0) and are called **standard discs**. After gluing some of the standard discs will survive, namely those which correspond to the end of $(\hat{Z}, \hat{\alpha})$ in the symplectisation (W, ω) of $(\hat{Z}, \hat{\alpha})$. We will study the corresponding moduli space $\mathcal{W}$ of holomorphic discs

$$u = (a, f)\colon\; \mathbb{D} \longrightarrow W$$

subject to varying Lagrangian boundary conditions, which will differ substantially from those used in [19,20]. This requires a different argument to obtain C^0-bounds for holomorphic discs, which at the end allows a wider class of base manifolds Q.

It will turn out that the evaluation map

$$\begin{aligned} \mathrm{ev}\colon \quad \mathcal{W}\times\mathbb{D} &\longrightarrow \hat{Z} \\ \big((a,f),z\big) &\longmapsto f(z) \end{aligned}$$

either is proper of degree one or there will be breaking off of finite energy planes. The first alternative allows conclusions on the diffeomorphism type of M with the s-cobordism theorem as in [13]. The second results in the existence of a **short** contractible periodic Reeb orbit of α on M by a result of Hofer [9]. Short here means that the action of the Reeb orbit is bounded by the area of $\mathbb{D}$.

The condition $\inf_0(\alpha)\geq\pi$ will exclude breaking of holomorphic discs along periodic Reeb orbits of action less than π. But in fact, under the assumptions of Theorem 2.1 the shape S of M actually is contained in $\mathbb{R}\times T^*Q\times B_r(0)\times\mathbb{C}^{n-1-d}$ for $r\in(0,1)$. Working out the proof of Theorem 2.1 with that slightly smaller radius r we will see that requiring non-existence of short periodic Reeb orbits with period bounded by πr^2 will be sufficient. In other words, we can assume that $\inf_0(\alpha)>\pi r^2$ to prove properness of the evaluation map ev. To simplify notation, we will assume $r=1$, i.e. from now on we assume $\inf_0(\alpha)>\pi$.

4. Standard holomorphic discs

In this section, we construct standard holomorphic discs. We will follow [19, Section 2] and [20, Section 2] adding adjustments to the current situation.

4.1. The contactisation

We consider the Liouville manifold

$$(V,\lambda_V):=\Big(T^*Q\times\mathbb{D}\times\mathbb{C}^{n-1-d},\,\lambda+\frac{1}{2}\big(x_0\mathrm{d}y_0-y_0\mathrm{d}x_0\big)-\mathbf{y}\mathrm{d}\mathbf{x}\Big),$$

whose contactisation $(\mathbb{R}\times V,\mathrm{d}b+\lambda_V)$ is (Z,α_0). The induced contact structure $\xi_0=\ker\alpha_0$ on Z is spanned by tangent vectors of the form $v-\lambda_V(v)\partial_b$, $v\in TV$.

4.2. Liouville manifold and potential

Denote by J_{T^*Q} the almost complex structure on T^*Q that is compatible with $\mathrm{d}\lambda$ and satisfies $\lambda=-\mathrm{d}F\circ J_{T^*Q}$. Here F is a strictly plurisubharmonic potential in the sense of [23, Section 3.1] that coincides with the kinetic energy function near the zero section of T^*Q and interpolates to the length function on the complement of a certain disc bundle in T^*Q, see [24, Section 3.1]. In Sect. 5, we will present a construction of (F,J_{T^*Q}).

Define an almost complex structure on the Liouville manifold (V,λ_V) by setting

$$J_V:=J_{T^*Q}\oplus\mathrm{i}\oplus\mathrm{i}.$$

J_V is compatible with the symplectic form $\mathrm{d}\lambda_V$ and satisfies $\lambda_V=-\mathrm{d}\psi\circ J_V$, where ψ is the strictly plurisubharmonic potential

$$\psi(\mathbf{w},z_0,\mathbf{z}):=F(\mathbf{w})+\frac{1}{4}|z_0|^2+\frac{1}{2}|\mathbf{y}|^2$$

denoting by $\mathbf{w} \in T^*Q$ a co-vector of Q, $z_0 \in \mathbb{D}$ and using complex coordinates $z_j = x_j + \mathrm{i}y_j$, $j = 1, \ldots, n-1-d$ on $\mathbb{C}^{n-1-d}$. Again the tuple $(z_1, \ldots, z_{n-1-d})$ is abbreviated by $\mathbf{z}$ so that $\frac{1}{2}|\mathbf{y}|^2$ reads as

$$\frac{1}{2}\sum_{j=1}^{n-1-d} y_j^2.$$

In particular, (V, J_V) is foliated by holomorphic discs $\{\mathbf{w}\} \times \mathbb{D} \times \{\mathbf{s} + \mathrm{i}\mathbf{t}\}$.

4.3. The symplectisation

Let $\tau \equiv \tau(a)$ be a strictly increasing smooth function $\mathbb{R} \to (0, \infty)$. We consider the symplectisation

$$\big(\mathbb{R} \times Z, \mathrm{d}(\tau\alpha_0)\big)$$

of (Z, α_0). Define a compatible, translation invariant almost complex structure J that preserves the contact hyperplanes ξ_0 on all slices $\{a\} \times Z$ by requiring that $J(\partial_a) = \partial_b$ and that

$$J\big(v - \lambda_V(v)\partial_b\big) = J_V v - \lambda_V(J_V v)\partial_b$$

for all $v \in TV$. The **Niederkrüger map** is the biholomorphism

$$\begin{aligned} \Phi\colon\ (\mathbb{R} \times \mathbb{R} \times V, J) &\longrightarrow \ (\mathbb{C} \times V, \mathrm{i} \oplus J_V) \\ (a, b, \mathbf{z}) \quad &\longmapsto \big(a - \psi(\mathbf{z}) + \mathrm{i}b, \mathbf{z}\big) \end{aligned}$$

recalling that $Z = \mathbb{R} \times V$, see [25, Proposition 5] and [19, Proposition 2.1].

4.4. The Niederkrüger transform

Using the inverse of Φ, we lift the holomorphic discs

$$\{a + \mathrm{i}b\} \times \{\mathbf{w}\} \times \mathbb{D} \times \{\mathbf{s} + \mathrm{i}\mathbf{t}\}$$

from $(\mathbb{C} \times V, \mathrm{i} \oplus J_V)$ to the symplectisation $(\mathbb{R} \times \mathbb{R} \times V, J)$ of (Z, α_0). For fixed $b \in \mathbb{R}$, $\mathbf{w} \in T^*Q$, and $\mathbf{s}, \mathbf{t} \in \mathbb{R}^{n-1-d}$, the resulting **standard holomorphic discs**

$$\mathbb{D} \longrightarrow \mathbb{R} \times \mathbb{R} \times T^*Q \times \mathbb{D} \times \mathbb{C}^{n-1-d}$$

are parametrised by

$$u^{\mathbf{t},\mathbf{w}}_{\mathbf{s},b}(z) = \Big(\tfrac{1}{4}\big(|z|^2 - 1\big), b, \mathbf{w}, z, \mathbf{s} + \mathrm{i}\mathbf{t}\Big),$$

cf. [19, Section 2.2].

To set boundary conditions for the standard discs we define a $(n-1)$-dimensional family of cylinders

$$L^{\mathbf{t}}_q := \{0\} \times \mathbb{R} \times T^*_q Q \times \partial\mathbb{D} \times \mathbb{R}^{n-1-d} \times \{\mathbf{t}\},$$

where $\mathbf{t} \in \mathbb{R}^{n-1-d}$ and $q \in Q$ are the parameters. Observe, that the $L^{\mathbf{t}}_q$ foliate $\{0\} \times \partial Z$. Furthermore the restriction of $\mathrm{d}(\tau\alpha_0)$ to the tangent bundle of $\{0\} \times Z$ equals $\tau(0)\mathrm{d}\alpha_0$, which is a positive multiple of

$$\mathrm{d}\lambda + \mathrm{d}x_0 \wedge \mathrm{d}y_0 + \mathrm{d}\mathbf{x} \wedge \mathrm{d}\mathbf{y}.$$

Therefore, $L^{\mathbf{t}}_q$ is a Lagrangian cylinder because the dimension of $L^{\mathbf{t}}_q$ is $n+1$.

4.5. Class independence

Preparing the definition of the moduli space $\mathcal{W}$ we consider the space $\mathbb{R} \times T_q^*Q \times \mathbb{R}^{n-1-d}$ of tuples $(b, \mathbf{w}, \mathbf{s})$. Assuming $n \geq 2$ this space is at least 2-dimensional, so that the complement of any ball in $\mathbb{R} \times T_q^*Q \times \mathbb{R}^{n-1-d}$ is path-connected. Therefore, we find $R > 0$ such that

(1) the shape S is contained in the closed disc bundle $D_R\big(T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}\big)$ of radius R, and
(2) all standard discs $u_{\mathbf{s},b}^{\mathbf{t},\mathbf{w}}$ of level $(q, \mathbf{t})$, $\mathbf{w} \in T_q^*Q$, that are contained in

$$\mathbb{R} \times \Big(Z \setminus D_R\big(T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}\big)\Big)$$

are homotopic therein relative $L_q^{\mathbf{t}}$ via a homotopy inside

$$\{0\} \times \mathbb{R} \times T_q^*Q \times \mathbb{D} \times \mathbb{R}^{n-1-d} \times \{\mathbf{t}\}.$$

5. Symplectic potentials on cotangent bundles

We prepare the proof of geometric bounds on holomorphic discs that belong to the moduli space $\mathcal{W}$. The aim of this section is to construct an almost complex structure on T^*Q.

The almost complex structure on T^*Q that belongs to the Levi-Civita connection of Q is the one that is induced by the kinetic energy function. The one coming from symplectising the unit cotangent bundle in contrast belongs to the length functional and does not extend over the zero section. Here we want to interpolate the two in order to obtain C^0-bounds on holomorphic curves in the complement of the unit codisc bundle that we after all can identify with the positive symplectisation also holomorphically.

5.1. Dual connection

We denote the covariant derivative of the Levi-Civita connection of Q by ∇. The corresponding covariant derivative ∇^* of the dual connection is defined via chain rule by

$$\big(\nabla^*\beta\big)(X, Y) := \big(\nabla_X^*\beta\big)(Y) := X\big(\beta(Y)\big) - \beta(\nabla_X Y)$$

for 1-forms β and vector fields X, Y on Q, cf. [21, Section 4]. Denoting the Christoffel symbols of ∇ by Γ_{ij}^k the Christoffel symbols $(\Gamma^*)_{ij}^k$ of ∇^* can be expressed by $(\Gamma^*)_{ij}^k = -\Gamma_{ik}^j$. The connection map of the dual connection $K\colon TT^*Q \to T^*Q$ and the tangent functor T are related via $K \circ T = \nabla^*$ and defines a splitting of

$$TT^*Q = \mathcal{H} \oplus \mathcal{V}$$

into horizontal

$$\mathcal{H} := \ker\big(K\colon\, TT^*Q \longrightarrow T^*Q\big)$$

and vertical distribution

$$\mathcal{V} = \ker\big(T\tau\colon\, TT^*Q \longrightarrow TQ\big),$$

where $T\tau$ is the linearisation of the cotangent map $\tau\colon T^*Q \to Q$. Observe that $T\tau$ defines a bundle isomorphism from $\mathcal{H}$ onto τ^*TQ and that $\mathcal{V}$ can be identified with τ^*T^*Q canonically.

5.2. Orthogonal splitting

Denoting the metric of Q by g, contraction defines a bundle isomorphism

$$\begin{aligned} G: TQ &\longrightarrow T^*Q \\ v &\longmapsto i_v g. \end{aligned}$$

The dual metric $g^\flat$ is defined by

$$g^\flat(\alpha,\beta) = g\big(G^{-1}(\alpha), G^{-1}(\beta)\big)$$

for co-vectors $\alpha, \beta \in T^*Q$ on Q, so that the dual norm $\alpha \mapsto |\alpha|_\flat$ defines the **length function** on T^*Q. The **kinetic energy function** reads as

$$k(\beta) = \frac{1}{2}|\beta|_\flat^2.$$

For a smooth, strictly increasing function $\chi\colon \mathbb{R} \to \mathbb{R}$ with $\chi(0) = 0$ we define

$$F = \chi \circ k\colon T^*Q \to [0,\infty).$$

This leads to a Riemannian metric h on T^*Q defined by

$$h\big(v \oplus \alpha, w \oplus \beta\big) := \frac{1}{\chi' \circ k} \cdot g\Big(T\tau(v), T\tau(w)\Big) + (\chi' \circ k) \cdot g^\flat(\alpha,\beta),$$

where $v, w \in \mathcal{H}$ and $\alpha, \beta \in \mathcal{V}$. The metric h turns $TT^*Q = \mathcal{H} \oplus \mathcal{V}$ into an orthogonal splitting.

5.3. Taming structure

The Liouville form λ on T^*Q is given by $\lambda_{\mathbf{w}} = \mathbf{w} \circ T\tau$ for $\mathbf{w} \in T^*Q$ and defines a symplectic form via $\mathrm{d}\lambda$. Observe that for $v, w \in \mathcal{H}$ and $\alpha, \beta \in \mathcal{V}$

$$\lambda_u\big(v \oplus \alpha\big) = \mathbf{w}\big(T\tau(v)\big)$$

and

$$\mathrm{d}\lambda\big(v \oplus \alpha, w \oplus \beta\big) = \alpha\big(T\tau(w)\big) - \beta\big(T\tau(v)\big).$$

In view of the splitting $TT^*Q = \mathcal{H} \oplus \mathcal{V}$, we define the almost complex structure J_{T^*Q} by setting

$$J_{T^*Q}\big(v \oplus \alpha\big) := (\chi' \circ k) \cdot G^{-1}(\alpha) \oplus \frac{-1}{\chi' \circ k} \cdot G(v)$$

for $v \in \mathcal{H}$ and $\alpha \in \mathcal{V}$. This yields

$$h = \mathrm{d}\lambda\big(.,J_{T^*Q}.\,\big),$$

i.e. J_{T^*Q} is compatible with the symplectic form $\mathrm{d}\lambda$. Non-degeneracy of the metric h and the symplectic form $\mathrm{d}\lambda$ shows that the almost complex structure J_{T^*Q} is uniquely determined.

5.4. Potentials

We claim that the function F is a symplectic potential on the tame symplectic manifold $(T^*Q, \mathrm{d}\lambda, J_{T^*Q})$ in the sense that

$$\lambda = -\mathrm{d}F \circ J_{T^*Q}.$$

Indeed, in local $(\mathbf{q},\mathbf{p})$-coordinates on T^*Q induced by Riemann coordinates on Q about $\mathbf{q} \equiv \mathbf{0}$ we have

$$\mathcal{H}_{(\mathbf{0},\mathbf{p})} = \left\{ \big(\mathbf{0},\mathbf{p},\dot{\mathbf{q}},\mathbf{0}\big) \mid \dot{\mathbf{q}} \in \mathbb{R}^d \right\}, \quad \mathcal{V}_{(\mathbf{0},\mathbf{p})} = \left\{ \big(\mathbf{0},\mathbf{p},\mathbf{0},\dot{\mathbf{p}}\big) \mid \dot{\mathbf{p}} \in \mathbb{R}^d \right\},$$

as well as

$$\lambda_{(\mathbf{0},\mathbf{p})} = \mathbf{p}\,\mathrm{d}\mathbf{q}, \quad \mathrm{d}\lambda_{(\mathbf{0},\mathbf{p})} = \mathrm{d}\mathbf{p} \wedge \mathrm{d}\mathbf{q},$$

and

$$\big(J_{T^*Q}\big)_{(\mathbf{0},\mathbf{p})} = \begin{pmatrix} 0 & \chi'\big(\frac{1}{2}p^ip^i\big) \\ -\Big(\chi'\big(\frac{1}{2}p^ip^i\big)\Big)^{-1} & 0 \end{pmatrix}$$

using block matrix notation and writing e.g. $\chi'\big(\frac{1}{2}p^ip^i\big)$ instead of $\chi'\big(\frac{1}{2}p^ip^i\big)\mathbb{1}$. Because of

$$\mathrm{d}F|_{(\mathbf{0},\mathbf{p})} = \chi'\big(\tfrac{1}{2}p^ip^i\big) \cdot p^j\mathrm{d}p^j$$

we get, therefore,

$$-\mathrm{d}F \circ J_{T^*Q}|_{(\mathbf{0},\mathbf{p})} = p^j\mathrm{d}q^j|_{(\mathbf{0},\mathbf{p})} = \lambda_{(\mathbf{0},\mathbf{p})}$$

as claimed.

5.5. Interpolating geodesic and normalised geodesic flow

We choose the strictly increasing function $\chi\colon \mathbb{R} \to \mathbb{R}$ from Sect. 5.2 to satisfy $\chi(t) = t$ for $t \le \frac{1}{4}$ and $\chi(t) = \sqrt{2t}$ for $t \ge \frac{1}{2}$ to interpolate the kinetic energy with the length function.

We would like to understand the interpolation given by χ in terms of symplectisation. For that, we consider the diffeomorphism

$$\begin{array}{rccc} \Phi : & \big(\mathbb{R} \times ST^*Q, \mathrm{e}^a\alpha\big) & \longrightarrow & \big(T^*Q \setminus Q, \lambda\big) \\ & (a,\mathbf{w}) & \longmapsto & \mathrm{e}^a\mathbf{w} \end{array}$$

of Liouville manifolds, where $\alpha := \lambda|_{TST^*Q}$. Observe, that

$$\Phi^*F(a,\mathbf{w}) = \chi \circ k\big(\mathrm{e}^a\mathbf{w}\big) = \chi\big(\tfrac{1}{2}\mathrm{e}^{2a}\big)$$

equals e^a for $a \ge 0$. Since Φ is a symplectomorphism $I := \Phi^*J_{T^*Q}$ is a compatible almost complex structure on the symplectisation $\big(\mathbb{R}\times ST^*Q, \mathrm{d}(\mathrm{e}^a\alpha)\big)$. Moreover, on the positive part $\{a > 0\}$ of the symplectisation, where $\Phi^*F = \mathrm{e}^a$, we obtain $\Phi^*\mathrm{d}F = \mathrm{e}^a\mathrm{d}a$. Therefore,

$$\mathrm{e}^a\alpha = \Phi^*\lambda = \Phi^*\big(-\mathrm{d}F \circ J_{T^*Q}\big) = -\mathrm{e}^a\mathrm{d}a \circ I,$$

which implies

$$\alpha = -\mathrm{d}a \circ I.$$

Consequently, I preserves the contact structure $\xi = \ker\alpha \cap \ker(\mathrm{d}a)$ induced by α on all slices. Moreover, denoting the Reeb vector field of α by R we get

$$1 = \alpha(R) = -\mathrm{d}a(IR).$$

Hence,

$$I\partial_a = R.$$

We remark that ∂_a is the Liouville vector field of $\big(\mathbb{R}\times ST^*Q, \mathrm{e}^a\alpha\big)$. Therefore, $\Phi_*\partial_a = Y$, where Y is the Liouville vector field on T^*Q determined by $\lambda = i_Y \mathrm{d}\lambda$.

We claim that the almost complex structure I is invariant under translation in $\mathbb{R}$-direction along $\mathbb{R}^+\times ST^*Q$. Indeed, using local Riemann coordinates as in Sect. 5.4 the restriction of J_{T^*Q} to $\{|\mathbf{p}|_\flat > 1\}$ is given by

$$\big(J_{T^*Q}\big)_{(\mathbf{0},\mathbf{p})} = \begin{pmatrix} 0 & \frac{1}{|\mathbf{p}|} \\ -|\mathbf{p}| & 0 \end{pmatrix}$$

abbreviating, e.g. $|\mathbf{p}| = |\mathbf{p}|_\flat \mathbb{1}$. As the flow of Y scales by e^t in $\mathbf{p}$-direction the pullback of J_{T^*Q} with respect to the flow of $Y = \mathbf{p}\partial_{\mathbf{p}}$ at $(\mathbf{0},\mathbf{p})$ equals

$$\begin{pmatrix} \mathbb{1} & 0 \\ 0 & \mathrm{e}^{-t} \end{pmatrix} \begin{pmatrix} 0 & \frac{\mathrm{e}^{-t}}{|\mathbf{p}|} \\ -\mathrm{e}^t|\mathbf{p}| & 0 \end{pmatrix} \begin{pmatrix} \mathbb{1} & 0 \\ 0 & \mathrm{e}^{t} \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{|\mathbf{p}|} \\ -|\mathbf{p}| & 0 \end{pmatrix}.$$

This shows that the Lie derivative $L_Y J_{T^*Q}$ vanishes. Hence, $\Phi_*\partial_a = Y$ implies $L_{\partial_a} I = 0$, i.e. $I_{(a,\mathbf{p})} = I_{(a+t,\mathbf{p})}$ for all $a, a+t > 0$.

In other words, I is a compatible almost complex structure on the positive part of the symplectisation $\big(\mathbb{R}^+ \times ST^*Q, \mathrm{d}(\mathrm{e}^a\alpha)\big)$. I is translation invariant, preserves the contact structure $\xi = \ker\alpha$, and sends the Liouville vector field ∂_a to the Reeb vector field R of α.

6. A boundary value problem

Following [19, Section 3] and [20, Section 3] we introduce the moduli space $\mathcal{W}$ of holomorphic discs to understand the topology of the manifold M. We consider the glued strict contact manifold $(\hat{Z}, \hat{\alpha})$ introduced in Sect. 3.1 and form its symplectisation (W, ω), i.e. we set

$$(W,\omega) := \Big(\mathbb{R}\times\hat{Z}, \mathrm{d}(\tau\hat{\alpha})\Big)$$

for a positive, strictly increasing smooth function τ defined on $\mathbb{R}$ such that $\tau(a) = \mathrm{e}^a$ for all $a \geq 0$. Compared to the constructions in [19,20], there will be a substantial difference in setting up the boundary conditions for the holomorphic discs.

6.1. An almost complex structure

We denote by $\hat{\xi}$ the contact structure defined by $\hat{\alpha}$. On the symplectisation (W,ω) we choose a compatible almost complex structure J that is $\mathbb{R}$-invariant, sends ∂_a to the Reeb vector field of $\hat{\alpha}$, and restricts to a complex bundle structure on $\big(\hat{\xi}, \mathrm{d}\hat{\alpha}\big)$.

To incorporate standard holomorphic discs we define the **box** B by

$$B := [-b_0, b_0]\times D_R T^*Q \times D^2_r \times D^{2n-2-2d}_R,$$

where $0 < b_0, r \in (0,1), 1 \leq R$ are real numbers chosen such that $S \subset \operatorname{Int} B$. Here, $D^{2\ell}_\rho \subset \mathbb{C}^\ell$ denotes the closed 2ℓ-disc of radius ρ and $D_\rho T^*Q$ is the closed ρ-disc subbundle of T^*Q. Set

$$\hat{B} := (B \setminus \operatorname{Int} D) \cup_\varphi M.$$

We require the almost complex structure J to be the one defined in Sect. 4 on the complement of $\mathbb{R} \times \operatorname{Int}(\hat{B})$ in $\mathbb{R} \times \hat{Z}$. On $\mathbb{R} \times \operatorname{Int}(\hat{B})$ we will choose J generically, see Sect. 8.

6.2. The moduli space

The **moduli space** $\mathcal{W}$ is the set of all holomorphic discs

$$u = (a, f) \colon \mathbb{D} \longrightarrow (W, J)$$

that satisfy the following conditions:

(w_1) There exists a **level** $(q, \mathbf{t}) \in Q \times \mathbb{R}^{n-1-d}$ such that

$$u(\partial\mathbb{D}) \subset L^{\mathbf{t}}_q .$$

(w_2) There exist $b \in \mathbb{R}$, $\mathbf{w} \in T^*_q Q$, $\mathbf{s} \in \mathbb{R}^{n-1-d}$ such that

$$[u] = [u^{\mathbf{t},\mathbf{w}}_{\mathbf{s},b}] \in H_2(W, L^{\mathbf{t}}_q) ,$$

where $(q, \mathbf{t})$ is the level of u.

(w_3) u maps the marked points $1, \mathrm{i}, -1$ to the characteristic leaves $L^{\mathbf{t}}_q \cap \{z_0 = 1\}$, $L^{\mathbf{t}}_q \cap \{z_0 = \mathrm{i}\}$, and $L^{\mathbf{t}}_q \cap \{z_0 = -1\}$, resp., i.e. for $k = 0, 1, 2$ we have

$$f(\mathrm{i}^k) \in \mathbb{R} \times T^*_q Q \times \{\mathrm{i}^k\} \times \mathbb{R}^{n-1-d} \times \{\mathbf{t}\} .$$

The parameters $b, \mathbf{w}, \mathbf{s}$ in condition (w_2) are assumed to be sufficiently large so that the standard disc $u^{\mathbf{t},\mathbf{w}}_{\mathbf{s},b}$ defines a holomorphic disc in (W, J). With Sect. 4.5 the relative homology class of $u^{\mathbf{t},\mathbf{w}}_{\mathbf{s},b}$ is independent of the choice of $b, \mathbf{w}, \mathbf{s}$.

6.3. Uniform energy bounds

The **symplectic energy** $\int_{\mathbb{D}} u^*\omega$ is bounded by π for all $u = (a, f) \in \mathcal{W}$. Indeed, by Stokes theorem, the symplectic energy of u is equal to the **action** $\int_{\partial\mathbb{D}} f^*\hat{\alpha}$ of the boundary circle. This also holds for any standard disc homologous to u. The claim follows as the symplectic energy is the same for all holomorphic discs of the same level and as the action of the boundary circle of standard discs equals π.

By a similar argument, we obtain that the symplectic energy of any non-constant holomorphic disc that takes boundary values in some Lagrangian cylinder $L^{\mathbf{t}}_q$ is a positive multiple of π.

6.4. Maximum principle

Let $u = (a, f) \in \mathcal{W}$ be a holomorphic disc of level $(q, \mathbf{t})$. By [19, Lemma 3.6.(i)], the function a is subharmonic and, hence, $a < 0$ on $\operatorname{Int} \mathbb{D}$.

The set $G := f^{-1}\big(\hat{Z} \setminus \hat{B}\big)$ is an open subset of $\mathbb{D}$ that contains a neighbourhood of $\partial\mathbb{D}$ in $\mathbb{D}$. Restricting f to G, we can write

$$f = \big(b, \mathbf{w}, h_0, \mathbf{h}\big)$$

w.r.t. coordinate functions on $\mathbb{R} \times T^*Q \times \mathbb{D} \times \mathbb{C}^{n-1-d}$. As the Niederkrüger map is biholomorphic, the function b is harmonic and the maps $\mathbf{w}, h_0, \mathbf{h}$ are holomorphic, see Sect. 4.3.

In particular, if $G = \mathbb{D}$, then u is one of the discs $u^{\mathbf{t},\mathbf{w}}_{\mathbf{s},b}$. This follows as in [19, Lemma 3.7]. Simply use the fact that a holomorphic map $\mathbf{w}\colon \mathbb{D} \to T^*Q$ with boundary on T^*_qQ is constant by Stokes theorem and $\mathbf{w}^*\lambda = 0$ on $\partial\mathbb{D}$.

Motivated by this, we introduce the notion of standard holomorphic discs to the glued manifold W:

Definition 6.1. A holomorphic disc $u = (a, f) \in \mathcal{W}$ is a called a **standard disc** if $f(\mathbb{D}) \subset \hat{Z} \setminus \operatorname{Int} \hat{B}$. Holomorphic discs $u = (a, f) \in \mathcal{W}$ with $f(\mathbb{D}) \cap \operatorname{Int} \hat{B} \neq \emptyset$ are called **non-standard**.

Applying the strong maximum principle and the boundary lemma by E. Hopf to h_0 we obtain as in [19, Lemma 3.6.(ii)] and on [19, p. 669 and p. 671]:

(1) $f(\operatorname{Int} \mathbb{D}) \subset \operatorname{Int} \hat{Z}$.
(2) $u|_{\partial\mathbb{D}}$ is an embedding.

Remark 6.2. In the situation, u is a non-constant holomorphic disc (W, J) that satisfies just the boundary condition $u(\partial\mathbb{D}) \subset L^{\mathbf{t}}_q$ the conclusions from this section that rely on the maximum principle continue to hold. The corresponding replacement of the statement in (2) which does not use the homological assumption is the following: h_0 restricts to an immersion on $\partial\mathbb{D}$ so that $u(\partial\mathbb{D})$ is positively transverse to each of the characteristic leaves $L^{\mathbf{t}}_q \cap \{z_0 = \mathrm{e}^{\mathrm{i}\theta}\}$, $\theta \in [0, 2\pi)$.

Remark 6.3. The monotonicity argument used in [19, Lemma 3.9] implies that there exists a compact ball $K \subset \mathbb{C}^{n-1-d}$ such that $\mathbf{h}(G) \subset K$ for all non-standard disc $u \in \mathcal{W}$, i.e. with $u = (a, f)$ we have

$$f^{-1}\Big(\mathbb{R} \times T^*Q \times \mathbb{D} \times \big(\mathbb{C}^{n-1-d} \setminus K\big)\Big) = \emptyset.$$

6.5. Integrated maximum principle

Let $u = (a, f) \in \mathcal{W}$ be a holomorphic disc of level $(q, \mathbf{t})$. As in Sect. 6.4 we consider $G := f^{-1}\big(\hat{Z} \setminus \hat{B}\big)$ so that we can write $f = \big(b, \mathbf{w}, h_0, \mathbf{h}\big)$ on G. In Sect. 6.4 we obtained uniform C^0-bounds on h_0 and $\mathbf{h}$ relying on the maximum principle from [19,20]. As the boundary conditions in T^*Q-direction are considerably different form the one used in [20] uniform C^0-bounds on $\mathbf{w}$ require a new argument.

First of all we remark that by Stokes theorem, the symplectic energy of u (which we computed in Sect. 6.3 to be equal to π) is equal to the area $\int_{\mathbb{D}} f^*\mathrm{d}\hat{\alpha}$

of f. Because $f^*\mathrm{d}\hat{\alpha}$ is an area density by our compatibility assumptions we obtain

$$\int_G \mathbf{w}^*\mathrm{d}\lambda \leq \int_G f^*\mathrm{d}\alpha_0 \leq \pi.$$

Recall the diffeomorphism $\Phi\colon \big(\mathbb{R}\times ST^*Q, \mathrm{e}^a\alpha\big) \to \big(T^*Q\setminus Q, \lambda\big)$ of Liouville manifolds from Sect. 5.5, which pulls J_{T^*Q} back to I. Define $v := \Phi^{-1}\circ\mathbf{w}$ and replace G by the subset $(|\mathbf{w}|)^{-1}\big((R,\infty)\big)$, $R\geq 1$ appearing in the definition of the box in Sect. 6.1, so that

$$v = (c,k)\colon\ G \longrightarrow (\ln R,\infty)\times ST^*Q$$

is an I-holomorphic map subject to the following boundary conditions:

$$c\big(\partial G\setminus\partial\mathbb{D}\big) = \{\ln R\},\quad k\big(\partial\mathbb{D}\cap G\big)\subset ST^*_qQ.$$

Further we have

$$\int_G v^*\mathrm{d}(\mathrm{e}^a\alpha)\leq\pi$$

for the symplectic energy of v.

We consider the subdomain

$$G_t := c^{-1}\big((t,\infty)\big)$$

of G for $t\geq\ln R$. Note that $G_{\ln R}=G$. In order to allow partial integration we denote by $\mathcal{R}$ the set of all regular values $t\in(\ln R,\infty)$ of the functions c and $c|_{\partial\mathbb{D}\cap G}$. By Sard's theorem $\mathcal{R}$ has full measure. Therefore, the open set $\mathcal{R}$ is dense in $(\ln R,\infty)$.

For $t\in\mathcal{R}$ the domain G_t has piecewise smooth boundary

$$\partial G_t = \partial\mathbb{D}\cap G_t + \partial G_t\setminus\partial\mathbb{D},$$

which we equip with the boundary orientation. Up to a null set the interior boundary $\partial G_t\setminus\partial\mathbb{D}$ is given by $c^{-1}(t)$. Observe that ST^*_qQ is a Legendrian sphere in the unit cotangent bundle so that the restrictions of $k^*\alpha$ to the tangent spaces of $\partial\mathbb{D}\cap G_t$ vanish. Stokes theorem applied twice implies

$$\int_{G_t} v^*\mathrm{d}(\mathrm{e}^a\alpha) = \mathrm{e}^t\int_{c^{-1}(t)} k^*\alpha = \mathrm{e}^t\int_{G_t} k^*\mathrm{d}\alpha,$$

where we used $v^*\mathrm{d}(\mathrm{e}^a\alpha)=\mathrm{d}(\mathrm{e}^c k^*\alpha)$.

On the other hand, using Leibniz rule, we have a decomposition

$$v^*\mathrm{d}(\mathrm{e}^a\alpha) = \mathrm{e}^c\mathrm{d}c\wedge k^*\alpha + \mathrm{e}^c k^*\mathrm{d}\alpha$$

into energy densities. Define the α-**energy** functional by

$$e(t) := \int_{G_t}\mathrm{e}^c\mathrm{d}c\wedge k^*\alpha\geq 0.$$

Therefore,

$$\int_{G_t} v^*\mathrm{d}(\mathrm{e}^a\alpha) = e(t) + \int_{G_t}\mathrm{e}^c k^*\mathrm{d}\alpha \geq e(t) + \mathrm{e}^t\int_{G_t} k^*\mathrm{d}\alpha$$

using $\mathrm{e}^c\geq\mathrm{e}^t$ on G_t.

Combining these expressions for the symplectic energy, we get $e(t) \leq 0$. Hence, $e(t) = 0$ for all $t \in \mathcal{R}$, i.e. the α-energy functional $e = e(t)$ vanishes identically. Because of

$$\mathrm{d}c \wedge f^*\alpha = \left(c_x^2 + c_y^2\right) \mathrm{d}x \wedge \mathrm{d}y,$$

we deduce that $c|_{G_t} = \text{const}$ and, since $k^*\alpha = -\mathrm{d}c \circ \mathrm{i}$, that $k^*\alpha|_{G_t} = 0$ as well as $k^*\mathrm{d}\alpha|_{G_t} = 0$. We conclude that $v|_{G_t} = \text{const}$ for all $t \in (\ln R, \infty)$. An open and closed argument for $G = (|\mathbf{w}|)^{-1}\big((R,\infty)\big)$ implies that either $G = \emptyset$ or $v = \text{const}$ on all of $G = \mathbb{D}$, which in turn implies that $u \in \mathcal{W}$ was a standard disc. This shows uniform C^0-bounds in T^*Q-direction for all non-standard discs $u \in \mathcal{W}$:

Proposition 6.4. *If $u = (a, f) \in \mathcal{W}$ is a non-standard holomorphic discs, then*

$$f^{-1}\Big(\mathbb{R} \times \big(T^*Q \setminus D_R T^*Q\big) \times \mathbb{D} \times \mathbb{C}^{n-1-d}\Big) = \emptyset.$$

7. Compactness

Consider a non-standard disc $u = (a, f) \in \mathcal{W}$ of level $(q, \mathbf{t})$. On the preimage $G := f^{-1}\big(\hat{Z} \setminus \hat{B}\big)$, we write

$$f = \big(b, \mathbf{w}, h_0, \mathbf{h}\big)$$

w.r.t. to the decomposition $\mathbb{R} \times T^*Q \times \mathbb{D} \times \mathbb{C}^{n-1-d}$. In Sects. 6.4 and 6.5, we obtained uniform bounds on

(i) a from above by 0,
(ii) h_0 in the sense $|h_0| \leq 1$,
(iii) $\mathbf{w}$ and $\mathbf{h}$ in the sense that $|\mathbf{w}|_\flat$ and $|\mathbf{h}|$, resp., are bounded by a geometric constant.

The coordinate function b completes to a holomorphic function

$$a - F(\mathbf{w}) - \frac{1}{4}|h_0|^2 - \frac{1}{2}|\operatorname{Im} \mathbf{h}|^2 + \mathrm{i}b$$

on G, where the restriction of the real part to $\partial\mathbb{D}$ equals $F(\mathbf{w})|_{\partial\mathbb{D}}$ up to a constant. In [19, Lemma 3.8], where no T^*Q-component appears, we used Schwarz reflection and the maximum principle to establish uniform bounds on $|b|$. In our situation this would require real analyticity of $F(\mathbf{w})|_{\partial\mathbb{D}}$, which in general does not hold.

We will work around this utilising a bubbling off analysis that uses target rescaling along the Reeb vector field ∂_b on $\hat{Z} \setminus \hat{B}$. This will require ideas from [21]. In fact, by the elliptic nature of the holomorphic curves equation the bubbling off analysis directly yields compactness properties of holomorphic curves. Therefore, we will combine the target rescaling in b-direction with the usual target rescaling along the Liouville vector field ∂_a:

By the maximum principle $|b|$ attains its maximum on ∂G. Observe that because of $f(\partial\mathbb{D}) \subset \hat{Z} \setminus \hat{B}$ the boundary of G decomposes

$$\partial G = \partial\mathbb{D} \sqcup f^{-1}(\partial\hat{B}).$$

Assuming $|b| \not\leq b_0$ we get therefore that $|b|$ attains its maximum on $\partial\mathbb{D}$.

Suppose there exist sequences $\zeta_\nu \in \mathbb{D}$ and $u_\nu = (a_\nu, f_\nu) \in \mathcal{W}$ of non-standard such that

$$|b_\nu(\zeta_\nu)| \longrightarrow \infty$$

writing $f_\nu = \big(b_\nu, \mathbf{w}_\nu, h_0^\nu, \mathbf{h}_\nu\big)$. We may assume that $\zeta_\nu \in \partial\mathbb{D}$ for all ν and that $\zeta_\nu \to \zeta_0$ in $\partial\mathbb{D}$. By the mean value theorem, we find a sequence z_ν in $\mathbb{D}$ such that $|\nabla u_\nu(z_\nu)| \to \infty$. This implies that uniform gradient bounds for non-standard holomorphic discs in $\mathcal{W}$ result in uniform bounds on b.

Proposition 7.1. *Under the assumptions of Theorem 2.1 each sequence of non-standard discs $u_\nu \in \mathcal{W}$ has a C^∞-converging subsequence.*

Proof. Consider a sequence of non-standard discs $u_\nu = (a_\nu, f_\nu) \in \mathcal{W}$ of level $(q_\nu, \mathbf{t}_\nu)$ such that $|\nabla u_\nu(z_\nu)| \to \infty$ for a sequence $z_\nu \to z_0$ in $\mathbb{D}$. By compactness of Q and Remark 6.3 we can assume that $(q_\nu, \mathbf{t}_\nu) \to (q_0, \mathbf{t}_0)$. Observe that modifications as made in [19, Section 4.1] that fix the varying boundary conditions we will mention in Sect. 8.3 are not necessary for the following compactness argument.

Up to a choice of a subsequence, we distinguish two cases:

(1) $f_\nu(z_\nu) \in \hat{Z} \setminus \hat{B}$ for all ν, and
(2) $f_\nu(z_\nu) \in \hat{B}$ for all ν.

In the first case, additionally, we can assume that the sequences $\mathbf{w}_\nu(z_\nu)$, $h_0^\nu(z_\nu)$, and $\mathbf{h}_\nu(z_\nu)$ converge and that either

(1.1) $b_\nu(z_\nu) \to \pm\infty$, or
(1.2) $b_\nu(z_\nu) \to b_\infty \in \mathbb{R}$.

In case (1.1), we use bubbling off analysis as in [26, Section 6], but this time applied to the holomorphic maps

$$\big(a_\nu - a_\nu(z_\nu), b_\nu - b_\nu(z_\nu), \mathbf{w}_\nu, h_0^\nu, \mathbf{h}_\nu\big)$$

defined on $G_\nu := f_\nu^{-1}\big(\hat{Z} \setminus \hat{B}\big)$ for interior bubbling; for bubbling along the boundary perform the shift w.r.t. the real parts x_ν of the z_ν. For both observe that shift in b-direction is a strict contactomorphism of (Z, α_0) and does not effect the Hofer energy. To have enough space inside G_ν during the domain rescaling use the trick in [26, Case 1.2.b] explained on [26, p. 547]; this time make use of the stretching of the holomorphic discs u_ν in b-direction instead of the a-direction. In the cases (2) and (1.2) apply the usual bubbling off analysis as in [9,27–29], cf. [26, Cases 1.1, 1.2.a, 2 in Section 6].

Finally, in all cases, we can argue as in [19, Section 4]. By the aperiodicity assumption $\inf_0(\alpha) \geq \pi$, which with Sect. 3.1 implies $\inf_0(\hat{\alpha}) \geq \pi$, there is no bubbling off of finite energy planes. This is because finite energy planes asymptotically converge to contractible periodic Reeb orbits. The asymptotic analysis of the finite energy planes possibly requires a bubbling off analysis that involves target rescaling in b-direction as explained above, cf. [21, Section 5.2].

Because there are no bubble spheres by exactness of (W, ω) we are left with bubbling off of holomorphic discs, cf. [21, Section 5.3]. This will lead us to a contradiction as in [19, Section 4.2]. Indeed, the Hofer energy of a bubble discs is a positive multiple of π, see Sect. 6.3. As the Hofer energy of all u_ν equals π by Sect. 6.3 there is at most one bubble discs. Hence, we can assume that u_ν converge in C^∞_{loc} on $\mathbb{D}\setminus\{z_0\}$ for some $z_0 \in \partial\mathbb{D}$. By our assumption on the 3 fixed marked points in the definition of $\mathcal{W}$ after removing the singularity z_0 the limiting holomorphic disc will be non-constant; and, therefore, will also have energy equal to a positive multiple of π. But the sum of energies of the bubble disc and the limiting disc can not exceed π. This contradiction shows uniform gradient bounds for any sequence u_ν of holomorphic discs in $\mathcal{W}$. $\square$

8. Transversality

In Sect. 7, we established properness of the evaluation map

$$\begin{array}{rccc} \mathrm{ev}\colon & \mathcal{W}\times\mathbb{D} & \longrightarrow & \hat{Z} \\ & \big(u=(a,f),z\big) & \longmapsto & f(z). \end{array}$$

The aim of this section is to show that ev has degree 1. We will follow the considerations from [19, Section 5] and [20, Section 3.5] and just indicate the adaptations to the present situation.

8.1. Maslov index

For all $u\in\mathcal{W}$ the Maslov index of the bundle pair

$$\big(u^*TW,(u|_{\partial\mathbb{D}})^*TL_q^{\mathbf{t}}\big)$$

equals 2, where $(\mathbf{t},q)$ is the level of u. Indeed, following [19, Lemma 3.1], by homotopy invariance it suffices to show the claim for standard discs

$$u(z)=u_{\mathbf{s},b}^{\mathbf{t},\mathbf{w}}(z)=\Big(\tfrac{1}{4}\big(|z|^2-1\big),b\,,\mathbf{w},z,\mathbf{s}+\mathrm{i}\mathbf{t}\Big),$$

$\mathbf{w}\in T_qQ$, assuming $W=\mathbb{R}\times\mathbb{R}\times T^*Q\times\mathbb{D}\times\mathbb{C}^{n-1-d}$. In particular, $u^*TW\cong\underline{\mathbb{C}}^{n+1}$. Moreover, $(u|_{\partial\mathbb{D}})^*TL_q^{\mathbf{t}}$ is isomorphic to $\mathrm{i}\mathbb{R}\oplus\mathrm{i}\mathbb{R}^d\oplus\mathrm{e}^{\mathrm{i}\theta}\mathbb{R}\oplus\mathbb{R}^{n-1-d}$ over $\mathrm{e}^{\mathrm{i}\theta}\in\partial\mathbb{D}$. Hence, the Maslov index equals 2 by normalisation.

8.2. Simplicity

First of all, we remark that the classes $[u]\in H_2(W,L_q^{\mathbf{t}})$, $u\in\mathcal{W}$, are J-indecomposable. Otherwise, we would find a decomposition

$$[u]=\sum_{j=1}^N m_j[v_j]$$

in $H_2(W,L_q^{\mathbf{t}})$, for simple holomorphic discs v_j with boundary on $L_q^{\mathbf{t}}$ and multiplicities $m_j\geq 1$. Writing $v_j=(a_j,f_j)$ we get for the energy

$$\pi=\sum_{j=1}^N m_j\int_{\partial\mathbb{D}} f_j^*\alpha_0.$$

Writing $\big(b_j, \mathbf{w}_j, h_0^j, \mathbf{x}_j + \mathrm{i}\mathbf{t}_j\big)$ for the restriction of $f_j|_{\partial\mathbb{D}}$ the left hand side reads as

$$\sum_{j=1}^{N} m_j \int_{\partial\mathbb{D}} \Big[b_j^*\mathrm{d}b + \mathbf{w}_j^*\lambda + (h_0^j)^*\tfrac{1}{2}\big(x_0\mathrm{d}y_0 - y_0\mathrm{d}x_0\big) - (\mathbf{x}_j + \mathrm{i}\mathbf{t}_j)^*(\mathbf{y}\mathrm{d}\mathbf{x})\Big].$$

The first and last summand vanish by exactness of the form we pull back to the circle $\partial\mathbb{D}$; the second vanishes because $\mathbf{w}_j(\partial\mathbb{D}) \subset T_q^*Q$. Hence, writing r_j for the winding number of $h_0^j|_{\partial\mathbb{D}}$, which is positive for non-constant h_0^j by the argument principle, we get

$$\pi = \pi \cdot \sum_{j=1}^{N} m_j r_j \geq N \cdot \pi.$$

We conclude that $N = 1$, $m_1 = 1$, i.e. $[u]$ is J-indecomposable.

Consulting [19, Lemma 3.4] we see that u must be simple. Because $u|_{\partial\mathbb{D}}$ is an embedding, see Sect. 6.4, we obtain as in [19, Lemma 3.5] that the set of f-injective points is open and dense in $\mathbb{D}$.

8.3. Variable boundary conditions

There is a natural way to identify the boundary conditions

$$L_q^{\mathbf{t}} = \{0\} \times \mathbb{R} \times T_q^*Q \times \partial\mathbb{D} \times \mathbb{R}^{n-1-d} \times \{\mathbf{t}\}$$

for the holomorphic discs in $\mathcal{W}$. Observe, that the union of $L_q^{\mathbf{t}}$ over all parameters $\mathbf{t} \in \mathbb{R}^{n-1-d}$ and $q \in Q$ equals

$$\{0\} \times \partial\hat{Z} = \{0\} \times \mathbb{R} \times T^*Q \times \partial\mathbb{D} \times \mathbb{C}^{n-1-d}$$

so that flows induced by tangent vectors $\mathbf{v} \in T_{\mathbf{t}}\mathbb{R}^{n-1-d}$ and $v \in T_qQ$ can be taken for the identifications: Consider a chart $(\mathbb{R}^d, 0) \to (Q, q)$ of Q about q and extend v to a vector field on $\mathbb{R}^d$ that has compact support and is constant near 0. The induced flow on Q naturally lifts to a fibre and Liouville form preserving flow on T^*Q, see [30, p. 92]. Similarly, extend $\mathbf{v} \in T_{\mathbf{t}}\mathbb{R}^{n-1-d}$ to a compactly supported vector field on $\mathbb{R}^{n-1-d}$ that is constant near $\mathbf{t} \in \mathbb{R}^{n-1-d}$.

We regard $(v, \mathbf{v})$ as a vector field on $\mathbb{R}\times\mathbb{R}\times T^*Q\times\partial\mathbb{D}\times\mathbb{C}^{n-1-d}$ cutting off $(v, \mathbf{v})$ with a bump function that has support on a small neighbourhood of $\{0\} \times [-b_0, b_0] \times T^*Q \times \partial\mathbb{D} \times \mathbb{C}$ and equals 1 on a smaller neighbourhood. We denote the corresponding flow on W by $\psi_t^{(v,\mathbf{v})}$. Given a level $(q_0, \mathbf{t}_0)$ we find a neighbourhood U of $(q_0, \mathbf{t}_0) \in Q \times \mathbb{R}^{n-1-d}$ and a vector field $(v, \mathbf{v})$ as above such that the time-1 map $\psi_1^{(v,\mathbf{v})}$ sends $L_{q_0}^{\mathbf{t}_0}$ to $\psi_1^{(v,\mathbf{v})}(L_{q_0}^{\mathbf{t}_0}) = L_q^{\mathbf{t}}$ for all $(q, \mathbf{t}) \in U$. Simply define $(v, \mathbf{v})$ to be $(q - q_0, \mathbf{t} - \mathbf{t}_0)$ on U.

8.4. Admissible functions

Denote by $\mathcal{B}$ the separable Banach manifold consisting of all continuous maps $u : (\mathbb{D}, \partial\mathbb{D}) \to \big(W, \{0\} \times \hat{C}\big)$ of Sobolev class $W^{1,p}$, $p > 2$, that satisfy the conditions (w_1)–(w_3) in the definition of the moduli space $\mathcal{W}$, see Sect. 6.2.

The Banach manifold structure is given as follows: The subset $\mathcal{B}_q^{\mathbf{t}} \subset \mathcal{B}$ of all u of level $(q, \mathbf{t})$ is a separable Banach manifold whose tangent spaces are

$$T_u\mathcal{B}_q^{\mathbf{t}} = W^{1,p}\big(u^*TW, (u|_{\partial\mathbb{D}})^*TL_q^{\mathbf{t}}\big).$$

Consider the level projection map $\mathcal{B} \to Q \times \mathbb{R}^{n-1-d}$ that assigns to all $u \in \mathcal{B}$ the corresponding level $(q, \mathbf{t})$. Using the identifying maps the $\psi_1^{(v,\mathbf{v})}$ from Sect. 8.3 these defines a locally trivial fibration on the Banach manifold $\mathcal{B}$ with fibres $\mathcal{B}_q^{\mathbf{t}}$.

8.5. Linearised Cauchy–Riemann operator

In particular,

$$T_u\mathcal{B} = T_u\mathcal{B}_q^{\mathbf{t}} \oplus \big(T_qQ \oplus \mathbb{R}^{n-1-d}\big)$$

so that the linearised Cauchy–Riemann operator at $u \in \mathcal{B}$ of level $(q, \mathbf{t})$ splits as

$$D_u = D_u^{(q,\mathbf{t})} \oplus K_u,$$

where $D_u^{(q,\mathbf{t})} := D_u|_{T_u\mathcal{B}_q^{\mathbf{t}}}$ is the linearised Cauchy–Riemann operator in fibre direction and $K_u : T_qQ \oplus \mathbb{R}^{n-1-d} \to L^p(u^*TW)$ is a compact perturbation, see [19, Section 5.1]. The index of $D_u^{(q,\mathbf{t})}$ equals n, as the Maslov index of the problem with fixed boundary level was 2 (see Sect. 8.1), so that the total index equals $\operatorname{ind} D_u = 2n - 1$.

If Q is oriented, we can orient D_u via the determinant bundle

$$\det D_u = \det D_u^{(q,\mathbf{t})} \otimes \det\big(T_qQ \oplus \mathbb{R}^{n-1-d}\big)$$

as follows: The line bundle $\det D_u^{(q,\mathbf{t})}$ is oriented by the construction in [31, Section 8.1] via the trivial bundle $TL_q^{\mathbf{t}} \cong T_q^*Q \oplus \underline{\mathbb{R}}^{n+1-d}$ and the orientation of $T_q^*Q \cong \mathbb{R}^d$ so that the bundle pair

$$\big(u^*TW, (u|_{\partial\mathbb{D}})^*TL_q^{\mathbf{t}}\big)$$

admits a natural trivialisation. The line bundle $\det\big(T_qQ \oplus \mathbb{R}^{n-1-d}\big)$ is oriented via the orientation of $Q \times \mathbb{R}^{n-1-d}$.

8.6. Lifting topology

As in [19, Section 5.2], we choose J to be regular by perturbing the induced complex structure on $\hat{\xi}$ over $\hat{B}$. Regularity of J along standard discs is obvious. Hence, the moduli space $\mathcal{W}$ is a smooth oriented manifold of dimension $2n - 1$ whose end is made out of standard holomorphic discs. Therefore, the evaluation map ev, which is proper, has degree 1. With [19, Section 6] and [13, Section 2] we see that ev induces surjections of homology groups and of π_1.

Identify Q with the subset

$$Q \equiv \{0\} \times Q \times \{1\} \times \{0\}$$

of

$$\mathbb{R} \times T^*Q \times \{1\} \times \mathbb{C}^{n-1-d} \subset \partial\hat{Z}.$$

Observe that M is a strong deformation retract of $\hat{Z}$. We choose a deformation retraction such that the inclusion $Q \subset \hat{Z}$ is isotoped to an embedding $Q \to M$. Combining this with the following commutative diagram

$$\begin{tikzcd}
\mathcal{W} \times \mathbb{D} \arrow[rr, "\mathrm{ev}"] & & \hat{Z} \\
\mathcal{W} \times \{1\} \arrow[u, "\subset"'] \arrow[r, "\mathrm{ev}"] & \mathbb{R} \times T^*Q \times \{1\} \times \mathbb{C}^{n-1-d} \arrow[ur, phantom, ""] & \arrow[u, "\subset"']
\end{tikzcd}$$

yields:

Proposition 8.1. *Under the assumptions of Theorem 2.1 the isotoped inclusion $Q \to M$ induces a surjection of homology and fundamental groups.*

Proof. This follows with the homology epimorphism argument from [13, Section 2.3] and the covering argument from [13, Section 2.5]. □

9. The homotopy type

We compute the homotopy type of M in terms of $D\big(T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}\big)$. For that we assume that, up to fibre preserving isotopy, the shape S is equal to the shape given by the unit sphere bundle in $T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}$. This results into the same construction for $\hat{Z}$ as in Sect. 3.1 up to ambient diffeotopy.

We identify Q with the section of the sphere bundle

$$\partial M = S\big(T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}\big)$$

given by

$$Q \equiv \{0\} \times Q \times \{1\} \times \{0\}$$

in

$$\mathbb{R} \times T^*Q \times \mathbb{D} \times \mathbb{C}^{n-1-d}.$$

Observe that this defines a natural embedding of $D\big(T^*Q \oplus \underline{\mathbb{R}}^{2n+1-2d}\big)$ into M via a small disc bundle about

$$\{0\} \times Q \times \{(1-\varepsilon)\} \times \{0\},$$

$\varepsilon > 0$ small. Indeed, simply shift a small disc bundle in $\mathbb{R} \times T^*Q \times \mathbb{D} \times \mathbb{C}^{n-1-d}$ in direction of $\{0\} \times Q \times \{(1-\varepsilon)\} \times \{0\}$. The image is denoted by M_0.

9.1. Homology type and fundamental group

Proposition 8.1 implies that the inclusion $Q \subset M$ is surjective in homology and π_1. Based on that we show:

Proposition 9.1. *Under the assumptions of Theorem 2.1, the inclusion $M_0 \subset M$ induces isomorphisms of homology groups.*

Proof. The arguments are similar to [20, p. 42] and [13, Section 2.4]. Recall the general assumption $n-1 \geq d$.

From Proposition 8.1, we immediately obtain $H_k M = 0$ for $k \geq d+1$ so that the homology isomorphism property of the inclusion $M_0 \subset M$ is automatic in all degrees $k \geq d+1$.

By general position, any section $Q \to \partial M$ of the sphere bundle induces an isomorphism in homology in degree $k \leq 2n-1-d$. Therefore, the inclusion of the sphere bundle into the disc bundle of $T^*Q \oplus \mathbb{R}^{2n+1-2d}$ is isomorphic in homology of degree $k \leq 2n-1-d$. We claim that the inclusion $\partial M \to M$ shares the same property. With $d+1 \leq 2n-1-d$, the proposition will be immediate.

By Poincaré duality and the universal coefficient theorem, we have

$$H_k(M, \partial M) \cong H^{2n+1-k}M \cong FH_{2n+1-k}M \oplus TH_{2n-k}M,$$

where FH_* and TH_* denote the free and the torsion part of H_*, respectively. By the above $H_k(M, \partial M) = 0$ for $k \leq 2n-d-1$. The long exact sequence of the pair $(M, \partial M)$ implies that $\partial M \to M$ is isomorphic in degree $k \leq 2n-2-d$ and epimorphic in degree $k = 2n-1-d$. Because the homology of the sphere bundle ∂M vanishes in degree $k = 2n-1-d$ the epimorphism is in fact injective. □

Corollary 9.2. *Under the assumptions of Theorem 2.1 the inclusion $M_0 \subset M$ induces an epimorphism on fundamental groups. If in addition $\pi_1 Q$ is abelian, then the inclusion $M_0 \subset M$ will be π_1-isomorphic.*

Proof. Using the π_1-isomorphism $M_0 \simeq Q \subset \partial M$, the claim follows from Proposition 8.1 and 9.1 as in [13, Section 2.5]. □

Proof of Theorem 2.1 (i). The claim directly follows from Proposition 9.1 and Corollary 9.2. Simply observe that the specific choice of section into the sphere bundle is irrelevant here. □

9.2. A cobordism

Implementing the construction from [20, Section 4.2] in the situation at hand we define a cobordism

$$X := M \setminus \operatorname{Int} M_0.$$

The construction comes with the following diagram

$$\begin{tikzcd}
 & & & X \arrow[dl, "\text{gen. pos.}"] \\
 & M_0 \arrow[r, "\text{time-1 map}", "\text{of isotopy}"'] & M & \\
\partial M_0 \arrow[ur, "\text{gen. pos.}"'] \arrow[uurrr, dashed, bend left] \arrow[rrr, bend right=80, "\text{time-1 map}", "\text{of former isotopy}"'] & & & \partial M \arrow[ul, Rightarrow] \arrow[uu, dashed] \\
 & Q_0 \arrow[ul] \arrow[uu, "\simeq"'] \arrow[r, "\text{time-1 map}", "\text{of isotopy}"'] & Q \arrow[uu] \arrow[ur] &
\end{tikzcd}$$

that commutes up to homotopy. We explain the diagram: Set

$$Q_0 \equiv \{0\} \times Q \times \{(1-\varepsilon')\} \times \{0\},$$

where $\varepsilon' \in (0, \varepsilon)$ is chosen such that $Q_0 \subset \partial M_0$. All arrows are given by inclusion except those whose label refers to an isotopy. The mentioned isotopy is an isotopy of Q_0 inside M that is the restriction of a diffeotopy on $\mathbb{R} \times T^*Q \times \mathbb{D} \times \mathbb{C}^{n-1-d}$ obtained by shifting and rescaling that brings Q_0 to Q and ∂M_0 to ∂M. The arrow $M_0 \to M$ is obtained from an extension of the isotopy of $Q_0 \subset M$ to M_0.

Proposition 9.3. *Under the assumptions of Theorem 2.1 the inclusion maps $\partial M_0, \partial M \subset X$ induce isomorphisms of homology groups. If in addition $\pi_1 Q$ is abelian (or more generally the inclusion $Q \subset M$ is π_1-injective) then the inclusions $\partial M_0, \partial M \subset X$ will be π_1-isomorphic.*

Proof. The argumentation is the one given at the end of [20, Section 4.2]: For low degrees $k \leq 2n - d - 1$ use general position arguments as indicated in the diagram and the results from Sect. 9.1. In higher degrees $k \geq d + 1$ essentially this is Poincaré duality and excision. □

Proof of Theorem 2.1 part (a) in (ii) and (iii). We have to establish homotopy equivalence, resp., a diffeomorphism between M and M_0. With Proposition 9.3 this essentially follows from the relative Hurewicz and the s-cobordism theorem. The arguments are precisely as in the proof of [13, Theorem 1.5] for Q simply connected and [13, Theorem 5.3] via finite coverings in the non-simply connected case. □

9.3. Infinite coverings

We assume the inclusion map $\partial M \subset M$ to be π_1-injective. This will be satisfied if $\pi_1 Q$ is abelian for example. If Q is simply connected vanishing in relative homology of the cobordism $\{\partial M_0, X, \partial M\}$, which will be simply connected too, implies triviality of relative homotopy groups. If Q is not simply connected, one way to work around this is to lift along the universal covering of X. For $\pi_1 Q$ finite the universal covering space $\widetilde{X}$ will be compact so that we are in the situation of the previous sections. This was used in the proof of Theorem 2.1 part (a) in (ii) and (iii) in Sect. 9.2.

If $\pi_1 Q$ is infinite, we reset the moduli space: The π_1-isomorphism $\partial M \subset M$ ensures that the universal cover of $\hat{Z}$ is obtained by gluing similarly to Sect. 3.1; this time we glue the universal covers of the involved objects along a lift of φ. This makes it possible to consider the moduli space $\mathcal{W}'$ of holomorphic discs in $\widetilde{W}$ defined as in Sect. 6.2; just replace Q with $\widetilde{Q}$ in the definition of the Lagrangian boundary cylinders. This places us into the situation of [20, Section 4.4]. The change of the boundary condition is inessential and the special choice $Q = T^d$ is not really used. Hence, we obtain a covering $\mathcal{W}' \to \mathcal{W}$ together with a proper degree 1 evaluation map

$$\begin{array}{rccc} \mathrm{ev}\colon & \mathcal{W}' \times \mathbb{D} & \longrightarrow & \widetilde{\hat{Z}} \\ & \big(u = (a, f), z\big) & \longmapsto & f(z), \end{array}$$

see [13, Lemma 6.1]. Similar to Proposition 8.1 and [13, Proposition 6.2 and Lemma 6.3] we obtain:

Proposition 9.4. *Under the assumptions of Theorem 2.1 the inclusion $\widetilde{Q} \to \widetilde{M}$ of universal covers induces a surjection of homology and fundamental groups. Further, the inclusion $\partial\widetilde{M}_0 \to \widetilde{X}$ is homology surjective.*

Because the universal cover $\widetilde{X}$ is not compact for $\pi_1 Q$ infinite Poincaré duality delivers no information about relative homology groups in contrary to our argument in Proposition 9.1. But we can say the following:

Theorem 2.1 part (b) in (ii) and (iii). Because the universal cover of Q is contractible so is $\widetilde{M}$ by Proposition 9.4. Hence, the inclusion $\widetilde{M}_0 \subset \widetilde{M}$ is a homotopy equivalence. This follows with the arguments from the proof of [13, Theorem 7.2]. With the proof of [13, Theorem 9.1], which in our situation is particularly easy because of the extra codimension, it follows that the boundary inclusions of $\widetilde{X}$ are homotopy equivalences. Hence, X is in fact an h-cobordism. For the diffeomorphism type, then apply the s-cobordism theorem. □

If ∂M is a simple space, which for example is satisfied whenever Q is a simple space and $\partial M \to Q$ a trivial sphere bundle, then vanishing of relative homology of $(X, \partial M_0)$ and $(X, \partial M)$, resp., implies homotopy equivalence of each of the boundary inclusions of the cobordism $\{\partial M_0, X, \partial M\}$. The basic idea here is that the kernel of the Hurewicz homomorphism is made out of the action of the fundamental group, which we now assume to be trivial, see [13, Section 8]:

Theorem 2.1 part (c) in (ii) and (iii). Follows with the same arguments as in [13, Theorem 1.7 and Example 9.3 (b)]. □

Acknowledgements

This work is inspired by the Ph.D. Thesis [32] of Kilian Barth and the Master Thesis [33] of Jay Schneider who obtained the crucial C^0-bounds via a uniform relative monotonicity lemma in similar situations. We are grateful to Peter Albers, Hansjörg Geiges, and Sefan Suhr for their valuable comments and to the University of Heidelberg for providing such a wonderful work environment that allowed us to finish this work.

References

[1] Gromov, M.: Pseudoholomorphic curves in symplectic manifolds. Invent. Math. **82**, 307–347 (1985)

[2] Eliashberg, Ya.: Filling by holomorphic discs and its applications, In: Geometry of Low-Dimensional Manifolds, vol. 2 (Durham, 1989), London Mathematical Society Lecture Note Series, vol. 151, pp. 45–67. Cambridge University Press, Cambridge (1990)

[3] McDuff, D.: The structure of rational and ruled symplectic 4-manifolds. J. Am. Math. Soc. **3**, 679–712 (1990)

[4] Hind, R.: Holomorphic filling of $\mathbb{R}P^3$. Commun. Contemp. Math. **2**, 349–363 (2000)

[5] Li, T.-J., Mak, C.Y., Yasui, K.: Clabi–Yau caps, uniruled caps and symplectic fillings. Proc. Lond. Math. Soc. (3) **114**, 159–187 (2017)

[6] Sivek, S., van Horn-Morris, J.: Fillings of unit cotangent bundles. Math. Ann. **368**, 1063–1080 (2017)

[7] Stipsicz, A.I.: Gauge theory and Stein fillings of certain 3-manifolds. Turk. J. Math. **26**, 115–130 (2002)

[8] Wendl, C.: Strongly fillable contact manifolds and J-holomorphic foliations. Duke Math. J. **151**, 337–384 (2010)

[9] Hofer, H.: Pseudoholomorphic curves in symplectizations with applications to the Weinstein conjecture in dimension three. Invent. Math. **114**, 515–563 (1993)

[10] Eliashberg, Ya.., Hofer, H.: A Hamiltonian characterization of the three-ball. Differ. Integral Equ. **7**, 1303–1324 (1994)

[11] Kegel, M., Schneider, J., Zehmisch, K.: Symplectic dynamics and the 3-sphere. Isr. J. Math. **235**, 245–254 (2020)

[12] McDuff, D.: Symplectic manifolds with contact type boundaries. Invent. Math. **103**, 651–671 (1991)

[13] Barth, K., Geiges, H., Zehmisch, K.: The diffeomorphism type of symplectic fillings. J. Symplectic Geom. **17**, 929–971 (2019)

[14] Bowden, J., Gironella, F., Moreno, A.: Bourgeois contact structures: tightness, fillability and applications. arXiv:1908.05749

[15] Geiges, H., Kwon, M., Zehmisch, K., Diffeomorphism type of symplectic fillings of unit cotangent bundles. J. Topol. Anal. https://doi.org/10.1142/S1793525321500424

[16] Kwon, M., Zehmisch, K.: Fillings and fittings of unit cotangent bundles of odd-dimensional spheres. Q. J. Math. **70**, 1253–1264 (2019)

[17] Bramham, B., Hofer, H.: First steps towards a symplectic dynamics. In: Surveys in Differential Geometry, vol. XVII, Surveys in Differential Geometry, 17, pp. 127–177. Int. Press, Boston, MA (2012)

[18] Geiges, H., Röttgen, N., Zehmisch, K.: Trapped Reeb orbits do not imply periodic ones. Invent. Math. **198**, 211–217 (2014)

[19] Geiges, H., Zehmisch, K.: Reeb dynamics detects odd balls. Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **15**, 663–681 (2016)

[20] Barth, K., Schneider, J., Zehmisch, K.: Symplectic dynamics of contact isotropic torus complements. Münster J. Math. **12**, 31–48 (2019)

[21] Bae, Y., Wiegand, K., Zehmisch, K.: Periodic orbits in virtually contact structures. J. Topol. Anal. **12**, 371–418 (2020)

[22] Geiges, H.: An Introduction to Contact Topology, Cambridge Studies in Advanced Mathematics, vol. 109, p. xvi+440. Cambridge University Press, Cambridge (2008)

[23] Geiges, H., Zehmisch, K.: Symplectic cobordisms and the strong Weinstein conjecture. Math. Proc. Cambr. Philos. Soc. **153**, 261–279 (2012)

[24] Membrez, C., Opshtein, E.: C^0-Rigidity of Lagrangian submanifolds and punctured holomorphic discs in the cotangent bundle. arXiv:1712.06404

[25] Niederkrüger, K.: The plastikstufe—a generalization of the overtwisted disk to higher dimensions. Algebraic Geom. Topol. **6**, 2473–2508 (2006)

[26] Geiges, H., Zehmisch, K.: How to recognize a 4-ball when you see one. Münster J. Math. **6**, 525–554 (2013) (erratum: pp. 555–556)

[27] Frauenfelder, U.: Gromov convergence of pseudoholomorphic disks. J. Fixed Point Theory Appl. **3**, 215–271 (2008)

[28] Frauenfelder, U., Zehmisch, K.: Gromov compactness for holomorphic discs with totally real boundary conditions. J. Fixed Point Theory Appl. **17**, 521–540 (2015)

[29] Hofer, H.: Holomorphic curves and dynamics in dimension three. In: Symplectic Geometry and Topology (Park City, UT, 1997), IAS/Park City Mathematical Series, 7, pp. 35–101. American Mathematical Society, Providence, RI (1999)

[30] McDuff, D., Salamon, D.: Introduction to Symplectic Topology, Oxford Mathematical Monographs, Second, pp. 1–486. The Clarendon Press, Oxford University Press, New York (1998)

[31] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Lagrangian intersection Floer theory: anomaly and obstruction. In: Part II, AMS/IP Studies in Advanced Mathematics, vol. 46.2. American Mathematical Society, Providence, RI (2009)

[32] Barth, K.: Homotopy type of fillings of unit stabilized cotangent bundles. Ph.D. thesis, Westfälische Wilhelms-Universität Münster, Münster, p. viii+53 (2016)

[33] Schneider, J.: Geometric bounds on holomorphic discs in contact handles. Master thesis, Westfälische Wilhelms-Universität Münster, Münster, p. iv+57 (2017)

Myeonggi Kwon
Department of Mathematics Education
Sunchon National University
Suncheon 57922
Korea
e-mail: mkwon@scnu.ac.kr

Kevin Wiegand
Mathematisches Institut
Ruprecht-Karls-Universität Heidelberg
Im Neuenheimer Feld 205
69120 Heidelberg
Germany
e-mail: kwiegand@mathi.uni-heidelberg.de

Kai Zehmisch
Fakultät für Mathematik
Ruhr-Universität Bochum
Universitäts- straße 150
44801 Bochum
Germany
e-mail: kai.zehmisch@rub.de

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:20
https://doi.org/10.1007/s11784-022-00936-x
Published online April 1, 2022

Journal of Fixed Point Theory and Applications

Conservative surface homeomorphisms with finitely many periodic points

Patrice Le Calvez

Abstract. The goal of the article is to characterize the conservative homeomorphisms of a closed orientable surface S of genus ≥ 2, that have finitely many periodic points. By conservative, we mean a map with no wandering point. As a particular case, when S is furnished with a symplectic form, we characterize the symplectic diffeomorphisms of S with finitely many periodic points.

Mathematics Subject Classification. 37C25, 37E30, 37E45.

Contents

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday".

1. Introduction

Let S be a closed surface furnished with an area form ω and its associated Borel measure λ_ω. What are the simplest examples of diffeomorphisms that preserve ω (or homeomorphisms that preserve λ_ω) and that have finitely many periodic points? If S is the 2-sphere, the irrational rotations provide a natural family of examples. For every $\alpha \in \mathbb{T} = \mathbb{R}/\mathbb{Z}$, denote $R_\alpha : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$, the extension of the rotation $z \mapsto e^{2i\pi\tilde{\alpha}}z$ to the Riemann sphere, where $\tilde{\alpha} + \mathbb{Z} = \alpha$. An irrational rotation is a map conjugate to R_α, where $\alpha \notin \mathbb{Q}/\mathbb{Z}$. In the case where S is the 2-torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ and $\omega = dx \wedge dy$, there are two natural families of ω-preserving diffeomorphisms with no periodic points: the irrational rotations

$$R_{\beta,\gamma} : (x, y) \mapsto (x + \beta, y + \gamma), \quad (\beta, \gamma) \notin \mathbb{Q}^2/\mathbb{Z}^2,$$

and the skew products over an irrational rotation of the circle

$$S_{\delta,k} : (x, y) \mapsto (x + ky, y + \delta), \quad \delta \notin \mathbb{Q}/\mathbb{Z}, \quad k \in \mathbb{Z}\backslash\{0\}.$$

In the case where the genus of S is larger than 1, examples can be constructed, conjugate to the time one map of the flow of a minimal direction for a translation surface, giving birth to a much wider family of "classical examples". Recall that such flows admit a one dimensional section, with a return map that is an interval exchange transformation.

One can modify some of the previous examples to enlarge our class of conservative maps with finitely many periodic points. The rotation $R_{\beta,\gamma}$ is the time one map of the flow induced by the constant vector field $X_{\tilde{\beta},\tilde{\gamma}}$: $(x, y) \mapsto (\tilde{\beta}, \tilde{\gamma})$, where $\beta = \tilde{\beta} + \mathbb{Z}$ and $\gamma = \tilde{\gamma} + \mathbb{Z}$. Suppose that $\tilde{\beta}$ and $\tilde{\gamma}$ are rationally independent and let $\psi : \mathbb{T}^2 \to [0, +\infty)$ be a continuous function with finitely many zeros. The time one map of the flow induced by the vector field $\psi X_{\tilde{\beta},\tilde{\gamma}}$ has no periodic point except the zeros of ψ. Moreover, it preserves the measure $\psi^{-1}\lambda_\omega$. This measure being finite if the function ψ^{-1} is integrable, one can construct in that way a homeomorphism of the 2-torus, preserving a finite measure equivalent to the Lebesgue measure λ_ω, having an arbitrarily large number of fixed points and no other periodic point. There is a well known alternative way to construct a smooth example: in a neighborhood W of a given point, choose a symplectic system of coordinates (u, v), vanishing at this point, such that the vector field can be written $\partial/\partial u$ and is associated to the Hamiltonian function v. One can replace this function with a new Hamiltonian function having a unique singular point, being equal to $v(u^2 + v^2)$ in a neighborhood of $(0, 0)$, giving birth to a foliation (with a singular leaf) such that every new leaf intersects the boundary of W in a pair of points belonging to an old leaf. A similar construction can be done on a surface of genus ≥ 2 for a ω-preserving vector field to obtain a smooth symplectic map with an arbitrarily large number of fixed points and no other periodic point. Note that the map f obtained in that way is isotopic to the identity relative to its fixed point set. Consider now a finite cyclic covering $\hat{S}$ of S. There is a unique lift $\hat{f}$ of f to $\hat{S}$ that is isotopic to the identity. For every covering automorphism T, the map $T \circ \hat{f}$ preserves the lifted form $\hat{\omega}$ and has finitely many periodic points, all of them with the same period.

Beyond the previous examples, one can find symplectic maps having finitely many periodic points, with much richer dynamics. The approximation method by conjugation, introduced by Anosov-Katok, permits to construct smooth symplectic diffeomorphisms on the 2-sphere with exactly two fixed points z_0, z_1 and no other periodic point, such that λ_ω is ergodic (see [4]). In these examples, there exists an irrational number $\tilde{\alpha}$ (in fact a Liouville number) such that every point of $S\backslash\{z_0, z_1\}$ "turns" in the annulus $S\backslash\{z_0, z_1\}$ with an angular speed equal to $2\pi\tilde{\alpha}$. In fact, for every Liouville number, one can make the construction in such a way that λ_ω is weakly mixing (see [10]). Examples of symplectic maps having finitely many periodic points, with a rich dynamics can also be constructed in the 2-torus (see [28] for instance).

The main issue of this article is to understand to what extent the examples given above describe all symplectic diffeomorphisms of surfaces that have finitely many periodic points. We will see that they permit to get a classification: every symplectic diffeomorphism with finitely many periodic points is naturally associated to one of the examples above. Moreover, this classification is valid for a wider set: the set of homeomorphisms with no wandering point (that we will call *non wandering homeomorphisms*). Furthermore we will see that non wandering homeomorphisms with infinite many periodic points can be divided into maps of finite order and maps with periodic points of arbitrarily large period. Before stating the precise results, recall that a non empty open set $U \subset S$ is wandering if $U \cap f^{-n}(U) = \emptyset$ for every $n \geq 1$ and that points of U are wandering points. So, f is non wandering if, for every non empty open set $U \subset S$, there exists $n \geq 1$ such that $U \cap f^{-n}(U) \neq \emptyset$. By Poincaré's Recurrence Theorem, every symplectic diffeomorphism, or more generally every λ_ω-preserving homeomorphism, is non wandering.

1.1. The case of the sphere

Denote $\pi : (r, \theta) \mapsto re^{2i\pi\theta}$ the covering projection defined on the universal covering space $(0, +\infty) \times \mathbb{R}$ of $\mathbb{C}\backslash\{0\}$. An *irrational pseudo-rotation of* $\mathbb{C}$ *of rotation number* $\alpha \in \mathbb{T}\backslash\mathbb{Q}/\mathbb{Z}$ is a non wandering homeomorphism f of $\mathbb{C}$ that fixes 0 and that satisfies the following property:

if $\widetilde{f}$ is a lift of $f|_{\mathbb{C}\backslash\{0\}}$ to $(0, +\infty) \times \mathbb{R}$, there exists $\tilde{\alpha} \in \mathbb{R}$ satisfying $\tilde{\alpha} + \mathbb{Z} = \alpha$ such that for every compact set $\Xi \subset \mathbb{C}\backslash\{0\}$, and every $\varepsilon > 0$, there exists $N \geq 1$ such that

$$(P_{\tilde{\alpha}}): \quad n \geq N \quad \text{and} \quad \widetilde{z} \in \pi^{-1}(\Xi) \cap \widetilde{f}^{-n}(\pi^{-1}(\Xi)) \Rightarrow \left| \frac{p_1(\widetilde{f}^n(\widetilde{z})) - p_1(\widetilde{z})}{n} - \tilde{\alpha} \right| \leq \varepsilon,$$

where $p_1 : (\theta, r) \mapsto \theta$ is the projection on the first factor.[1] By extension, every homeomorphism of a 2-sphere that is conjugate to the extension of f to $\hat{\mathbb{C}}$ would be called an irrational pseudo-rotation of rotation number α.

[1]Requiring the whole set $\Xi = \mathbb{C}\backslash\{0\}$ to satisfy $(P_{\tilde{\alpha}})$ would be too strong, and in opposition to this weaker definition, would define a property that is not invariant by conjugacy in the group of orientation preserving homeomorphisms of $\mathbb{C}$ that fix 0.

The following result gives a very precise description of non wandering homeomorphisms of the 2-sphere with finitely many periodic points.

Theorem 1.1. *Let f be an orientation preserving and non wandering homeomorphism of the 2-sphere. Then exactly one of the following assertions holds:*

(1) *The map f has periodic points of period arbitrarily large.*
(2) *There exists $\alpha \in \mathbb{Q}/\mathbb{Z}$ such that f is conjugate to R_α.*
(3) *There exists $\alpha \in \mathbb{T}\backslash\mathbb{Q}/\mathbb{Z}$ such that f is an irrational pseudo-rotation of rotation number α.*

Theorem 1.1 is well known. Let us briefly explain why it is true. It is known (see Franks [14]) that an area preserving homeomorphism of the 2-sphere that has at least three fixed points has infinitely many periodic points (with periods arbitrarily large if the map has infinite order). Moreover, this result is also true for non wandering homeomorphisms (see [19]). So, if f is an orientation preserving and non wandering homeomorphism of the 2-sphere and if neither (1) nor (2) holds, then f has no periodic point but two fixed points. To get (3) it remains to study the non wandering homeomorphisms of the annulus $\mathbb{T} \times \mathbb{R}$ that are isotopic to the identity (this means that the orientation is preserved and the ends are fixed) and that have no periodic point. Assertion (3) is related to the Poincaré–Birkhoff Theorem and its many generalizations, its meaning is that there is a unique rotation number (for whatever reasonable definition) and moreover that it cannot be rational (see [11,18]).

1.2. The case of the torus

A classification of the area preserving homeomorphisms of $\mathbb{T}^2$ with finitely many periodic points has been done by Addas-Zanata and Tal [2]. This classification is still valid for non wandering homeomorphisms. We will state the result here but will give the definitions appearing in the statement in the last section of this article (rotation set, vertical rotation set, automorphism of $\mathbb{T}^2$). We will give the proof in the same section. The proof is nothing but the original proof of [2] except at one point where a later result of Addas-Zanata, Garcia and Tal [3] is needed to replace the area preserving condition with the non wandering condition.

Theorem 1.2. *Let f be an orientation preserving and non wandering homeomorphism of $\mathbb{T}^2$. Then exactly one of the following assertions holds:*

(1) *The map f has periodic points of period arbitrarily large.*
(2) *There exist $g \in \mathrm{Aut}(\mathbb{T}^2)$, $k \in \mathbb{Z}\backslash\{0\}$ and $\delta \notin \mathbb{Q}/\mathbb{Z}$ such that $g \circ f \circ g^{-1}$ is isotopic to $S_{0,k}$, with a vertical rotation set reduced to δ. In this case f has no periodic point.*
(3) *The map f is isotopic to the identity and its rotation set (for every lift) is a point or a segment that does not meet $\mathbb{Q}^2/\mathbb{Z}^2$. In this case f has no periodic point.*
(4) *There exists an integer $q \geq 1$ such that:*
    - *the periodic points of f^q are fixed;*
    - *the fixed point set of f^q is non empty and f^q is isotopic to the identity relative to it;*

- *the rotation set of the lift of f^q that has fixed points is reduced to* 0 *or is a segment with irrational slope that has zero as an end point.*[2]

1.3. The case of high genus

Let us state the main result of the article that gives a characterization of non wandering homeomorphisms of a surface of genus ≥ 2 that have finitely many periodic points:

Theorem 1.3. *Let S be an orientable closed surface of genus $g \geq 2$ and f an orientation preserving and non wandering homeomorphism of S. Then the following alternative holds:*

(1) *The map f has periodic points of period arbitrarily large.*
(2) *There exists an integer $q \geq 1$ such that:*
  - *the periodic points of f^q are fixed;*
  - *the fixed point set of f^q is non empty and f^q is isotopic to the identity relative to it.*[3]

Remark. Suppose that S is furnished with a symplectic form ω and that f preserves λ_ω. If f satisfies (2) and if f^q is not the identity, then the rotation vector of λ_ω (for f^q) is a non zero element of $H_1(S, \mathbb{R})$ (in other words f^q is not Hamiltonian). This is a particular case of the Conley Conjecture (see [15] or [19]). The examples we must have in mind are the ones given at the beginning of the introduction where a section exists.

Let us explain the structure of the proof of Theorem 1.3.

Definition. Let S be an orientable closed surface of genus ≥ 2. A *Dehn twist map* of S is an orientation preserving homeomorphism h of S that satisfies the following properties:

- there exists a non empty finite family $(A_i)_{i\in I}$ of pairwise disjoint invariant essential closed annuli (meaning sets homeomorphic to $\mathbb{T} \times [0, 1]$ with boundary loops non homotopic to zero in S);
- no connected component of $S \setminus \cup_{i\in I} A_i$ is an annulus (meaning a set homeomorphic to $\mathbb{T} \times (0, 1)$);
- h fixes every point of $S \setminus \cup_{i\in I} A_i$;
- for every $i \in I$, there exists $r_i \in \mathbb{Z}\setminus\{0\}$ such that $h_{|A_i}$ is conjugate to τ^{r_i}, where τ is the homeomorphism of $\mathbb{T} \times [0, 1]$ that is lifted to the universal covering space $\mathbb{R} \times [0, 1]$ by $\tilde{\tau} : (x, y) \mapsto (x + y, y)$.

The annuli A_i will be called the *twisted annuli* and r_i the *twist coefficients.*

The Nielsen–Thurston Decomposition Theorem (see [6], [9] or [27]) tells us the following: if f is an orientation preserving homeomorphism of S, then there exists a finite family $(A_i)_{i\in I}$ of pairwise disjoint essential closed annuli and a homeomorphism f_* isotopic to f such that:

[2]This last case contains the case where f has finite order.

[3]Here again the second case contains the case where f has finite order.

- no connected component of $S\setminus \cup_{i\in I} A_i$ is an annulus;
- the family $(A_i)_{i\in I}$ is invariant by f_*;
- for every connected component W of $S\setminus \cup_{i\in I} A_i$, there exists q such that ${f_*}^q(W) = W$ and ${f_*}^q{}_{|W}$ is isotopic to a map of finite order or to a pseudo-Anosov map;
- for every $i \in I$, there exists q_i such that ${f_*}^{q_i}(A_i) = A_i$ and $k_i \in \mathbb{Z}$ such that ${f_*}^{q_i}{}_{|A_i}$ is conjugate to τ^{k_i}.

In particular, the following classification holds for an orientation preserving homeomorphism f of S:

- there exists at least one component of pseudo-Anosov type in the Nielsen–Thurston classification;
- there exists $q \geq 1$ such that f^q is isotopic to a Dehn twist map;
- there exists $q \geq 1$ such that f^q is isotopic to the identity.

It is well known and folklore that if there exists a component of pseudo-Anosov type, then f has infinitely periodic points of arbitrarily large period (see [9] or [16]). As we will see in the next section, it is easy to prove that a power of a non wandering map is still non wandering. So, Theorem 1.3 can be deduced from the two following results:

Proposition 1.4. *Let S be an orientable closed surface of genus $g \geq 2$ and f a non wandering homeomorphism of S. If f is isotopic to a Dehn twist map, then it has periodic points of period arbitrarily large.*

Proposition 1.5. *Let S be an orientable closed surface of genus $g \geq 2$ and f a non wandering homeomorphism of S. If f is isotopic to the identity, then f has fixed points and*

- *either f has periodic points of period arbitrarily large;*
- *or every periodic point of f is fixed and f is isotopic to the identity relative to its fixed point set.*

1.4. Plan of the article

Proposition 1.4 is a kind of Poincaré–Birkhoff Theorem in surfaces of high genus. Its proof will be given in Sect. 3. Proposition 1.5 tells us that, roughly speaking, a non wandering homeomorphism with finitely many periodic points is "modelized" by the time one map of the flow of a minimal direction for a translation surface, after adding stopping points and lifting to a cyclic finite covering. Its proof will be given in Sect. 4. The proof of Theorem 1.2 will be given in Sect. 5. A certain number of definitions and preliminary results will be given in Sect. 2, most of the results being well known. However we will state two "new technical results" in this section, a fixed point theorem and a forcing result, which are inspired by common works with Fabio Tal ([21] and [22]).

Sections 3 and 4 are the more technical parts of the article. Surprisingly, the proofs of Proposition 1.4 and 1.5 are very similar. Let us give more details about the proof of Proposition 1.4. In a recent work with Martín Sambarino [20], we have proved that a generic symplectic diffeomorphism of a surface of genus ≥ 2, for the C^k-topology, $k \geq 1$, has transverse homoclinic intersections. A crucial argument is to prove that a generic symplectic diffeomorphism of a surface of genus ≥ 2 has more than $2g-2$ periodic points. The case of diffeomorphisms isotopic to a Dehn twist is studied in a long section of the article. We note that such a diffeomorphism f has a lift $\hat{f}$ to a certain annular covering space $\hat{S}$ that satisfies a "twist condition" and so, if f has finitely many periodic points, then $\hat{f}$ cannot satisfy the "intersection property": there exists an essential simple loop $\hat{\lambda} \subset \hat{S}$ that is disjoint from its image by $\hat{f}$. Lifting $\hat{f}$ to a diffeomorphism $\tilde{f}$ of the universal covering space $\tilde{S}$, looking at the action of the dynamics of $\tilde{f}$ on the set of lifts of $\hat{\lambda}$ and using the properties of the stable and unstable manifolds of the fixed saddle points (there exists at least $2g-2$ such points) we succeed to prove that homoclinic intersections exist. What is done in the present article is to show that the existence of infinitely many periodic points can be obtained without using the saddle points. Looking at the dynamics of $\tilde{f}$ on the set of lifts of $\hat{\lambda}$ is sufficient if we use the fixed point theorem and the forcing result stated in Sect. 2. In particular the proof is valid for area preserving homeomorphisms, or more generally for non wandering homeomorphisms.

To conclude this introduction, let us state some other known results about the dynamics of a non wandering homeomorphism f with finitely many periodic points on a surface S of genus ≥ 2. Theorem 1.3 tells us that it is sufficient to look at the case where f is isotopic to the identity relative to its fixed point set and has no other periodic point. A result of Lellouch [23] says that if μ_1 and μ_2 are two invariant Borel probability measures, then the rotation vectors of μ_1 and μ_2 do not intersect (for the canonical intersection form $\wedge$ on $H_1(S,\mathbb{R})$). Another result, that can be found in [20], is the existence, in the case where f has finitely many fixed points, of a section in the following sense: there exists a simple oriented loop $\lambda \subset S$ non homologous to zero, such that if $\check{S}$ is the infinite cyclic covering space of S associated to λ and $\check{f}$ the natural lift of f to $\check{S}$, then for every loop $\check{\lambda} \subset \hat{S}$ that lifts λ, the points that are on $\check{\lambda}$ and not fixed by $\check{f}$ are sent by $\check{f}$ on the left of $\check{\lambda}$ and by $\check{f}^{-1}$ on its right. It would be a natural challenge to look for further dynamical properties of these maps (what looks like the rotation set? does there always exist a section if f is not the identity?)

I would like to thank the referee for the useful comments.

2. Definitions, notations and preliminaries

2.1. Loops and paths

Let S be an orientable connected surface (not necessarily closed, not necessarily boundaryless). A *loop* of S is a continuous map $\gamma : \mathbb{T} \to S$. It will be called *essential* if it is not homotopic to a constant loop. A *path* of S is a continuous map $\gamma : I \to S$ where $I \subset \mathbb{R}$ is an interval. A loop or a path will be called *simple* if it is injective. A *segment* is a simple path $\sigma : [a,b] \to X$, where $a < b$. The points $\sigma(a)$ and $\sigma(b)$ are the *ends* of σ and the set $\sigma((a,b))$ its *interior*. We will say that σ *joins* $\sigma(a)$ to $\sigma(b)$. More generally if A and B are disjoint, we will say that σ joins A to B, if $\sigma(a) \in A$ and $\sigma(b) \in B$.

A *line* is a proper simple path $\lambda : \mathbb{R} \to S$, a half line a proper simple path $\lambda : I \to S$, where $I = [a, +\infty)$ or $I = (-\infty, a]$. In that case, $\gamma(a)$ is its end. As it is usually done we will use the same name and the same notation to refer to the image of a loop or a path γ.

Note that a simple loop or a simple path is naturally oriented. If γ is a simple loop that separates S (meaning that its complement has two connected components) the one that is located on the right of γ will be denoted $R(\gamma)$ and the other one $L(\gamma)$. We will use the same notation for a line that separates S, in particular for a line of $\mathbb{R}^2$.

Let f be an orientation preserving homeomorphism of $\mathbb{R}^2$. A *Brouwer line* of f is a line $\lambda \subset \mathbb{R}^2$ such that $f(\lambda) \subset L(\lambda)$ and $f^{-1}(\lambda) \subset R(\lambda)$. Equivalently it means that $f(\overline{L(\lambda)}) \subset L(\lambda)$ or that $f^{-1}(\overline{R(\lambda)}) \subset R(\lambda)$.

2.2. Homeomorphisms of hyperbolic surfaces

Let S be a connected closed orientable surface of genus $g \geq 2$. Furnishing S with a Riemannian metric of constant negative curvature, one can suppose that the universal covering space of S is the disk $\mathbb{D} = \{z \in \mathbb{C} \,|\, |z| < 1\}$ and that the group of covering transformations, denoted G, is composed of Mőbius automorphisms of $\mathbb{D}$. Every element $T \in G$ is hyperbolic: it can be extended to a homeomorphism of $\overline{\mathbb{D}} = \{z \in \mathbb{C} \,|\, |z| \leq 1\}$ having two fixed points, both on the boundary: a repelling fixed point $\alpha(T)$ and an attracting fixed point $\omega(T)$. For every $z \in \overline{\mathbb{D}} \setminus \{\alpha(T), \omega(T)\}$, it holds that

$$\lim_{k \to -\infty} T^k(z) = \alpha(T), \lim_{k \to +\infty} T^k(z) = \omega(T).$$

It is well known that for every α, ω in the boundary of $\mathbb{D}$, the set G' of covering transformations T such that $\alpha(T) = \alpha$ and $\omega(T) = \omega$, is an infinite cyclic subgroup of G, if not empty, because G' acts freely and properly on the geodesic joining α to ω. One also knows that if $T' \notin G'$, then $\alpha(T')$ and $\omega(T')$ do not belong to $\{\alpha, \beta\}$. Indeed, suppose for example that one of the points $\alpha(T')$ or $\omega(T')$ is equal to α. Choose $T \in G$. Using the fact that ω is an attracting fixed point of T, one deduces that one of the points $\alpha(T^n \circ T' \circ T^{-n})$ or $\omega(T^n \circ T' \circ T^{-n})$ is equal to α and the other one tends to ω, when n tends to $+\infty$. It contradicts the fact that G is discrete. We define a *T-line* to be a line of $\mathbb{D}$ invariant by T and oriented in such a way that it can be extended to a segment of $\overline{\mathbb{D}}$ that joins $\alpha(T)$ to $\omega(T)$.

It is well known that a homeomorphism $\tilde{f}$ of $\mathbb{D}$ that lifts a homeomorphism f of S can be extended to a homeomorphism of $\overline{\mathbb{D}} = \{z \in \mathbb{C} \,|\, |z| \leq 1\}$. If $[\tilde{f}]$ is the automorphism of G defined by the relation:

$$\tilde{f}(T(z)) = [\tilde{f}](T)(\tilde{f}(z)), \quad \text{for all} \quad z \in \mathbb{D},$$

then the extension of $\tilde{f}$ satisfies

$$\tilde{f}(\alpha(T)) = \alpha([\tilde{f}](T)) \quad \text{and} \quad \tilde{f}(\omega(T)) = \omega([\tilde{f}](T)) \quad \text{for all} \quad T \in G.$$

If h is a homeomorphism of S that is isotopic to f, then every isotopy from f to h can be lifted to an isotopy from $\tilde{f}$ to a certain lift $\tilde{h}$ of h. The lift $\tilde{h}$ does not depend on the initial isotopy and has the same extension as $\tilde{f}$ on the boundary circle, because the automorphisms $[\tilde{f}]$ and $[\tilde{h}]$ coincide. For conveniency, we will write $\tilde{S}$ for the universal covering space of S and $\partial\tilde{S}$ for its boundary defined via the identification of $\tilde{S}$ with $\mathbb{D}$. We will write $\overline{\tilde{S}}=\tilde{S}\cup\partial(\tilde{S})$.

2.3. Non wandering homeomorphisms

Let us recall some very classical easy results that we will use in the article. Recall that if f is a homeomorphism of a surface S, a point $z\in S$ is recurrent if there exists a subsequence of $(f^n(z))_{n\geq 0}$ that converges to z.

Lemma 2.1. *Let f be a non wandering homeomorphism of a surface S. For every non empty open set U and every $q\geq 1$, there exists an increasing sequence $(n_i)_{0\leq i\leq q}$ in $\mathbb{N}$, satisfying $n_0=0$, such that $\bigcap_{0\leq i\leq q} f^{-n_i}(U)\neq\emptyset$.*

Proof. Let us prove the lemma by induction. By definition of a non wandering homeomorphism, the lemma is true for $q=1$. Suppose that the lemma is true for every $q'<q$, where $q\geq 2$. Let $U\subset S$ be a non empty open set. There exists an increasing sequence $(n_i)_{0\leq i<q}$ in $\mathbb{N}$, satisfying $n_0=0$, such that $\bigcap_{0\leq i<q} f^{-n_i}(U)\neq\emptyset$. As $V=\bigcap_{0\leq i<q} f^{-n_i}(U)$ is open and non empty, there exists $n>0$ such that $V\cap f^{-n}(V)\neq\emptyset$. In particular, it holds that $\bigcap_{0\leq i<q} f^{-n_i}(U)\neq\emptyset$, where $n_q=n_{q-1}+n$. So, the lemma is true for q. ☐

Proposition 2.2. *Let f be a non wandering homeomorphism of a surface S. Then:*

1. *every power f^k, $k\in\mathbb{Z}$, is non wandering;*
2. *if $\hat{S}$ is a finite covering of S, every lift of f to $\hat{S}$ is non wandering;*
3. *the set of recurrent points is a dense G_δ set.*

Proof. By definition, f is non wandering if and only if f^{-1} is non wandering. Moreover the identity is non wandering. So, to prove (1) it is sufficient to prove it for $k\geq 2$. Let $U\subset S$ be a non empty open set. By Lemma 2.1, there exists an increasing sequence $(n_i)_{0\leq i\leq k}$ in $\mathbb{N}$, satisfying $n_0=0$, such that $\bigcap_{0\leq i\leq k} f^{-n_i}(U)\neq\emptyset$. There exists $i_0<i_1$ such that $n_{i_1}-n_{i_0}\in k\mathbb{Z}$. Write $n_{i_1}-n_{i_0}=nk$, where $n>0$. It holds that $U\cap f^{-nk}(U)=f^{n_{i_0}}(f^{-n_{i_0}}(U)\cap f^{-n_{i_1}}(U))\neq\emptyset$. So U is a non wandering open set of f^k.

Let us prove (2). Suppose that $\hat{S}$ is a r-cover of S and denote $\hat{\pi}:\hat{S}\to S$ the covering projection. To prove that $\hat{f}$ is non wandering, it is sufficient to prove that if $U\subset S$ is an open disk such that $\hat{\pi}^{-1}(U)=\bigsqcup_{1\leq j\leq r}\hat{U}_j$, where each $\hat{U}_j$ is mapped homeomorphically onto U by $\hat{\pi}$, then every $\hat{U}_j$ is non wandering. By Lemma 2.1, there exists an increasing sequence $(n_i)_{0\leq i\leq r!}$ in $\mathbb{N}$, satisfying $n_0=0$, such that $\bigcap_{0\leq i\leq r!} f^{-n_i}(U)\neq\emptyset$. Let us choose $z\in\bigcap_{0\leq i\leq r!} f^{-n_i}(U)$ and denote $\hat{z}_j$ the preimage of z belonging to $\hat{U}_j$. For every $i\in\{0,\dots,r!\}$ denote $\sigma_i\in\mathcal{S}_r$ the permutation such that $\hat{f}^{n_i}(\hat{z}_j)\in\hat{U}_{\sigma_i(j)}$. There exists $i_0<i_1$ such that $\sigma_{i_0}=\sigma_{i_1}$. Setting $n_{i_1}-n_{i_0}=n>0$, one deduces that $\hat{U}_j\cap\hat{f}^{-n}(\hat{U}_j)\neq\emptyset$, for every $j\in\{1,\dots,r\}$.

To prove (3) furnish S with a distance d. If $m\geq 1$ and $q\geq 1$, define

$$O_{m,q}=\{x\in S,\quad \exists n\geq q,\quad d(f^n(x),x)<\frac{1}{m}\}.$$

Applying Lemma 2.1 to a ball $B(z,\varepsilon)$, $\varepsilon < 1/2m$, and to q, we obtain that $O_{m,q} \cap B(z,\varepsilon) \neq \emptyset$. So $O_{m,q}$ is a dense open set. Noting that $\bigcap_{m\geq 1, q\geq 1} O_{m,q}$ is the set of recurrent points and that S is a Baire space, we can conclude. □

We will often use the following result:

Proposition 2.3. *Let f be a non wandering homeomorphism of a surface S and $\hat{f}$ a lift of f to a covering space $\hat{S}$. If $\hat{U}$ is a non empty wandering open set of $\hat{f}$, then $\bigcup_{n\geq 0} \hat{f}^n(\hat{U})$ and $\bigcup_{n\geq 0} \hat{f}^{-n}(\hat{U})$ are not relatively compact.*

Proof. Of course it is sufficient to prove that $\bigcup_{n\geq 0} \hat{f}^n(\hat{U})$ is not relatively compact. We will argue by contradiction and suppose that $\bigcup_{n\geq 0} \hat{f}^n(\hat{U})$ is relatively compact. The frontier $\mathrm{fr}\left(\bigcup_{n\geq 0} \hat{f}^n(\hat{U})\right)$ of $\bigcup_{n\geq 0} \hat{f}^n(\hat{U})$ is a compact set with empty interior. The covering map $\hat{\pi} : \hat{S} \to S$ being continuous, $\hat{\pi}\left(\mathrm{fr}\left(\bigcup_{n\geq 0} \hat{f}^n(\hat{U})\right)\right)$ is compact. The map $\hat{\pi}$ being a local homeomorphism, $\hat{\pi}\left(\mathrm{fr}\left(\bigcup_{n\geq 0} \hat{f}^n(\hat{U})\right)\right)$ has empty interior. Indeed, one can cover $\mathrm{fr}\left(\bigcup_{n\geq 0} \hat{f}^n(\hat{U})\right)$ with finitely many open sets that are sent homeomorphically by $\hat{\pi}$ on their image. The map $\hat{\pi}$, being continuous and open, $\hat{\pi}^{-1}\left(\hat{\pi}\left(\mathrm{fr}\left(\bigcup_{n\geq 0} \hat{f}^n(\hat{U})\right)\right)\right)$ is a closed set of $\hat{S}$ with empty interior. Denote $\mathrm{rec}(f)$ the set of recurrent points of f, which is a dense G_δ set. The map $\hat{\pi}$ being continuous and open, $\hat{\pi}^{-1}(S\backslash\mathrm{rec}(f))$ is a F_σ set with empty interior. The contradiction comes from the fact that $\hat{U}$ is contained in the union of $\hat{\pi}^{-1}\left(\hat{\pi}\left(\mathrm{fr}\left(\bigcup_{n\geq 0} \hat{f}^n(\hat{U})\right)\right)\right)$ and $\hat{\pi}^{-1}(S\backslash\mathrm{rec}(f))$. Indeed, suppose that the image by $\hat{\pi}$ of $\hat{z} \in U$ is recurrent. The closure of $\bigcup_{n\geq 0} \hat{f}^n(\hat{U})$ contains finitely many preimages of $\hat{\pi}(\hat{z})$. At least one of them belongs to $\omega(\hat{z})$, meaning that it is a limit of a subsequence of $(\hat{f}^n(\hat{z}))_{n\geq 0}$. It cannot belong to $\bigcup_{n\geq 0} \hat{f}^n(\hat{U})$ because $\hat{U}$ is wandering and so is on the frontier.[4] □

Remark. In case f preserves a totally supported finite measure μ, the results above are obvious. Indeed, in Proposition 2.2, the maps f^k preserve μ and $\hat{f}$ preserves the lift of μ that is a totally supported finite measure. Moreover, as a consequence of Poincaré's Recurrence Theorem, it is known that almost every point is recurrent, and so the set of recurrent points is dense. In Proposition 2.3, $\hat{f}$ preserves a totally supported locally finite measure.

2.4. Poincaré–Birkhoff Theorem

We suppose in this subsection that I is a non trivial interval of $\mathbb{R}$. We define the annuli $\mathbb{A} = \mathbb{T} \times I$ and $\mathrm{int}(\mathbb{A})$, where $\mathrm{int}(\mathbb{A})$ is obtained from $\mathbb{A}$ by taking out the possible boundary circles. Writing $\tilde{\mathbb{A}} = \mathbb{R}\times I$ for the universal covering space of $\mathbb{A}$, we define the covering projection $\pi : \tilde{\mathbb{A}} \to \mathbb{A}, (x,y) \mapsto (x+\mathbb{Z}, y)$ and the generating covering automorphism $T : \tilde{\mathbb{A}} \to \tilde{\mathbb{A}}, (x,y) \mapsto (x+1, y)$.

[4] In the case where $\hat{S}$ is a normal covering space and $\hat{f}$ commutes with the covering automorphisms, a much simpler proof can be given. Unfortunately, this will not be always the case when we will apply Proposition 2.3.

Let f be a homeomorphism of $\mathbb{A}$ isotopic to the identity (meaning orientation preserving and fixing the possible ends or boundary circles) and $\tilde{f}$ a lift of f to $\tilde{\mathbb{A}}$. Let us recall the definition of the rotation number of a compactly supported Borel probability measure μ invariant by f. Denote $p_1 : \tilde{\mathbb{A}} \to \mathbb{R}$ the projection on the first factor. The maps $\tilde{f}$ and T commute, so $p_1 \circ \tilde{f} - p_1$ lifts a continuous function $\psi_{\tilde{f}} : \mathbb{A} \to \mathbb{R}$. The *rotation number* $\mathrm{rot}_{\tilde{f}}(\mu) = \int_{\mathbb{A}} \psi_{\tilde{f}}\, d\mu \in \mathbb{R}$ measures the mean horizontal displacement of $\tilde{f}$.

If z is a periodic point of f of period q and if $\tilde{z}$ is a lift of z in $\tilde{\mathbb{A}}$, then there exists an integer p, independent of the choice of $\tilde{z}$, such that $\tilde{f}^q(\tilde{z}) = T^p(\tilde{z})$. We will say that p/q is the rotation number of z for the lift $\tilde{f}$, it coincides with the rotation vector of the equidistributed measure supported on the orbit of z.

We will use many times the following extension of the classical Poincaré–Birkhoff Theorem (see [18]):

Theorem 2.4. *Let f be a homeomorphism of $\mathbb{A}$ isotopic to the identity and $\tilde{f}$ a lift of f to $\tilde{\mathbb{A}}$. We suppose that there exist two invariant ergodic compactly supported Borel probability measures μ_1 and μ_2, such that $\mathrm{rot}_{\tilde{f}}(\mu_1) < \mathrm{rot}_{\tilde{f}}(\mu_2)$. Then:*

- *either, for every rational number $p/q \in (\mathrm{rot}_{\tilde{f}}(\mu_1), \mathrm{rot}_{\tilde{f}}(\mu_2))$, written in an irreducible way, there exists a periodic point z of f of period q and rotation number p/q for $\tilde{f}$;*
- *or there exists an essential simple loop $\lambda \in \mathrm{int}(\mathbb{A})$ such that $f(\lambda) \cap \lambda = \emptyset$.*

2.5. A fixed point theorem for a planar homeomorphism

In this sub-section, we will give a criterion of existence of a fixed point for a planar homeomorphism, which is a slight generalization of a result proved in Proposition 12 of [22]. It will be an essential tool in the proofs of Propositions 1.4 and 1.5.

Let $(\lambda_i)_{1 \le i \le r}$ be a finite family of pairwise disjoint lines of $\mathbb{R}^2$. Say that the family is *cyclically ordered* if:

- one can choose, for every $i \in \{1, \dots, r\}$, a connected component E_i of $\mathbb{R}^2 \backslash \lambda_i$ in such a way that the sets $\overline{E}_i$ are pairwise disjoint;
- for every $i \in \{1, \dots, r-1\}$, one can find two disjoint connected sets, the first one containing λ_i and λ_{i+1}, the other one containing every line λ_j, $j \notin \{i, i+1\}$.

Note that the complement of $\bigcup_{1 \le i \le r} E_i$ is a connected sub-surface Σ whose boundary is equal to $\bigcup_{1 \le i \le r} \lambda_i$. Note also that one can find two disjoint connected sets, the first one containing λ_1 and λ_r, the other one containing every line λ_j, $1 < j < r$. By the extension of Schoenflies Theorem due to Homma (see [17]), one knows that $(\lambda_i)_{1 \le i \le r}$ is cyclically ordered if and only if there exists a homeomorphism of $\mathbb{R}^2$ that sends λ_i on the graph of the function

$$\psi_i : (2i-1, 2i+1) \to \mathbb{R}, \quad x \mapsto \frac{1}{1-(x-2i)^2}.$$

Proposition 2.5. *Let f be a homeomorphism of $\mathbb{R}^2$. Suppose that there exists a cyclically ordered family $(\lambda_i)_{1\leq i\leq 4}$ of pairwise disjoint lines and for every $i\in\{1,3\}$ a segment $\sigma_i\subset\lambda_i$ such that:*

- $f(\sigma_i)\cap\lambda_i=\emptyset$ *if* $i\in\{1,3\}$*;*
- $f(\lambda_j)\cap\lambda_j=\emptyset$ *if* $j\in\{2,4\}$*;*
- $f(\sigma_i)\cap\lambda_j\neq\emptyset$ *if* $i\in\{1,3\}$ *and* $j\in\{2,4\}$.

Then f has a fixed point.

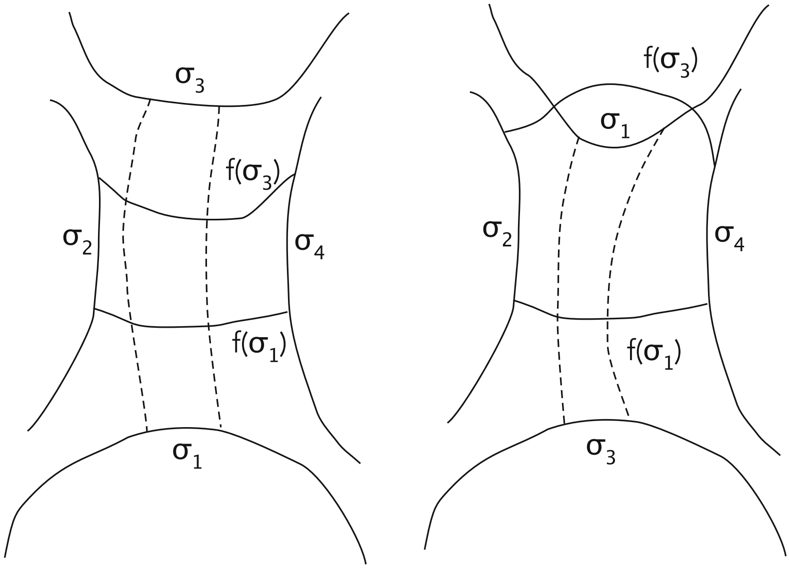
Proof. Recall that E_i is the connected component of $\mathbb{R}^2\backslash\lambda_i$ that does not contain the λ_j, $j\neq i$. Taking a sub-segment of σ_1 if necessary, one can suppose that one end of $f(\sigma_1)$ belongs to λ_2, the other end to λ_4 and the other points of $f(\sigma_1)$ neither to λ_2 nor to λ_4. The segment $f(\sigma_1)$, being disjoint from λ_1, belongs to the connected component of $\mathbb{R}^2\backslash\lambda_1$ that contains λ_2 and λ_4, it does not meet $\overline{E}_1$. Moreover, the interior of $f(\sigma_1)$ is contained in the connected component of $\mathbb{R}^2\backslash(\lambda_2\cup\lambda_4)$ that contains λ_1 and λ_3, it does not meet neither $\overline{E}_2$, nor $\overline{E}_4$. One can suppose that σ_3 satisfies similar properties and of course $f(\sigma_1)$ and $f(\sigma_3)$ are disjoint. Let $\sigma_2\subset\lambda_2$ be the segment that joins the two ends of $f(\sigma_1)$ and $f(\sigma_3)$ that are on λ_2. By hypothesis, $f^{-1}(\sigma_2)$ is disjoint from λ_2 and so is included in the connected component of $\mathbb{R}^2\backslash\lambda_2$ that contains λ_1 and λ_3, it is disjoint from $\overline{E}_2$. The segment $\sigma_4\subset\lambda_4$ that joins the two ends of $f(\sigma_1)$ and $f(\sigma_3)$ that are on λ_4, satisfies similar properties. One gets a loop C by taking the union of $f(\sigma_1)$, $f(\sigma_3)$, σ_2 and σ_4. The vector field $z\mapsto f^{-1}(z)-z$ does not vanish on C, let us explain why its index on C is equal to 1 or -1, which implies that there exists at least one fixed point of f in the bounded component of $\mathbb{R}^2\backslash C$. The set of orientation preserving homeomorphisms h of $\mathbb{R}^2$ being path connected for the compact open topology, the value of the index of the vector field $z\mapsto h\circ f^{-1}\circ h^{-1}(z)-z$ on $h(C)$ does not depend on the choice of h. Applying Homma's theorem, one can find an orientation preserving homeomorphism h such that :

$$\begin{aligned} h(\lambda_2)&=\{0\}\times\mathbb{R},\\ h(\lambda_4)&=\{1\}\times\mathbb{R},\\ h(f(\sigma_1))&=[0,1]\times\{0\},\\ h(f(\sigma_3))&=[0,1]\times\{1\}, \end{aligned}$$

or

$$\begin{aligned} h(\lambda_2)&=\{0\}\times\mathbb{R},\\ h(\lambda_4)&=\{1\}\times\mathbb{R},\\ h(f(\sigma_3))&=[0,1]\times\{0\},\\ h(f(\sigma_1))&=[0,1]\times\{1\}. \end{aligned}$$

There is no loss of generality by supposing that the first situation occurs. The family $(\lambda_i)_{1\leq i\leq 4}$ being cyclically ordered, there are two possibilities: in the first case, $\overline{E}_1$ is contained in $(0,1)\times(-\infty,0)$ and $\overline{E}_3$ in $(0,1)\times(1,+\infty)$; in the second case, $\overline{E}_1$ is contained in $(0,1)\times(0+\infty)$ and $\overline{E}_3$ in $(0,1)\times(-\infty,1)$. It is very easy to compute the index of the vector field $z\mapsto h\circ f^{-1}\circ h^{-1}(z)-z$ on the square $h(C)$: in the first case, the vector field is pointing on the right on $\{0\} \times [0, 1]$, pointing downwards on $[0, 1] \times \{0\}$, pointing on the left on $\{1\} \times [0, 1]$, pointing upwards on $[0, 1] \times \{1\}$ and the index is -1; in the second case, the vector field is pointing on the right on $\{0\} \times [0, 1]$, pointing upwards on $[0, 1] \times \{0\}$, pointing on the left on $\{1\} \times [0, 1]$, pointing downwards on $[0, 1] \times \{1\}$ and the index is $+1$.



□

Remark. A special case where this criterion can be applied is when every line λ_i is free (meaning that $f(\lambda_i) \cap \lambda_i = \emptyset$) and $f(\lambda_i) \cap \lambda_j \neq \emptyset$ if $i \in \{1, 3\}$ and $j \in \{2, 4\}$. The proposition above was proved in [22] when every λ_i is a Brouwer line and $f(\lambda_i) \cap \lambda_j \neq \emptyset$ if $i \in \{1, 3\}$ and $j \in \{2, 4\}$. The proof above tells us that the arguments given in [22] are still valid with some weaker hypothesis.

2.6. Some forcing results

In this sub-section, we will state another result which will be essential in the proofs of Propositions 1.4 and 1.5. It is naturally related to the forcing lemma for Brouwer lines that is stated in Proposition 20 of [21], but will concern free lines instead of Brouwer lines and will “include” a dynamics among these lines.

Denote

$$p_1 : \tilde{\mathbb{A}} \to \mathbb{R} \quad \text{and} \quad p_2 : \tilde{\mathbb{A}} \to [0, 1]$$

the horizontal and vertical projections defined on $\tilde{\mathbb{A}} = \mathbb{R} \times [0, 1]$.

Write $\mathrm{Homeo}_*(\tilde{\mathbb{A}})$ for the set of orientation preserving homeomorphisms of $\tilde{\mathbb{A}}$ that lets invariant each boundary line. The boundary lines can be naturally ordered, transposing by p_1 the usual order of the real line.

Denote $\tilde{\mathcal{E}}_0$ the set of lines $\tilde{\lambda}_0 : \mathbb{R} \to \mathbb{R} \times (0,1)$ such that there exist $(\tilde{a}_0^-, 0) \neq (\tilde{a}_0^+, 0)$ in $\mathbb{R} \times \{0\}$ satisfying

$$\lim_{t\to -\infty} \tilde{\lambda}_0(t) = (\tilde{a}_0^-, 0), \quad \lim_{t\to +\infty} \tilde{\lambda}_0(t) = (\tilde{a}_0^+, 0).$$

We will say that $(\tilde{a}_0^-, 0)$ and $(\tilde{a}_0^+, 0)$ are the *ends* of $\tilde{\lambda}$.[5] We can define a relation $\prec$ on $\tilde{\mathcal{E}}_0$, writing $\tilde{\lambda}_0 \prec \tilde{\lambda}'_0$ if:

- $\tilde{\lambda}_0 \cap \tilde{\lambda}'_0 = \emptyset$;
- the smallest end of $\tilde{\lambda}_0$ is smaller than the smallest end of $\tilde{\lambda}'_0$ and the highest end of $\tilde{\lambda}_0$ is smaller than the highest end of $\tilde{\lambda}'_0$.

Note that if $\tilde{\lambda}_0 \prec \tilde{\lambda}'_0$, then the highest end of $\tilde{\lambda}_0$ is not higher than the smallest end of $\tilde{\lambda}'_0$ because $\tilde{\lambda}_0 \cap \tilde{\lambda}'_0 = \emptyset$ but it can be equal. The relation $\prec$ is not transitive owing to the first condition. Nevertheless, if $\tilde{\mathcal{F}}_0$ is a subset of $\tilde{\mathcal{E}}_0$ such that the lines $\tilde{\lambda}_0 \in \tilde{\mathcal{F}}_0$ are pairwise disjoint, then the restriction of $\prec$ to $\tilde{\mathcal{F}}_0$ induces an order $\preceq$ (not necessarily total) defined as follows:

$$\tilde{\lambda}_0 \preceq \tilde{\lambda}'_0 \Leftrightarrow \tilde{\lambda}_0 = \tilde{\lambda}'_0 \quad \text{or} \quad \tilde{\lambda}_0 \prec \tilde{\lambda}'_0.$$

Every $\tilde{f} \in \mathrm{Homeo}_*(\tilde{\mathbb{A}})$ naturally acts on $\tilde{\mathcal{E}}_0$ and it holds that

$$\tilde{\lambda}_0 \prec \tilde{\lambda}'_0 \Longrightarrow \tilde{f}(\tilde{\lambda}_0) \prec \tilde{f}(\tilde{\lambda}'_0).$$

Similarly, one can define the set $\tilde{\mathcal{E}}_1$ of lines $\tilde{\lambda}_1 : \mathbb{R} \to \mathbb{R} \times (0,1)$ such that there exist $(\tilde{a}_1^-, 1) \neq (\tilde{a}_1^+, 1)$ in $\mathbb{R} \times \{1\}$, the ends of $\tilde{\lambda}_1$, satisfying

$$\lim_{t\to -\infty} \tilde{\lambda}_1(t) = (\tilde{a}_1^-, 1), \quad \lim_{t\to +\infty} \tilde{\lambda}_1(t) = (\tilde{a}_1^+, 1).$$

Moreover, one can define a relation $\prec$ on $\tilde{\mathcal{E}}_1$ like we did on $\tilde{\mathcal{E}}_0$.

We now fix $\tilde{f} \in \mathrm{Homeo}_*(\tilde{\mathbb{A}})$ until the end of the section. Consider $\tilde{\lambda}_0 \in \tilde{\mathcal{E}}_0$, $\tilde{\lambda}_1 \in \tilde{\mathcal{E}}_1$ and $n \geq 1$. We will write $\tilde{\lambda}_0 \xrightarrow{n} \tilde{\lambda}_1$ in the case where $\tilde{f}^n(\tilde{\lambda}_0) \cap \tilde{\lambda}_1 \neq \emptyset$. The following result is immediate:

Lemma 2.6. *Suppose that $\tilde{\lambda}_0 \in \tilde{\mathcal{E}}_0$, $\tilde{\lambda}_1 \in \tilde{\mathcal{E}}_1$ and $n \geq 1$ satisfy $\tilde{\lambda}_0 \xrightarrow{n} \tilde{\lambda}_1$. Then it holds that*

$$\tilde{\lambda}_0 \xrightarrow{n+1} \tilde{f}(\tilde{\lambda}_1), \quad \tilde{f}^{-1}(\tilde{\lambda}_0) \xrightarrow{n+1} \tilde{\lambda}_1.$$

Moreover, if $\tilde{h} \in \mathrm{Homeo}_*(\tilde{\mathbb{A}})$ commutes with $\tilde{f}$, then

$$\tilde{h}(\tilde{\lambda}_0) \xrightarrow{n} \tilde{h}(\tilde{\lambda}_1).$$

The next result is less obvious:

Lemma 2.7. *Suppose that $\tilde{\lambda}_0 \in \tilde{\mathcal{E}}_0$, $\tilde{\lambda}'_0 \in \tilde{\mathcal{E}}_0$, $\tilde{\lambda}_1 \in \tilde{\mathcal{E}}_1$, $\tilde{\lambda}'_1 \in \tilde{\mathcal{E}}_1$, $n \geq 1$, $n' \geq 1$ satisfy*

$$\tilde{\lambda}_0 \xrightarrow{n} \tilde{\lambda}_1, \quad \tilde{\lambda}'_0 \xrightarrow{n'} \tilde{\lambda}'_1$$

and

$$\tilde{f}^n(\tilde{\lambda}_0) \prec \tilde{\lambda}'_0, \quad \tilde{\lambda}'_1 \prec f^{n'}(\tilde{\lambda}_1), \quad \tilde{\lambda}'_0 \cap \tilde{\lambda}_1 = \emptyset.$$

[5] We could have defined $\tilde{\mathcal{E}}_0$ to be the set of segments of $\tilde{\mathbb{A}}$ whose ends are on $\mathbb{R} \times \{0\}$ and whose interior is in the interior of $\tilde{\mathbb{A}}$ but the object that will appear naturally for the applications are lines and not segments.

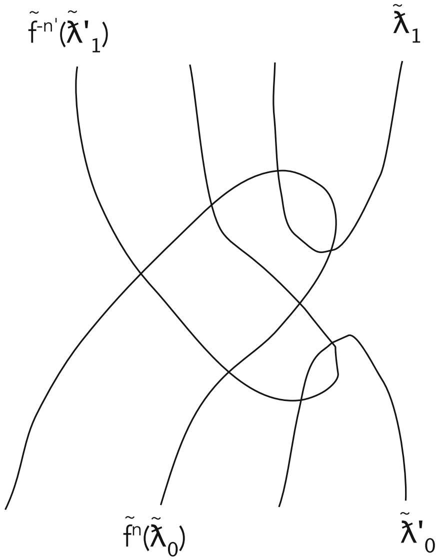
Then, one has

$$\tilde{\lambda}_0 \xrightarrow{n+n'} \tilde{\lambda}'_1.$$

Proof. The lines $\tilde{f}^n(\tilde{\lambda}_0)$ and $\tilde{\lambda}_1$ intersect, so $\tilde{f}^n(\tilde{\lambda}_0) \cup \tilde{\lambda}_1$ is connected. Similarly, $\tilde{\lambda}'_0$ and $\tilde{f}^{-n'}(\tilde{\lambda}'_1)$ intersect, so $\tilde{\lambda}'_0 \cup \tilde{f}^{-n'}(\tilde{\lambda}'_1)$ is connected. By hypothesis, one has $\tilde{f}^{-n'}(\tilde{\lambda}'_1) \prec \tilde{\lambda}_1$. The fact that $\tilde{f}^n(\tilde{\lambda}_0) \prec \tilde{\lambda}'_0$ and $\tilde{f}^{-n'}(\tilde{\lambda}'_1) \prec \tilde{\lambda}_1$ implies that $\tilde{f}^n(\tilde{\lambda}_0) \cup \tilde{\lambda}_1$ and $\tilde{\lambda}'_0 \cup \tilde{f}^{-n'}(\tilde{\lambda}'_1)$ intersect. Noting that all sets

$$\tilde{\lambda}'_0 \cap \tilde{\lambda}_1, \quad \tilde{f}^n(\tilde{\lambda}_0) \cap \tilde{\lambda}'_0, \quad \tilde{f}^{-n'}(\tilde{\lambda}'_1) \cap \tilde{\lambda}_1$$

are empty, we deduce that $\tilde{f}^n(\tilde{\lambda}_0)$ and $\tilde{f}^{-n'}(\tilde{\lambda}'_1)$ intersect, which means that $\tilde{\lambda}_0 \xrightarrow{n+n'} \tilde{\lambda}'_1$.



□

Remark. Lemma 2.7 is a slight modification of the forcing lemma given in [21]. Of course, its conclusion still holds under the assumptions

$$\tilde{\lambda}'_0 \prec \tilde{f}^n(\tilde{\lambda}_0), \quad \tilde{f}^{n'}(\tilde{\lambda}_1) \prec \tilde{\lambda}'_1, \quad \tilde{\lambda}'_0 \cap \tilde{\lambda}_1 = \emptyset.$$

Let us state now the main result, supposing that $\tilde{f}$ commutes with $T : (x, y) \mapsto (x+1, y)$ or equivalently supposing that it lifts a homeomorphism of $\mathbb{A} = \mathbb{T} \times [0, 1]$.

Proposition 2.8. *Let ρ_0 and ρ_1 be the rotation numbers induced by $\tilde{f}$ on $\mathbb{R} \times \{0\}$ and $\mathbb{R} \times \{1\}$ respectively. Suppose that $\tilde{\lambda}_0 \in \tilde{\mathcal{E}}_0$, $\tilde{\lambda}_1 \in \tilde{\mathcal{E}}_1$ satisfy the following:*

1. *for every $n > 0$ and every $p \in \mathbb{Z}$, one has $\tilde{f}^n(\tilde{\lambda}_0) \cap T^p(\tilde{\lambda}_0) = \emptyset$,*
2. *for every $n > 0$ and every $p \in \mathbb{Z}$, one has $\tilde{f}^n(\tilde{\lambda}_1) \cap T^p(\tilde{\lambda}_1) = \emptyset$,*
3. *for every $n > 0$ and every $p \in \mathbb{Z}$, one has $\tilde{f}^n(\tilde{\lambda}_1) \cap T^p(\tilde{\lambda}_0) = \emptyset$,*
4. *there exists $n_0 \geq 1$ such that $\tilde{\lambda}_0 \xrightarrow{n_0} \tilde{\lambda}_1$.*

Then, if $p \in \mathbb{Z}$ *and* $q \geq 1$ *satisfy*

$$\rho_0(n_0 + q) < p < \rho_1(n_0 + q),$$

or

$$\rho_1(n_0 + q) < p < \rho_0(n_0 + q),$$

it holds that

$$\tilde{\lambda}_0 \xrightarrow{2n_0+q} T^p \tilde{\lambda}_1.$$

Proof. We can suppose that $\rho_0(n_0 + q) < p < \rho_1(n_0 + q)$, the other case being similar. Let $(\tilde{z}_0^-, 0)$ and $(\tilde{z}_0^+, 0)$ be the ends of $\tilde{\lambda}_0$. One can suppose for instance that $\tilde{z}_0^- < \tilde{z}_0^+$. From the relation $\rho_0(n_0 + q) < p$, we obtain that

$$\tilde{f}^{n_0+q}(\tilde{z}_0^-, 0) < T^p(\tilde{z}_0^-, 0), \quad \tilde{f}^{n_0+q}(\tilde{z}_0^+, 0) < T^p(\tilde{z}_0^+, 0).$$

As a consequence of (1) we deduce that

$$\tilde{f}^{n_0+q}(\tilde{\lambda}_0) \prec T^p(\tilde{\lambda}_0).$$

Similarly, as a consequence of (2) and of the relation $p < \rho_1(n_0 + q)$, we deduce that

$$T^p(\tilde{\lambda}_1) \prec \tilde{f}^{n_0+q}(\tilde{\lambda}_1).$$

Writing $\tilde{f}^{n_0+q}(\tilde{\lambda}_1) = \tilde{f}^{n_0}(\tilde{f}^q(\tilde{\lambda}_1))$, using (3) and the relations

$$\tilde{\lambda}_0 \xrightarrow{n_0+q} \tilde{f}^q(\tilde{\lambda}_1), \quad T^p\tilde{\lambda}_0 \xrightarrow{n_0} T^p\tilde{\lambda}_1,$$

we conclude by Lemma 2.7, that

$$\tilde{\lambda}_0 \xrightarrow{2n_0+q} T^p \tilde{\lambda}_1.$$

□

Remark. Of course, what has been done in this section can be extended to any abstract annulus (meaning every topological space homeomorphic to $\mathbb{A}$) and its universal covering space, the sets $\tilde{\mathcal{E}}_0$ and $\tilde{\mathcal{E}}_1$ being defined relative to the two boundary lines of the universal covering space.

3. Dehn twist maps

The goal of this section is to prove Proposition 1.4, which means to prove that if S is an orientable closed surface of genus $g \geq 2$ and f a non wandering homeomorphism of S isotopic to a Dehn twist map, then f has periodic points of period arbitrarily large.

We will fix from now on a Dehn twist map h on S and a homeomorphism f isotopic to h (note that f is orientation preserving). We will begin by stating some results that can be found in [20]. We denote $(A_i)_{i\in I}$ the family of twisted annuli and $(r_i)_{i\in I}$ the family of twist coefficients. Fix an annulus $A = A_{i_0}$ and then choose a connected component $\tilde{A}$ of $\tilde{\pi}^{-1}(A)$, where $\tilde{\pi} : \tilde{S} \to S$ is the universal covering projection. The boundary of $\tilde{A}$ is the union of two lines $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$, each of them lifting a boundary circle of A, denoted respectively λ_1 and λ_2. We orient $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ in such a way that $\tilde{\lambda}_2 \subset L(\tilde{\lambda}_1)$ and $\lambda_1 \subset R(\tilde{\lambda}_2)$.

There exists $T_0 \in G$, uniquely defined, such that $\tilde{\lambda}_1$ and $\tilde{\lambda}_2$ are T_0-lines. Note that the stabilizer of $\tilde{A}$ in G is the infinite cyclic group generated by T_0. There exists a lift $\tilde{h}$ of h, uniquely defined, that fixes every point of $\tilde{\lambda}_1$. This lift coincides with $T_0^{r_{i_0}}$ or $T_0^{-r_{i_0}}$ on $\tilde{\lambda}_2$. Replacing r_{i_0} with $-r_{i_0}$ if necessary, we can suppose that we are in the first case. The map $\tilde{h}$ fixes every point of the unique connected component of $\tilde{\pi}^{-1}(S \setminus \cup_{i \in I} A_i)$ whose closure contains $\tilde{\lambda}_1$. This component is included in $R(\tilde{\lambda}_1)$ and its closure in $\overline{\tilde{S}}$ meets a (unique) component $\tilde{J}_1$ of $\partial \tilde{S} \setminus \{\alpha(T_0), \omega(T_0)\}$ because no connected component of $S \setminus \cup_{i \in I} A_i$ is an annulus. Consequently, the extension of $\tilde{h}$ to $\overline{\tilde{S}}$, still denoted $\tilde{h}$, admits fixed points in $\tilde{J}_1$. For the same reason $\tilde{h} \circ T_0^{-r_{i_0}}$ admits fixed points in the other component, denoted $\tilde{J}_2$.

Note that $\tilde{h}$ commutes with T_0 and so lifts a homeomorphism $\hat{h}$ of the open annulus $\hat{S} = \tilde{S}/T_0$ that preserves the orientation and fixes the two ends of $\hat{S}$. The map $\hat{h}$ can be extended to the compact annulus $\overline{\hat{S}} = \left(\overline{\tilde{S}} \setminus \{\alpha(T_0), \omega(T_0)\}\right)/T_0$. It contains fixed points on the added circles $\hat{J}_1 = \tilde{J}_1/T_0$ and $\hat{J}_2 = \tilde{J}_2/T_0$, the fixed points in $\hat{J}_1$ being lifted to fixed points of $\tilde{h}$, the fixed points in $\hat{J}_2$ being lifted to fixed points of $\tilde{h} \circ T_0^{-r_{i_0}}$.

The map f being isotopic to h admits a unique lift $\tilde{f}$ such that $[\tilde{f}] = [\tilde{h}]$ and this lift can be extended to a homeomorphism of $\overline{\tilde{S}}$ that coincides with $\tilde{h}$ on $\partial \tilde{S}$. The map $\tilde{f}$ commutes with T_0 and lifts a homeomorphism $\hat{f}$ of $\hat{S}$. The map $\hat{f}$ can be extended to a homeomorphism of $\overline{\hat{S}}$ that coincides with $\hat{h}$ on $\hat{J}_1$ and $\hat{J}_2$. Consequently, $\hat{f}$ admits fixed points on the boundary circles of $\overline{\hat{S}}$, the ones on $\widehat{J_1}$ having a rotation number equal to zero for the lift $\tilde{f}$, the ones on $\widehat{J_2}$ having a rotation number equal to r_{i_0} for the lift $\tilde{f}$: the map $\hat{f}$ satisfies a boundary twist condition.

Proposition 3.1. *At least one of the following situations holds:*

(1) *the map f has periodic points of period arbitrarily large;*
(2) *there exists an essential simple loop $\hat{\lambda} \subset \hat{S}$ such that $\hat{f}(\hat{\lambda}) \cap \hat{\lambda} = \emptyset$.*

This result was stated (in a weaker form) in Proposition 7.2 of [20], asking for infinitely many periodic points instead of periodic points of period arbitrarily large. The proof below is a slight modification of the proof in [20].

Proof. By Theorem 2.4, one knows that if (2) is not true, then for every $p/q \in [0, r_0]$, written in an irreducible way, there exists a periodic point $\hat{z}$ of period q and rotation number p/q for the lift $\tilde{f}$. One can easily prove (see [20]) that for every non trivial compact interval $J \subset (0, r_0)$, there exists a compact set $K \subset \hat{S}$ such that every periodic orbit of rotation number $p/q \in J$ meets K. Let $(p_m/q_m)_{m \geq 0}$ be a sequence in J, such that the sequence $(q_m)_{m \geq 0}$ is increasing and consists of prime numbers. For every $m \geq 0$, choose a periodic point $\hat{z}_m \in K$ of period q_m and rotation number p_m/q_m. Taking a subsequence if necessary, one can suppose that the sequence $(\hat{z}_m)_{m \geq 0}$ converges to a point $\hat{z}$. Moreover, $\hat{z}_m$ projects onto a point $z_m \in S$ satisfying $f^{q_m}(z_m) = z_m$. The integer q_m being prime, z_m has period q_m if it is not fixed. To prove the proposition it remains to show that z_m is not fixed if m is large enough. If it is not the case, taking a sub-sequence if necessary, one can suppose that all z_m are fixed. The sequence $(z_m)_{m\geq 0}$ converges to z, the projection of $\hat{z}$, and this point is a fixed point. Take a lift $\tilde{z} \in \tilde{S}$ of $\hat{z}$. There exists a sequence $(T_n)_{n\geq 1}$ in G such that $\tilde{f}^n(\tilde{z}) = T_n(\tilde{z})$. But if m is large enough and $\tilde{z}_m$ is the lift of $\hat{z}_m$ that is close to $\tilde{z}$, then one has $\tilde{f}^n(\tilde{z}_m) = T_n(\tilde{z}_m)$, for every $n \geq 1$. This is impossible because the integers q_m are all distinct. The fact that $\hat{f}^{q_m}(\hat{z}_m) = \hat{z}_m$ implies that T_{q_m} is a multiple of T_0; the fact that $\hat{f}^{q_m}(\hat{z}_{m'}) \neq \hat{z}_{m'}$ if $m' \neq m$, implies the opposite. □

By Proposition 3.1, it is sufficient to prove the following result to get Proposition 1.4:

Proposition 3.2. *If f is non wandering and if there exists an essential simple loop $\hat{\lambda}$ such that $\hat{f}(\hat{\lambda}) \cap \hat{\lambda} = \emptyset$, then f has periodic points of period arbitrarily large.*

Let us begin by proving:

Lemma 3.3. *Suppose that f is non wandering and that there exists an essential simple loop $\hat{\lambda}$ such that $\hat{f}(\hat{\lambda}) \cap \hat{\lambda} = \emptyset$. Then:*

(1) *the annulus A_{i_0} does not separate S (its complement is connected);*
(2) *the loop $\hat{\lambda}$ projects onto a simple loop $\check{\lambda} \subset \check{S}$ such that $\check{f}(\check{\lambda}) \cap \check{\lambda} = \emptyset$, where $\check{S}$ is the cyclic cover of S naturally associated to A_{i_0} and $\check{f}$ the homeomorphism of $\check{S}$ lifted by $\hat{f}$.*

The surface $\check{S}$ is the normal covering space of S, whose group of automorphisms is infinite cyclic, and such that the preimage of A_{i_0} by the covering projection is the union of disjoint separating annuli homeomorphic to A_{i_0}. The result is proved in Proposition 7.6 of [20] assuming f lets invariant a totally supported finite measure. We just need to verify that the arguments given in [20] are still valid in this wider situation.

Proof. The proof given in [20] is based on the fact that if the conclusions of Lemma 3.3 are not satisfied, there exists a non empty wandering open set $\hat{U}$ of $\hat{f}$ such that $\bigcup_{n\geq 0} \hat{f}^n(\hat{U})$ or $\bigcup_{n\geq 0} \hat{f}^{-n}(\hat{U})$ is relatively compact. So, by Proposition 2.3, the proof extends to the case of a non wandering homeomorphism. Let us explain briefly the arguments.

The loop $\hat{\lambda}$ can be lifted to a T_0-line $\tilde{\lambda}_0$ of $\tilde{S}$. The orientation of $\tilde{\lambda}_0$ induces an orientation on $\hat{\lambda}$. The loops $\hat{f}(\hat{\lambda})$ and $\hat{f}^{-1}(\hat{\lambda})$ belong to different components of $\hat{S}\backslash\hat{\lambda}$. Replacing f with f^{-1} if necessary, one can suppose that $\hat{f}(\hat{\lambda})$ is on the left of $\hat{\lambda}$ and $\hat{f}^{-1}(\hat{\lambda})$ on its right. This implies that $\tilde{\lambda}_0$ is a Brouwer line of $\tilde{f}$. The lift $\tilde{f}$ acts on the set of lifts of $\hat{\lambda}$ in a natural way. Indeed if $\tilde{\lambda} = T\tilde{\lambda}_0$, $T \in G$, is another lift, then $\tilde{\lambda}$ is a TT_0T^{-1}-line and one can define

$$[\tilde{f}](\tilde{\lambda}) = [\tilde{f}](T)(\tilde{\lambda}_0) = \tilde{f} \circ T \circ \tilde{f}^{-1}(\tilde{\lambda}_0).$$

The line $\tilde{\lambda}_0$ being a Brouwer line of $\tilde{f}$, it holds that:

$$\tilde{f}\left(\overline{L(\tilde{\lambda})}\right) \subset L([\tilde{f}](\tilde{\lambda})), \quad \tilde{f}^{-1}\left(\overline{R(\tilde{\lambda})}\right) \subset R([\tilde{f}]^{-1}(\tilde{\lambda})).$$

Define Λ^- as being the set of lifts $\tilde\lambda = T(\tilde\lambda_0)$ such that:

- $\alpha(TT_0T^{-1})$ and $\omega(TT_0T^{-1})$ are on the right of $\tilde\lambda_0$;
- $\alpha(T_0)$ and $\omega(T_0)$ are on the right of $\tilde\lambda$.

Similarly, define Λ^+ as being the set of lifts $\tilde\lambda = T(\tilde\lambda_0)$ such that:

- $\alpha(TT_0T^{-1})$ and $\omega(TT_0T^{-1})$ are on the left of $\tilde\lambda_0$;
- $\alpha(T_0)$ and $\omega(T_0)$ are on the left of $\tilde\lambda$.

It is not very difficult to see that the conclusions of Lemma 3.3 will occur if we prove that $\tilde\lambda\cap\tilde\lambda_0=\emptyset$ for every $\tilde\lambda\in\Lambda_-\cup\Lambda_+$. Indeed, for every couple $(T,T')\in G^2$, let us define the integer $T\wedge T'$ as follows. Choose $\tilde z\in\mathbb{D}$ and $\tilde z'\in\mathbb{D}$, then a path $\tilde\gamma$ joining $\tilde z$ to $T(\tilde z)$ and a path $\tilde\gamma'$ joining $\tilde z'$ to $T'(\tilde z')$. Denote Γ and Γ' the loops obtained by projecting $\tilde\gamma$ and $\tilde\gamma'$ in S, and define $T\wedge T'\in\mathbb{Z}$ to be the algebraic intersection number between Γ and Γ'. In particular, it holds that $\check S=\tilde S/\check G$, where $\check G=\{T\in G\,|\,T\wedge T_0=0\}$. If we knows that $\tilde\lambda\cap\tilde\lambda_0=\emptyset$ for every $\tilde\lambda\in\Lambda_-\cup\Lambda_+$, we can deduce that $T\wedge T_0\neq 0$, if $T(\tilde\lambda_0)\cap\tilde\lambda_0\neq\emptyset$. This implies that $\tilde\lambda_0$ projects onto a simple loop of $\check S$.

Now fix $\tilde\lambda\in\Lambda_-$. It projects onto a line of $\hat S$, denoted $\hat{\tilde\lambda}$, that joins the end on the right of $\hat\lambda$ to itself. It separates $\hat S$ into two components, $R(\hat{\tilde\lambda})$ on its right and $L(\hat{\tilde\lambda})$ on its left. Of course, one has $\widehat{T_0(\tilde\lambda)}=\hat{\tilde\lambda}$. There are finitely many lines $\hat{\tilde\lambda}$, $\tilde\lambda\in\Lambda_-$, that meet $\hat\lambda$, we want to prove that none of them does it. The set

$$\hat K=\overline{L(\hat\lambda)}\cap\left(\bigcup_{\tilde\lambda\in\Lambda_-}\overline{L(\hat{\tilde\lambda})}\right)$$

is compact and satisfies $\hat f(\hat K)\subset\operatorname{int}(\hat K)$. Indeed, if $\hat z\in\overline{L(\hat\lambda)}\cap\overline{L(\hat{\tilde\lambda})}$, then $\hat f(\hat z)\in L(\hat\lambda)\cap L(\widehat{[f](\tilde\lambda)})$. In case, there exists $\tilde\lambda\in\Lambda_-$ such that $\tilde\lambda\cap\tilde\lambda_0\neq\emptyset$, we deduce that $\hat U=\operatorname{int}(\hat K)\backslash\hat f(\hat K)$ is a non empty wandering open set of $\hat f$ such that $\bigcup_{n\geq 0}\hat f^n(\hat U)$ is contained in $\hat K$. We can apply Proposition 2.3 to get a contradiction. Similarly, we can construct a non empty wandering open set $\hat U$ of $\hat f$ such that $\bigcup_{n\geq 0}\hat f^{-n}(\hat U)$ is relatively compact in case there exists $\tilde\lambda\in\Lambda_+$ such that $\tilde\lambda\cap\tilde\lambda_0\neq\emptyset$. □

We will suppose until the end of the section that f satisfies the hypothesis of Proposition 3.2 and Lemma 3.3: f is non wandering and there exists an essential simple loop $\hat\lambda$ such that $\hat f(\hat\lambda)\cap\hat\lambda=\emptyset$. More precisely we will suppose that $\hat f(\hat\lambda)$ is on the left of $\hat\lambda$ and $\hat f^{-1}(\hat\lambda)$ on its right, meaning that the lift $\tilde\lambda_0$ of $\hat\lambda$ that is a T_0-line, is a Brouwer line of $\tilde f$.

We begin using Lemma 3.3. Let $\check T$ be the generator of the group of covering automorphisms of the covering map $\check\pi:\check S\to S$ such that $\check T(\check\lambda_1)$ is on the left of $\check\lambda_1$, where $\check\lambda_1$ is a lift of λ_1 (one of the boundary curves of A_{i_0}). It is possible that $\check T(\check\lambda)\cap\check\lambda\neq\emptyset$ but if $s\geq 1$ is large enough, then $\check T^s(\check\lambda)$ does not meet $\check\lambda$ and is on the left of $\check\lambda$. Replacing S with $\check S/\check T^s$ and f with the homeomorphism of $\check S/\check T^s$ that is lifted by $\check f$, we can always suppose that $s = 1$, meaning that $\hat{\lambda}$ projects onto a simple loop λ of S (indeed, by Proposition 2.2 the new map will be non wandering and if it has periodic points of period arbitrarily high, so will have f).

Let us explain now how we will apply the results of Sect. 2.6 (forcing theory). Considering the annulus $\overline{\hat{S}}$ and its universal covering space $\tilde{\overline{S}}\backslash\{\alpha(T_0), \omega(T_0)\}$, we can define the sets $\tilde{\mathcal{E}}_1$ and $\tilde{\mathcal{E}}_2$. Recall that there is a well defined relation $\prec$ on these sets. Every lift $\tilde{\lambda} \neq \tilde{\lambda}_0$ of λ is disjoint from $\tilde{\lambda}_0$. If it is on the right of $\tilde{\lambda}_0$ its two "ends" are on $\tilde{J}_1$, and so $\tilde{\lambda}$ belongs to $\tilde{\mathcal{E}}_1$, if it is on the left of $\tilde{\lambda}_0$ the ends are on $\tilde{J}_2$, and so $\tilde{\lambda}$ belongs to $\tilde{\mathcal{E}}_2$. For every integers $m < n$, denote $\check{S}_{[m,n]}$ the compact surface bordered by $\check{T}^m(\check{\lambda})$ and $\check{T}^n(\check{\lambda})$. Then note respectively $\tilde{S}_{[-1,0]}$ and $\tilde{S}_{[0,1]}$ the connected component of the preimage of $\check{S}_{[-1,0]}$ and $\check{S}_{[0,1]}$ by the universal covering projection that contains $\tilde{\lambda}_0$ in its boundary. The boundary of $\tilde{S}_{[-1,0]}$ is a disjoint union of lifts of λ, the ones that bound $\tilde{S}_{[-1,0]}$ on their right side being lifts of $\check{\lambda}$, the ones that bound $\tilde{S}_{[-1,0]}$ on their left side being lifts of $\check{T}^{-1}(\check{\lambda})$. Denote $\tilde{\mathcal{L}}_1$ the set of lifts of λ different from $\tilde{\lambda}_0$, that are on the boundary of $\tilde{S}_{[-1,0]}$. Denote $\tilde{\mathcal{L}}_{1,l} \subset \tilde{\mathcal{L}}_1$ the subset of lines in $\tilde{\mathcal{L}}_1$ that lift $\check{T}^{-1}(\check{\lambda})$ and $\tilde{\mathcal{L}}_{1,r} \subset \tilde{\mathcal{L}}_1$ the subset of lines in $\tilde{\mathcal{L}}_1$ that lift $\check{\lambda}$. Note also that the relation $\prec$ defines a total order on $\tilde{\mathcal{L}}_1$, setting

$$\tilde{\lambda}_1 \preceq \tilde{\lambda}'_1 \Leftrightarrow \tilde{\lambda}_1 \prec \tilde{\lambda}'_1 \quad \text{or} \quad \tilde{\lambda}_1 = \tilde{\lambda}'_1.$$

Similarly, denote $\tilde{\mathcal{L}}_2$ the set of lifts of λ different from $\tilde{\lambda}_0$ that are on the boundary of $\tilde{S}_{[0,1]}$, denote $\tilde{\mathcal{L}}_{2,l} \subset \tilde{\mathcal{L}}_2$ the subset of lines in $\tilde{\mathcal{L}}_2$ that lift $\check{\lambda}$ and $\tilde{\mathcal{L}}_{2,r} \subset \tilde{\mathcal{L}}_2$ the subset of lines in $\tilde{\mathcal{L}}_2$ that lift $\check{T}(\check{\lambda})$. Here again, $\prec$ induces a total order on $\tilde{\mathcal{L}}_2$. There is a natural action of $\mathbb{Z}^2$ on these sets. Note first that $\tilde{\mathcal{L}}_1$, $\tilde{\mathcal{L}}_{1,r}$, $\tilde{\mathcal{L}}_{1,l}$, $\tilde{\mathcal{L}}_2$, $\tilde{\mathcal{L}}_{2,r}$, $\tilde{\mathcal{L}}_{2,l}$, are invariant by T_0 because $\tilde{\lambda}_0$ is invariant by T_0. These sets are also invariant by the map $[\tilde{f}]$ defined on the set of lifts of λ. Moreover, T_0 and $[\tilde{f}]$ commute on our sets.

One deduces that:

$$\begin{aligned}
\tilde{\lambda}_1 \in \tilde{\mathcal{L}}_{1,r}, \quad \tilde{\lambda} \in \tilde{\mathcal{L}}_1 \cup \tilde{\mathcal{L}}_2, \quad k > 0 &\Rightarrow \tilde{f}^k(\tilde{\lambda}_1) \cap \tilde{\lambda} = \emptyset,\\
\tilde{\lambda}_1 \in \tilde{\mathcal{L}}_{1,l}, \quad \tilde{\lambda} \in \tilde{\mathcal{L}}_1 \cup \tilde{\mathcal{L}}_2, \quad k > 0 &\Rightarrow \tilde{f}^{-k}(\tilde{\lambda}_1) \cap \tilde{\lambda} = \emptyset,\\
\tilde{\lambda}_2 \in \tilde{\mathcal{L}}_{2,r}, \quad \tilde{\lambda} \in \tilde{\mathcal{L}}_1 \cup \tilde{\mathcal{L}}_2, \quad k > 0 &\Rightarrow \tilde{f}^k(\tilde{\lambda}_2) \cap \tilde{\lambda} = \emptyset,\\
\tilde{\lambda}_2 \in \tilde{\mathcal{L}}_{2,l}, \quad \tilde{\lambda} \in \tilde{\mathcal{L}}_1 \cup \tilde{\mathcal{L}}_2, \quad k > 0 &\Rightarrow \tilde{f}^{-k}(\tilde{\lambda}_2) \cap \tilde{\lambda} = \emptyset.
\end{aligned}$$

Consequently, it holds that

$$\begin{aligned}
\tilde{\lambda}_1 \in \tilde{\mathcal{L}}_{1,r}, \quad \tilde{\lambda}'_1 \in \tilde{\mathcal{L}}_{1,r}, \quad k \in \mathbb{Z}\backslash\{0\} &\Rightarrow \tilde{f}^k(\tilde{\lambda}_1) \cap \tilde{\lambda}'_1 = \emptyset,\\
\tilde{\lambda}_1 \in \tilde{\mathcal{L}}_{1,l}, \quad \tilde{\lambda}'_1 \in \tilde{\mathcal{L}}_{1,l}, \quad k \in \mathbb{Z}\backslash\{0\} &\Rightarrow \tilde{f}^k(\tilde{\lambda}_1) \cap \tilde{\lambda}'_1 = \emptyset,\\
\tilde{\lambda}_2 \in \tilde{\mathcal{L}}_{2,r}, \quad \tilde{\lambda}'_2 \in \tilde{\mathcal{L}}_{2,r}, \quad k \in \mathbb{Z}\backslash\{0\} &\Rightarrow \tilde{f}^k(\tilde{\lambda}_2) \cap \tilde{\lambda}'_2 = \emptyset,\\
\tilde{\lambda}_1 \in \tilde{\mathcal{L}}_{2,l}, \quad \tilde{\lambda}'_2 \in \tilde{\mathcal{L}}_{2,l}, \quad k \in \mathbb{Z}\backslash\{0\} &\Rightarrow \tilde{f}^k(\tilde{\lambda}_2) \cap \tilde{\lambda}'_2 = \emptyset.
\end{aligned}$$

Let us state now the key result implying Proposition 3.2. We will postpone its proof and begin by explaining how to get Proposition 3.2 from it.

Proposition 3.4. *There exist $\tilde\lambda_1 \in \tilde{\mathcal{L}}_{1,l}$, $\tilde\lambda_2 \in \tilde{\mathcal{L}}_{2,r}$, $n_0 \in \mathbb{N}$, $\rho_- \in \mathbb{R}$, $\rho_+ \in \mathbb{R}$, such that $\rho_- < \rho_+$ and such that for every $p \in \mathbb{Z}$ and every $q \geq 1$, it holds that*

$$\rho_-(n_0+q) < p < \rho_+(n_0+q) \Rightarrow \tilde f^{2n_0+q}(\tilde\lambda_1) \cap \tilde T_0^p(\tilde\lambda_2) \neq \emptyset.$$

Moreover, at least one of the two following properties holds:

(1) *If*

$$\rho_-(n_0+q) < p < p' < \rho_+(n_0+q),$$

then there exists a segment $\tilde\sigma_1 \subset \tilde\lambda_1$ such that:
- *$\tilde f^{2n_0+q}(\tilde\sigma_1)$ joins $T_0^p(\tilde\lambda_2)$ and $T_0^{p'}(\tilde\lambda_2)$;*
- *the interior of $\tilde f^{2n_0+q}(\tilde\sigma_1)$ is included in $R(T_0^p(\tilde\lambda_2)) \cap R(T_0^{p'}(\tilde\lambda_2))$;*
- *$\tilde f^{2n_0+q}(\tilde\sigma_1)$ is included in $\bigcup_{k\geq 0} \tilde f^{-k}(L(\tilde\lambda_0))$.*

(2) *If*

$$\rho_-(n_0+q) < p < p' < \rho_+(n_0+q),$$

then there exists a segment $\tilde\sigma_2 \subset \tilde\lambda_2$ such that:
- *$\tilde f^{-2n_0-q}(\tilde\sigma_2)$ joins $T_0^{-p}(\tilde\lambda_1)$ and $T_0^{-p'}(\tilde\lambda_1)$;*
- *the interior of $\tilde f^{-2n_0-q}(\tilde\sigma_2)$ is included in $L(T_0^{-p}(\tilde\lambda_1)) \cap L(T_0^{-p'}(\tilde\lambda_1))$;*
- *$\tilde f^{-2n_0-q}(\tilde\sigma_2)$ is included in $\bigcup_{k\geq 0} \tilde f^{k}(R(\tilde\lambda_0))$.*

We will suppose from now on, and until the end of the proof of Proposition 3.2, that Proposition 3.4 is true. We can deduce the following:

Proposition 3.5. *At least one of the two following statements is true:*

(1) *There exists $\tilde\lambda_1 \in \tilde{\mathcal{L}}_{1,l}$ such that for every $s \geq 2$, there exists $m_s \geq 0$ such that for every $m \geq m_s$, there exists $\tilde\lambda_2 \in \tilde{\mathcal{L}}_{2,r}$ such that for every $0 < p < p' \leq s$, there exists a segment $\tilde\sigma_1 \subset \tilde\lambda_1$ satisfying:*
- *$\tilde f^m(\tilde\sigma_1)$ joins $T_0^p(\tilde\lambda_2)$ and $T_0^{p'}(\tilde\lambda_2)$;*
- *the interior of $\tilde f^m(\tilde\sigma_1)$ is included in $R(T_0^p(\tilde\lambda_2)) \cap R(T_0^{p'}(\tilde\lambda_2))$;*
- *$\tilde f^m(\tilde\sigma_1)$ is included in $L(\tilde\lambda_0)$.*

(2) *There exists $\tilde\lambda_2 \in \tilde{\mathcal{L}}_{2,r}$ such that for every $s \geq 2$ there exists $m_s \geq 0$ such that for every $m \geq m_s$, there exists $\tilde\lambda_1 \in \tilde{\mathcal{L}}_{1,l}$ such that for every $0 < p < p' \leq s$, there exists a segment $\tilde\sigma_2 \subset \tilde\lambda_2$ satisfying:*
- *$\tilde f^{-m}(\tilde\sigma_2)$ joins $T_0^{-p}(\tilde\lambda_1)$ and $T_0^{-p'}(\tilde\lambda_1)$;*
- *the interior of $\tilde f^{-m}(\tilde\sigma_2)$ is included in $L(T_0^{-p}(\tilde\lambda_1)) \cap L(T_0^{-p'}(\tilde\lambda_1))$;*
- *$\tilde f^{-m}(\tilde\sigma_2)$ is included in $R(\tilde\lambda_0)$.*

Proof. Suppose that the first situation of Proposition 3.4 holds. Fix $s \geq 2$ and denote $\mathcal{K}$ the (finite) set of couples (p,p') such that $0 < p < p' \leq s$. There exist $b \in \mathbb{Z}$ and $q \geq 1$ such that, for every $(p,p') \in \mathcal{K}$ it holds that

$$\rho_-(n_0+q) < p+b < p'+b < \rho_+(n_0+q).$$

For each couple $\kappa = (p,p') \in \mathcal{K}$, one can choose a segment $\tilde\sigma_{1,\kappa}$ satisfying the three items of Proposition 3.4 (1). The third item tells us that there exists $a_\kappa \geq 0$ such that $\tilde{f}^{a+2n_0+q}(\tilde{\sigma}_{1,\kappa}) \subset L(\tilde{\lambda}_0)$ if $a \geq a_\kappa$. Moreover $\tilde{f}^{a+2n_0+q}(\tilde{\sigma}_{1,\kappa})$ joins $L([\tilde{f}]^a(T_0^{p+b}(\tilde{\lambda}_2)))$ and $L([\tilde{f}]^a(T_0^{p'+b}(\tilde{\lambda}_2)))$ and so contains a segment that joins $[\tilde{f}]^a(T_0^{p+b}(\tilde{\lambda}_2))$ and $[\tilde{f}]^a(T_0^{p'+b}(\tilde{\lambda}_2))$ and whose interior is included in $R([\tilde{f}]^a(T_0^{p+b}(\tilde{\lambda}_2)))$ and $R([\tilde{f}]^a(T_0^{p'+b}(\tilde{\lambda}_2)))$. This segment can be written $\tilde{f}^{a+2n_0+q}(\tilde{\sigma}'_{1,\kappa})$ where $\tilde{\sigma}'_{1,\kappa}$ is a segment included in $\tilde{\sigma}_{1,\kappa}$. Set $a_{\max} = \max_{\kappa\in K} a_\kappa$ and $m_s = a_\kappa + 2n_0 + q$. For every $m \geq m_s$ and for every $\kappa = (p,p') \in \mathcal{K}$, there exists a segment $\tilde{\sigma}'_{1,\kappa} \subset \tilde{\lambda}_1$ satisfying:

- $\tilde{f}^m(\tilde{\sigma}'_{1,\kappa})$ joins $T_0^p([\tilde{f}]^{m-2n_0-q}(T_0^b(\tilde{\lambda}_2)))$ and $T_0^{p'}([\tilde{f}]^{m-2n_0-q}(T_0^b(\tilde{\lambda}_2)))$;
- the interior of $\tilde{f}^m(\tilde{\sigma}'_{1,\kappa})$ is included in $R(T_0^p([\tilde{f}]^{m-2n_0-q}(T_0^b(\tilde{\lambda}_2))))$ and in $R(T_0^{p'}([\tilde{f}]^{m-2n_0-q}(T_0^b(\tilde{\lambda}_2))))$;
- $\tilde{f}^m(\tilde{\sigma}'_{1,\kappa})$ is included in $L(\tilde{\lambda}_0)$.

So (1) is satisfied. If the second situation of Proposition 3.4 holds, one prove similarly that (2) is satisfied. □

Recalling that m_5 has been defined in Proposition 3.5, finally we get:

Proposition 3.6. *There exists a sequence $(T_m)_{m\geq m_5}$ in G such that each T_m sends $\tilde{S}_{[-1,0]}$ onto $\tilde{S}_{[0,1]}$ and such that $\tilde{f}^m \circ T_m^{-1}$ has a fixed point.*

Proof. Suppose that the first item of Proposition 3.5 holds and denote $\tilde{\lambda}_1$ the line of $\tilde{\mathcal{L}}_{1,l}$ defined in Proposition 3.5. Fix $m \geq m_5$ and denote $\tilde{\lambda}_2$ the line of $\tilde{\mathcal{L}}_{2,r}$ defined in Proposition 3.5. If $T \in G$ sends $\tilde{\lambda}_1$ onto $\tilde{\lambda}_0$, it sends $\tilde{S}_{[-1,0]}$ onto $\tilde{S}_{[0,1]}$ and $T_0(\tilde{\lambda}_1)$ onto an element of $\tilde{\mathcal{L}}_{2,l}$. Moreover $T' \in G$ sends $\tilde{\lambda}_1$ onto $\tilde{\lambda}_0$ if and only if there exists $k \in \mathbb{Z}$ such that $T' = T_0^k T$. So, there exists an automorphism $T_m \in G$, uniquely defined, such that

$$T_m(\tilde{\lambda}_1) = \tilde{\lambda}_0 \quad \text{and} \quad T_0^3(\tilde{\lambda}_2) \prec T_m(T_0(\tilde{\lambda}_1)) \prec T_0^4(\tilde{\lambda}_2),$$

and T_m sends $\tilde{S}_{[-1,0]}$ onto $\tilde{S}_{[0,1]}$. Note that T_m sends $\tilde{\lambda}_0$ onto an element of $\tilde{\mathcal{L}}_{2,r}$. Consider the map $\tilde{g} = \tilde{f}^m \circ T_m^{-1}$. Every line $\tilde{\lambda}'_2 \in \tilde{\mathcal{L}}_{2,r} \backslash T_m(\tilde{\lambda}_0)$ is sent by T_m^{-1} onto an element of $\tilde{\mathcal{L}}_{1,r}$ and its image by $\tilde{g}$ belongs to $L([\tilde{f}]^m(T_m^{-1}(\tilde{\lambda}'_2)))$. In particular, $\tilde{g}(\tilde{\lambda}'_2) \cap \tilde{\lambda}'_2 = \emptyset$. Every line $\tilde{\lambda}'_2 \in \tilde{\mathcal{L}}_{2,l}$ is sent by T_m^{-1} onto an element of $\tilde{\mathcal{L}}_{1,l}$ and its image by $\tilde{f}^{-m}$ belongs to $R([\tilde{f}]^{-m}(\lambda'_2))$. The images of $\tilde{\lambda}'_2$ by T_m^{-1} and by $\tilde{f}^{-m}$ being disjoint, it holds that $\tilde{g}(\tilde{\lambda}'_2) \cap \tilde{\lambda}'_2 = \emptyset$. In fact the only lines on the boundary of $\tilde{S}_{[0,1]}$ that can meet their image by $\tilde{g}$ are $\tilde{\lambda}_0$ and $T_m(\tilde{\lambda}_0)$. One can choose $i \in \{2,3\}$ and $j \in \{4,5\}$ such that $T_0^i(\tilde{\lambda}_2) \neq T_m(\tilde{\lambda}_0)$ and $T_0^j(\tilde{\lambda}_2) \neq T_m(\tilde{\lambda}_0)$. Applying Proposition 3.5 to the pairs (i,j) and $(i-1,j-1)$ and using the fact that $\tilde{f}$ and T_0 commute, we deduce that there exist a segment $\tilde{\sigma}_1 \subset \tilde{\lambda}_1$ and a segment $\tilde{\sigma}'_1 \subset T_0(\tilde{\lambda}_1)$ satisfying:

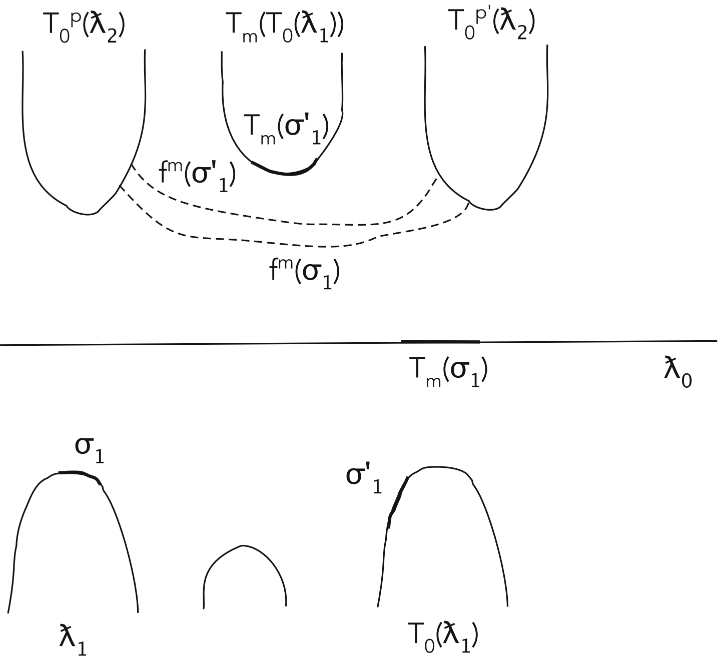
- $\tilde{f}^m(\tilde{\sigma}_1)$ and $\tilde{f}^m(\tilde{\sigma}'_1)$ join $T_0^i(\tilde{\lambda}_2)$ and $T_0^j(\tilde{\lambda}_2)$;
- the interior of $\tilde{f}^m(\tilde{\sigma}_1)$ and $\tilde{f}^m(\tilde{\sigma}'_1)$ are included in $R(T_0^i(\tilde{\lambda}_2))$ and in $R(T_0^j(\tilde{\lambda}_2))$;
- $\tilde{f}^m(\tilde{\sigma}_1)$ and $\tilde{f}^m(\tilde{\sigma}'_1)$ are included in $L(\tilde{\lambda}_0)$.

Let us summarize the situation:

- $T_m(\tilde{\sigma}_1)$ is a segment of $\tilde{\lambda}_0$ whose image by $\tilde{g}$ is disjoint from $\tilde{\lambda}_0$ and joins $T_0^i(\tilde{\lambda}_2)$ and $T_0^j(\tilde{\lambda}_2)$;
- $T_m(\tilde{\sigma}_1')$ is a segment of $T_mT_0(\tilde{\lambda}_1)$ whose image by $\tilde{g}$ is disjoint from $T_mT_0(\tilde{\lambda}_1)$ and joins $T_0^i(\tilde{\lambda}_2)$ and $T_0^j(\tilde{\lambda}_2)$;
- $T_0^i(\tilde{\lambda}_2)$ and $T_0^j(\tilde{\lambda}_2)$ are disjoint from there image by $\tilde{g}$;
- the lines $\tilde{\lambda}_0$, $T_0^j(\tilde{\lambda}_2)$, $T_mT_0(\tilde{\lambda}_1)$ and $T_0^i(\tilde{\lambda}_2)$ are cyclically ordered.

We can use Proposition 2.5 and deduce that g has a fixed point.

In the case where the second item of Proposition 3.5 holds, we can prove similarly that for every $m \geq m_5$, there exists $T_m \in G$ that sends $\tilde{S}_{[-1,0]}$ onto $\tilde{S}_{[0,1]}$, such that $\tilde{f}^{-m} \circ T_m$ has a fixed point. Note that the image by T_m of the fixed point of $\tilde{f}^{-m} \circ T_m$ is a fixed point of $\tilde{f}^m \circ T_m^{-1}$.



□

It remains to explain why Proposition 3.6 implies Proposition 3.2:

Proof of Proposition 3.2. For every $m \geq m_5$, one can find $\tilde{z}_m \in \tilde{S}$ such that $\tilde{f}^m(\tilde{z}_m) = T_m(\tilde{z}_m)$ (just take the image by T_m^{-1} of a fixed point of $\tilde{f}^m \circ T_m^{-1}$). It projects onto a point $\check{z}_m \in \check{S}$ such that $\check{f}^m(\check{z}_m) = \check{T}(\check{z}_m)$ which itself projects onto a point $z_m \in S$ such that $f^m(z_m) = z_m$. It remains to show that the period of z_m is m. Suppose that $f^q(z_m) = z_m$, where $q|m$. Then there exists $p \in \mathbb{Z}$ such that $\check{f}^q(\check{z}_m) = \check{T}^p(\check{z}_m)$ and so $\check{f}^m(\check{z}_m) = \check{T}^{pm/q}(\check{z}_m)$, which implies that $pm = q$. Of course this implies that $m = q$ and $p = 1$. □

To conclude the section, it remains to prove Proposition 3.4. Let us begin with the following result:

Lemma 3.7. *There exists* $\tilde{\lambda}_1 \in \tilde{\mathcal{L}}_{1,l}$, $\tilde{\lambda}_2 \in \tilde{\mathcal{L}}_{2,r}$ *and* $n \geq 1$ *such that* $\tilde{f}^n(\tilde{\lambda}_1) \cap \tilde{\lambda}_2 \neq \emptyset$.

Proof. One can find an open disk $\check{U} \subset \check{S}_{[-1,0]}$ whose image by $\check{f}$ is on the left of $\check{\lambda}$. It is a wandering disk of $\check{f}$ that projects onto an open disk U of S. There is a lift $\tilde{U} \subset \tilde{S}_{[-1,0]}$, uniquely defined up to the action of the iterates of T_0, whose image by $\tilde{f}$ is on the left of $\tilde{\lambda}_0$. By Lemma 2.1, there exists $z \in U$ and positive integers n_0, n_1, n_2 such that $f^{-n_0}(z)$, $f^{n_1}(z)$ and $f^{n_1+n_2}(z)$ belong to U. If $\check{z}$ is the lift of z that belongs to $\check{U}$, then $\check{f}^{-n_0}(\check{z})$ is on the right of $\check{T}^{-1}(\check{\lambda})$, $\check{f}^{n_1}(\check{z})$ on the left of $\check{\lambda}$ and $\check{f}^{n_1+n_2}(\check{z})$ on the left of $\check{T}(\check{\lambda})$. So, if $\tilde{z}$ is the lift of $\check{z}$ that belongs to $\tilde{U}$, there exists $\tilde{\lambda}_1 \in \tilde{\mathcal{L}}_{1,l}$ and $\tilde{\lambda}_2 \in \tilde{\mathcal{L}}_{2,r}$ such that $\tilde{f}^{-n_0}(\tilde{z}) \in R(\tilde{\lambda}_1)$ and $\tilde{f}^{n_1+n_2}(\tilde{z}) \in L(\tilde{\lambda}_2)$. Setting $n = n_0 + n_1 + n_2$, one gets that $\tilde{f}^n(R(\tilde{\lambda}_1)) \cap L(\tilde{\lambda}_2) \neq \emptyset$, which implies that $\tilde{f}^n(\tilde{\lambda}_1) \cap \tilde{\lambda}_2 \neq \emptyset$. □

Proof of Proposition 3.4. We consider the annuli:

$$\hat{U}_1 = \bigcup_{n\geq 0} \hat{f}^{-n}(L(\hat{\lambda})), \quad \hat{U}_2 = \bigcup_{n\geq 0} \hat{f}^{n}(R(\hat{\lambda})), \quad \hat{U}_0 = \hat{U}_1 \cap \hat{U}_2,$$

and the respective covering spaces:

$$\tilde{U}_1 = \bigcup_{n\geq 0} \tilde{f}^{-n}(L(\tilde{\lambda}_0)), \quad \tilde{U}_2 = \bigcup_{n\geq 0} \tilde{f}^{n}(R(\tilde{\lambda}_0)), \quad \tilde{U}_0 = \tilde{U}_1 \cap \tilde{U}_2.$$

The three annuli are invariant by $\hat{f}$. By Caratheodory's Prime End Theory (see [24] for instance), each annulus $\hat{U}_0$, $\hat{U}_1$, $\hat{U}_2$, can be compactified as an annulus in such a way that the restriction of $\hat{f}$ to the former annulus extends to a homeomorphism of the compact annulus. More precisely, to compactify $\hat{U}_0$, one must add the circle of prime ends $\hat{S}_1$ corresponding to the end on the right of $\hat{\lambda}$ and the circle of prime ends $\hat{S}_2$ corresponding to the end on the left of $\hat{\lambda}$; to compactify $\hat{U}_1$, one must add $\hat{S}_1$ and $\hat{J}_2$; to compactify $\hat{U}_2$, one must add $\hat{J}_1$ and $\hat{S}_2$. Furthermore, one can add the covering spaces $\tilde{S}_1$ and $\tilde{S}_2$ of $\hat{S}_1$ and $\hat{S}_2$ to $\tilde{U}_0$ in such a way that the restriction of $\tilde{f}$ extends continuously to the added space. Similarly, one can add $\tilde{S}_1$, $\tilde{J}_2$ to $\tilde{U}_1$ and $\tilde{J}_1$, $\tilde{S}_2$ to $\tilde{U}_2$. For every $i \in \{0, 1, 2\}$, one can consider the sets $\tilde{\mathcal{E}}_1(\tilde{U}_i)$ and $\tilde{\mathcal{E}}_2(\tilde{U}_i)$ like in Sect. 2.6, noting that $\tilde{\mathcal{E}}_1(\tilde{U}_0)$ is a subset of $\tilde{\mathcal{E}}_1(\tilde{U}_1)$ and $\tilde{\mathcal{E}}_2(\tilde{U}_0)$ a subset of $\tilde{\mathcal{E}}_2(\tilde{U}_2)$. One can define the rotation numbers ρ_1, ρ_2, defined respectively by $\tilde{f}$ on each spaces $\tilde{S}_1$ and $\tilde{S}_2$. Note that at least one of the following conditions is satisfied:

$$\rho_1 \neq r_{i_0}, \quad \rho_2 \neq 0, \quad \rho_1 \neq \rho_2,$$

meaning that a boundary twist condition is satisfied on at least one annulus $\hat{U}_0$, $\hat{U}_1$ or $\hat{U}_2$. The sets $\tilde{U}_0 \cap \left(\bigcup_{\tilde{\lambda}\in\tilde{\mathcal{L}}_1} \tilde{\lambda}\right)$ and $\tilde{U}_1 \cap \left(\bigcup_{\tilde{\lambda}\in\tilde{\mathcal{L}}_1} \tilde{\lambda}\right)$ coincide, we note $\tilde{\mathcal{D}}_1$ the set of its connected components. Each element $\tilde{\delta}_1$ of $\tilde{\mathcal{D}}_1$ is contained in a line $\tilde{\lambda}_1 \in \tilde{\mathcal{L}}_1$, moreover it holds that $\tilde{\lambda}_1 \in \tilde{\mathcal{L}}_{1,l}$. Note that $\tilde{\delta}_1$ is a line of $\tilde{U}_1$ that has two different limit points in the frontier of $\tilde{U}_1$ in $\overline{\tilde{S}}$ (called accessible points). An important property of prime end theory is that $\tilde{\lambda}_1$ has two different limit points in $\tilde{S}_1$. In particular $\tilde{\mathcal{D}}_1$ is a subset of $\tilde{\mathcal{E}}_1(\tilde{U}_0)$

(and consequently of $\tilde{\mathcal{E}}_1(\tilde{U}_1)$). Moreover T_0 naturally acts on it. Note that for every $k \neq 0$, every $\delta_1 \in \tilde{\mathcal{D}}_1$, and every $\tilde{\delta}_1' \in \tilde{\mathcal{D}}_1$, one has $\tilde{f}^k(\delta_1) \cap \delta_1' = \emptyset$. Indeed, we have a similar properties in $\tilde{\mathcal{L}}_{1,l}$. This means that one gets a subset $\bigcup_{k\in\mathbb{Z}} \tilde{f}^k(\tilde{\mathcal{D}}_1)$ of $\tilde{\mathcal{E}}_1(\tilde{U}_0)$ invariant by T_0 and $\tilde{f}$ that consists of pairwise disjoint lines. Consequently, $\prec$ is transitive on $\bigcup_{k\in\mathbb{Z}} \tilde{f}^k(\tilde{\mathcal{D}}_1)$ and induces an order $\preceq$ which is not necessarily total. Of course T_0 and $\tilde{f}$ preserve the order and one has $\tilde{\delta} \prec T_0(\tilde{\delta})$ for every $\tilde{\delta} \in \bigcup_{k\in\mathbb{Z}} \tilde{f}^k(\tilde{\mathcal{D}}_1)$. We can define similarly $\tilde{\mathcal{D}}_2 \subset \tilde{\mathcal{E}}_2(\tilde{U}_0)$ with the same properties.

By Lemma 3.7, there exist $\tilde{\lambda}_1 \in \tilde{\mathcal{L}}_{1,l}$, $\tilde{\lambda}_2 \in \tilde{\mathcal{L}}_{2,r}$ and $n_0 \geq 1$ such that $\tilde{f}^{n_0}(\tilde{\lambda}_1) \cap \tilde{\lambda}_2 \neq \emptyset$. Choose $\tilde{z} \in \tilde{\lambda}_1 \cap \tilde{f}^{-n_0}(\tilde{\lambda}_2)$. Its orbit is included in $\tilde{U}_0$. So, there exists $\tilde{\delta}_1 \in \tilde{\mathcal{D}}_1$ and $\tilde{\delta}_2 \in \tilde{\mathcal{D}}_2$ such that $\tilde{z} \in \tilde{\delta}_1$ and $\tilde{f}^{n_0}(\tilde{z}) \in \tilde{\delta}_2$.

We will begin by studying the case where $\rho_1 \neq r_{i_0}$ and will prove that the first item of Proposition 3.4 is satisfied. Then we will study the case where $\rho_2 \neq 0$ and will prove that the second item is satisfied. Eventually we will look at the case where $\rho_1 = r_{i_0}$ and $\rho_2 = 0$ and will see that both items are satisfied.

Suppose first that $\rho_1 \neq r_{i_0}$. Note that all properties of Proposition 2.8 are satisfied with $\tilde{\delta}_1 \in \tilde{\mathcal{E}}_1(\tilde{U}_1)$ and $\tilde{\lambda}_2 \in \tilde{\mathcal{E}}_2(\tilde{U}_1)$. Setting $\rho_- = \min(\rho_1, r_{i_0})$ and $\rho_+ = \max(\rho_1, r_{i_0})$ one gets that

$$\rho_-(n_0+q) < p < \rho_+(n_0+q) \Rightarrow \tilde{f}^{2n_0+q}(\tilde{\delta}_1) \cap T_0^p(\tilde{\lambda}_2) \neq \emptyset.$$

So, if $\rho_-(n_0+q) < p < p' < \rho_+(n_0+q)$, there exists a segment $\tilde{\sigma}_1 \subset \tilde{\delta}_1 \subset \tilde{\lambda}_1$ such that:

- $\tilde{f}^{2n_0+q}(\tilde{\sigma}_1)$ joins $T_0^p(\tilde{\lambda}_2)$ and $T_0^{p'}(\tilde{\lambda}_2)$;
- the interior of $\tilde{f}^{2n_0+q}(\tilde{\sigma}_1)$ is included in $R(T_0^p(\tilde{\lambda}_2)) \cap R(T_0^{p'}(\tilde{\lambda}_2))$;
- $\tilde{f}^{2n_0+q}(\tilde{\sigma}_1)$ is included in $\bigcup_{k\leq 0} \tilde{f}^{-k}(L(\tilde{\lambda}_0))$.

Suppose now that $\rho_2 \neq 0$. The properties of Proposition 2.8 are satisfied with $\tilde{\lambda}_1 \in \tilde{\mathcal{E}}_1(\tilde{U}_2)$ and $\tilde{\delta}_2 \in \tilde{\mathcal{E}}_2(\tilde{U}_2)$. Setting $\rho_- = \min(\rho_2, 0)$ and $\rho_+ = \max(\rho_2, 0)$ one gets that

$$\rho_-(n_0+q) < p < \rho_+(n_0+q) \Rightarrow \tilde{f}^{2n_0+q}(\tilde{\lambda}_1) \cap T_0^p(\tilde{\delta}_2) \neq \emptyset.$$

So, if $\rho_-(n_0+q) < p < p' < \rho_+(n_0+q)$, there exists a segment $\tilde{\sigma}_2 \subset \tilde{\delta}_2 \subset \tilde{\lambda}_2$ such that:

- $\tilde{f}^{-2n_0-q}(\tilde{\sigma}_2)$ joins $T_0^{-p}(\tilde{\lambda}_1)$ and $T_0^{-p'}(\tilde{\lambda}_1)$;
- the interior of $\tilde{f}^{-2n_0-q}(\tilde{\sigma}_2)$ is included in $L(T_0^{-p}(\tilde{\lambda}_1)) \cap L(T_0^{-p'}(\tilde{\lambda}_1))$;
- $\tilde{f}^{-2n_0-q}(\tilde{\sigma}_2)$ is included in $\bigcup_{k\leq 0} \tilde{f}^{-k}(R(\tilde{\lambda}_0))$.

Suppose now that $\rho_1 = r_{i_0}$ and $\rho_2 = 0$. The properties of Proposition 2.8 are satisfied with $\tilde{\delta}_1 \in \tilde{\mathcal{E}}_1(\tilde{U}_0)$ and $\tilde{\delta}_2 \in \tilde{\mathcal{E}}_2(\tilde{U}_0)$. Setting $\rho_- = \min(0, r_{i_0})$ and $\rho_+ = \max(0, r_{i_0})$ one gets that

$$\rho_-(n_0+q) < p < \rho_+(n_0+q) \Rightarrow \tilde{f}^{2n_0+q}(\tilde{\delta}_1) \cap T_0^p(\tilde{\delta}_2) \neq \emptyset.$$

So, if $\rho_-(n_0+q) < p < p' < \rho_+(n_0+q)$, there exists a segment $\tilde{\sigma}_1 \subset \tilde{\delta}_1 \subset \tilde{\lambda}_1$ such that:

- $\tilde{f}^{2n_0+q}(\tilde{\sigma}_1)$ joins $T_0^p(\tilde{\lambda}_2)$ and $T_0^{p'}(\tilde{\lambda}_2)$;

- the interior of $\tilde{f}^{2n_0+q}(\tilde{\sigma}_1)$ is included in $R(T_0^p(\tilde{\delta}_2)) \cap R(T_0^{p'}(\tilde{\delta}_2))$.
- $\tilde{f}^{2n_0+q}(\tilde{\sigma}_1)$ is included in $\bigcup_{k\leq 0} \tilde{f}^{-k}(L(\tilde{\lambda}_0))$.

  Similarly, there exists a segment $\tilde{\sigma}_2 \subset \tilde{\delta}_2 \subset \tilde{\lambda}_2$ such that:
- $\tilde{f}^{-2n_0-q}(\tilde{\sigma}_2)$ joins $T_0^{-p}(\tilde{\lambda}_1)$ and $T_0^{-p'}(\tilde{\lambda}_1)$;
- the interior of $\tilde{f}^{-2n_0-q}(\tilde{\sigma}_2)$ is included in $L(T_0^{-p}(\tilde{\lambda}_1)) \cap L(T_0^{-p'}(\tilde{\lambda}_1))$.
- $\tilde{f}^{-2n_0-q}(\tilde{\sigma}_2)$ is included in $\bigcup_{k\leq 0} \tilde{f}^{-k}(R(\tilde{\lambda}_0))$. □

Remark. The boundary twist condition is satisfied on the whole space $\tilde{S}$. Setting $\rho_- = \min(0, r_{i_0})$ and $\rho_+ = \max(0, r_{i_0})$ one knows that

$$\rho_-(n_0+q) < p < \rho_+(n_0+q) \Rightarrow \tilde{f}^{2n_0+q}(\tilde{\lambda}_1) \cap T_0^p(\tilde{\lambda}_2) \neq \emptyset.$$

So, if $\rho_-(n_0+q) < p < p' < \rho_+(n_0+q)$, there exists a segment $\tilde{\sigma}_1 \subset \tilde{\lambda}_1$ such that:

- $\tilde{f}^{2n_0+q}(\tilde{\sigma}_1)$ joins $T_0^p(\tilde{\lambda}_2)$ and $T_0^{p'}(\tilde{\lambda}_2)$;
- the interior of $\tilde{f}^{2n_0+q}(\tilde{\sigma}_1)$ is included in $R(T_0^p(\tilde{\lambda}_2)) \cap R(T_0^{p'}(\tilde{\lambda}_2))$.

The problem is that we need the supplementary condition

- $\tilde{f}^{2n_0+q}(\tilde{\sigma}_1)$ is included in $\bigcup_{k\leq 0} \tilde{f}^{-k}(L(\tilde{\lambda}_0))$;

to make right the argument of the proof of Proposition 3.6. There is no reason why such a condition will be satisfied even if q is large enough. This is why we must work in $\tilde{U}_0$, $\tilde{U}_1$ or in $\tilde{U}_2$.

4. Homeomorphisms isotopic to the identity

The goal of this section is to prove Proposition 1.5, which means to prove that if S is an orientable closed surface of genus $g \geq 2$ and f a non wandering homeomorphism of S isotopic to the identity, then either f has periodic points of period arbitrarily large, or every periodic point of f is fixed and f is isotopic to the identity relative to its fixed point set (the existence of at least one fixed point being a consequence of Lefschetz formula).

Proof of Proposition 1.5. We keep the same notations as before. The map f being isotopic to the identity and the genus of S being larger than 1, there exists a unique lift $\tilde{f}$ of f to $\tilde{S}$ that commutes with the covering automorphisms. A periodic point of f that is lifted to a periodic point of $\tilde{f}$ is called a *contractible periodic point*. It was proved in [19] that if f has a contractible periodic point that is not fixed, or if f is not isotopic to the identity relative to the contractible fixed point set, then f has contractible periodic points of arbitrarily large period. So, Proposition 1.5 is an extension of this result and to obtain Proposition 1.5 it remains to show that f has periodic points of arbitrarily large period if there exists a non contractible periodic point. We fix such a point z_0, denoting by q_0 its period. If $\tilde{z}_0$ is a lift of z_0, there exists $T_0 \in G\backslash\{\mathrm{Id}\}$ such that $\tilde{f}^q(\tilde{z}_0) = T_0(\tilde{z}_0)$. The map $\tilde{f}$ commutes with T_0 and lifts a homeomorphism $\hat{f}$ of $\hat{S} = \tilde{S}/T_0$. Recall that $\tilde{f}$ extends to a homeomorphism of $\overline{\tilde{S}}$ that fixes every point of ∂S. Consequently $\hat{f}$ extends to a homeomorphism of the compact annulus $\overline{\hat{S}}$ obtained by adding the two circles $\hat{J}_1 = \tilde{J}_1/T_0$ and $\hat{J}_2 = \tilde{J}_2/T_0$, where $\tilde{J}_1$ and $\tilde{J}_2$ are the two connected components of $\partial S\backslash\{\alpha(T_0), \omega(T_0)\}$, the first one on the right of every T_0-line, the second on the left. Note that every point of $\hat{J}_1 \cup \hat{J}_2$ is fixed, with a rotation number equal to zero for the lift $\tilde{f}$. Note also that $\tilde{z}_0$ projected onto a periodic point $\hat{z}_0$ of period q_0 and rotation number $1/q_0$. One can apply Theorem 2.4:

- either, for every rational number p/q between 0 and $1/q_0$, written in an irreducible way, there exists a periodic point z of $\hat{f}$ of period q and rotation number p/q for $\tilde{f}$;
- or there exists an essential simple loop $\hat{\lambda} \subset \hat{S}$ such that $\hat{f}(\hat{\lambda}) \cap \hat{\lambda} = \emptyset$.

In case the first item holds, we can prove, as we did in Sect. 3 for a map isotopic to a Dehn twist map, that f has periodic points of period arbitrarily large. So, we can assume that the second item holds. Consider the lift $\tilde{\lambda}_0 \subset \tilde{S}$ of $\hat{\lambda}$, oriented as a T_0-line. Replacing f with f^{-1} if necessary, we can suppose that it is a Brouwer line. The fact that $\tilde{f}$ commutes with the covering automorphisms implies that every line $T(\tilde{\lambda}_0)$, $T \in G$, is a Brouwer line of $\tilde{f}$.

The loop $\hat{\lambda}$ projects onto a loop λ of S and there is a natural partition $\tilde{\mathcal{L}} = \tilde{\mathcal{L}}_0 \cup \tilde{\mathcal{L}}_1 \cup \tilde{\mathcal{L}}_2$ of the set of lifts of λ in $\tilde{S}$, defined as follows:

- every $\tilde{\lambda} \in \tilde{\mathcal{L}}_0$ meets $\tilde{\lambda}_0$;
- every $\tilde{\lambda} \in \tilde{\mathcal{L}}_1$ is included in $R(\tilde{\lambda}_0)$;
- every $\tilde{\lambda} \in \tilde{\mathcal{L}}_2$ is included $L(\tilde{\lambda}_0)$.

Note that the ends of a line in $\tilde{\mathcal{L}}_1$ belong to $\tilde{J}_1$ and the ends of a line in $\tilde{\mathcal{L}}_2$ belong to $\tilde{J}_2$. Moreover, we have two partitions $\tilde{\mathcal{L}}_1 = \tilde{\mathcal{L}}_{1,r} \cup \tilde{\mathcal{L}}_{1,l}$ and $\tilde{\mathcal{L}}_2 = \tilde{\mathcal{L}}_{2,r} \cup \tilde{\mathcal{L}}_{2,l}$, where:

- $\tilde{\lambda} \in \tilde{\mathcal{L}}_{1,r} \cup \tilde{\mathcal{L}}_{2,r}$ if $\tilde{\lambda}_0 \subset R(\tilde{\lambda})$;
- $\tilde{\lambda} \in \tilde{\mathcal{L}}_{1,l} \cup \tilde{\mathcal{L}}_{2,l}$ if $\tilde{\lambda}_0 \subset L(\tilde{\lambda})$.

Note that the subsets defined above are all invariant by T_0. A last important remark is the following: there exists $N > 0$ such that if $\tilde{\lambda}$ is a lift of λ, there exists at most N other lifts that meet $\tilde{\lambda}$ up to the action of T, where T is the generator of the stabilizer of $\tilde{\lambda}$ (in particular the number of T_0-orbits in $\tilde{\mathcal{L}}_0$ is bounded by N). Indeed fix a segment $\tilde{\delta} \subset \tilde{\lambda}_0$ joining a point $\tilde{z}$ to $T_0(\tilde{z})$ the set of $T \in G$ such that $T(\tilde{\delta}) \cap \tilde{\delta} \neq \emptyset$ is finite, choose N to be its cardinal.

Lemma 4.1. *There exist $\tilde{\lambda}_1 \in \tilde{\mathcal{L}}_{1,l}$, $\tilde{\lambda}_2 \in \tilde{\mathcal{L}}_{1,r}$ and $n_0 \geq 1$ such that $\tilde{f}^{n_0}(\tilde{\lambda}_1) \cap \tilde{\lambda}_2 \neq \emptyset$.*

Proof. Consider an open disk $\tilde{U} \subset R(\tilde{\lambda}_0) \cap \tilde{f}^{-1}(L(\tilde{\lambda}_0))$ that projects onto an open disk U of S. Using Lemma 2.1, there exists $z \in U$ and two increasing sequences $(m_i)_{0\leq k\leq N}$, $(m'_i)_{i\leq k\leq N}$, with $m_0 = m'_0 = 0$, such that for every $k \in \{0, \dots, N\}$, the points $f^{m_k}(z)$ and $f^{-m'_k}(z)$ belong to U. Let $\tilde{z}$ be the lift of z that belongs to $\tilde{U}$. So there exist two sequences $(T^{(k)})_{0\leq k\leq N}$, $(T'^{(k)})_{0\leq k\leq N}$ of covering automorphisms such that $\tilde{f}^{m_k}(\tilde{z}) \in T^{(k)}(\tilde{U})$ and $\tilde{f}^{-m'_k}(\tilde{z}) \in T'^{(k)}(\tilde{U})$. Suppose that there exist $0 \leq k < k' \leq N$ and $m \in \mathbb{Z}$

such that $T^{(k')} = T_0^m T^{(k)}$. Then it holds that $\tilde{f}^{m_k}(T^{(k)-1}(z)) \in \tilde{U}$ and $\tilde{f}^{m_{k'}}(T^{(k)-1}(z)) \in T_0^m(\tilde{U})$. We have a contradiction because $T_0^m(\tilde{U}) \subset R(\tilde{\lambda}_0)$ and $\tilde{f}^{m_{k'}-m_k}(\tilde{U}) \subset L(\tilde{\lambda}_0)$. But we have $T^{(0)} = \mathrm{Id}$, and so, there exists $1 \le k \le N$ such that $T^{(k)}(\tilde{\lambda}_0) \notin \tilde{\mathcal{L}}_0$. The lines $\tilde{\lambda}_0$ and $T^{(k)}(\tilde{\lambda}_0)$ being Brouwer lines, one deduces that $T^{(k)}(\tilde{\lambda}_0) \in \tilde{\mathcal{L}}_{2,r}$. For the same reasons, one can find $1 \le k' \le N$ such that $T'^{(k')}(\tilde{\lambda}_0) \in \tilde{\mathcal{L}}_{1,l}$. Setting $n_0 = m_k + m'_{k'} + 1$, one deduce that $\tilde{f}^{n_0}(R(T'^{(k')}(\tilde{\lambda}_0)) \cap L(T^{(k)}(\tilde{\lambda}_0)) \neq \emptyset$ which easily implies that $\tilde{f}^{n_0}(T'^{(k')}(\tilde{\lambda}_0)) \cap T^{(k)}(\tilde{\lambda}_0) \neq \emptyset$. □

We would like to give a proof similar to the proof in Sect. 3. The fact that $\tilde{z}_0$ lifts a periodic point of $\hat{f}$ implies that the points $\tilde{f}^q \circ T_0^p(\tilde{z}_0)$, $p \in \mathbb{Z}$, $q \in \mathbb{Z}$, are all on the same side of $\tilde{\lambda}_0$. We will suppose that they are on the left side, meaning that they all belong to $\tilde{U}_1 = \bigcup_{n \ge 0} \tilde{f}^{-n}(L(\tilde{\lambda}_0))$, the covering space of $\hat{U}_1 = \bigcup_{n \ge 0} \hat{f}^{-n}(L(\hat{\lambda}_0))$. The case where they are on the right side can be treated similarly, replacing $\tilde{U}_1$ and $\hat{U}_1$ with $\tilde{U}_2 = \bigcup_{n \ge 0} \tilde{f}^n(R(\tilde{\lambda}_0))$ and $\hat{U}_2 = \bigcup_{n \ge 0} \hat{f}^n(R(\hat{\lambda}))$. Here again, we compactify $\hat{U}_1$ by adding $\hat{J}_2$ and $\hat{S}^1$, the circle of prime ends corresponding to the end on the right of $\hat{\lambda}$. To obtain the universal covering space, we add $\tilde{J}_2$ and $\tilde{S}_1$, the covering space of $\hat{S}_1$, to $\tilde{U}_1$. The map $\tilde{f}_{|\tilde{U}_1}$ extends continuously to the added lines and fixes every point of $\tilde{J}_2$. We denote ρ_1 the rotation numbers of $\widetilde{f}_{|\tilde{S}_1}$. Here again we can define the set $\tilde{\mathcal{D}}_1$ of connected components of $\tilde{U}_1 \cap \left(\bigcup_{\tilde{\lambda} \in \tilde{\mathcal{L}}_1} \tilde{\lambda}\right)$, noting that every element $\tilde{\delta}$ is a line of $\tilde{U}_1$ that is contained in a line $\tilde{\lambda} \in \tilde{\mathcal{L}}_{1,l}$.

If we want to repeat the arguments given in Sect. 3, we will meet a problem. In Sect. 3, the lines of $\tilde{\lambda} \in \tilde{\mathcal{L}}_1$ or $\tilde{\lambda} \in \tilde{\mathcal{L}}_2$ were pairwise disjoint, meaning that $\prec$ induces an order on these sets. This fact was very important in the proof because it was necessary for applying Proposition 2.8. It is no more the case here. Nevertheless, there is at most N lines $T_0^k(\tilde{\lambda}_1)$, $k \in \mathbb{Z}$, that intersect $\tilde{\lambda}_1$ and at most N lines $T^k(\tilde{\lambda}_2)$ that intersect $\tilde{\lambda}_2$. In particular if s is large enough, $\prec$ induces an order on the sets

$$\tilde{\mathcal{L}}'_{1,l} = \{T_0^{sk}(\tilde{\lambda}_1),\, k \in \mathbb{Z}\} \quad \text{and} \quad \tilde{\mathcal{L}}'_{2,l} = \{T_0^{sk}(\tilde{\lambda}_2),\, k \in \mathbb{Z}\}.$$

So we will have to work in the annulus $\tilde{S}/T_0^s$ instead of the annulus $\tilde{S}/T_0$. We define the set $\tilde{\mathcal{D}}'_1 \subset \tilde{\mathcal{D}}_1$ of connected components of $\tilde{U}_1 \cap \left(\bigcup_{k \in \mathbb{Z}} T_0^{sk}(\tilde{\lambda}_1)\right)$.

Like in Sect. 3, it holds that

$$\begin{aligned} \tilde{\lambda}_1 \in \tilde{\mathcal{L}}'_{1,l}, \quad \tilde{\lambda}'_1 \in \tilde{\mathcal{L}}'_{1,l}, \quad k \in \mathbb{Z}\backslash\{0\} &\Rightarrow \tilde{f}^k(\tilde{\lambda}_1) \cap \tilde{\lambda}'_1 = \emptyset, \\ \tilde{\delta}_1 \in \tilde{\mathcal{D}}'_1, \quad \tilde{\delta}'_1 \in \tilde{\mathcal{D}}'_1, \quad k \in \mathbb{Z}\backslash\{0\} &\Rightarrow \tilde{f}^k(\tilde{\delta}_1) \cap \tilde{\delta}'_1 = \emptyset, \\ \tilde{\lambda}_2 \in \tilde{\mathcal{L}}'_{2,r}, \quad \tilde{\lambda}'_2 \in \tilde{\mathcal{L}}'_{2,r}, \quad k \in \mathbb{Z}\backslash\{0\} &\Rightarrow \tilde{f}^k(\tilde{\lambda}_2) \cap \tilde{\lambda}'_2 = \emptyset. \end{aligned}$$

Moreover, as a consequence of Lemma 4.1, there exists $\tilde{\delta}_1 \in \tilde{\mathcal{D}}'_1$ such that $\tilde{f}^{n_0}(\tilde{\delta}_1) \cap \tilde{\lambda}_2 \neq \emptyset$.

There is two cases to study: the case where $\rho_1 \neq 0$ and the case where $\rho_1 = 0$.

Case where $\rho_1 \neq 0$. It is the case where there is a boundary twist condition. The rotation numbers in the new annulus are ρ_1/s and 0. All the arguments appearing in Sect. 3 are still valid. Setting $\rho_- = \min(0, \rho_1)$ and $\rho_+ = \max(0, \rho_1)$ one gets that

$$\rho_-(n_0+q) < sp < \rho_+(n_0+q) \Rightarrow \tilde{f}^{2n_0+q}(\tilde{\delta}_1) \cap T_0^{sp}(\tilde{\lambda}_2) \neq \emptyset.$$

Moreover, if

$$\rho_-(n_0+q) < sp < sp' < \rho_+(n_0+q),$$

then there exists a segment $\tilde{\sigma}_1 \subset \tilde{\delta}_1 \subset \tilde{\lambda}_1$ such that:

- $\tilde{f}^{2n_0+q}(\tilde{\sigma}_1)$ joins $T_0^{sp}(\tilde{\lambda}_2)$ and $T_0^{sp'}(\tilde{\lambda}_2)$;
- the interior of $\tilde{f}^{2n_0+q}(\tilde{\sigma}_1)$ is included in $R(T_0^{sp}(\tilde{\lambda}_2)) \cap R(T_0^{sp'}(\tilde{\lambda}_2))$;
- $\tilde{f}^{2n_0+q}(\tilde{\sigma}_1)$ is included in $\bigcup_{k\leq 0} \tilde{f}^{-k}(L(\tilde{\lambda}_0))$.

Like in Sect. 3, we deduce that for every $s \geq 2$, there exists $m_s \geq 0$ such that for every $m \geq m_s$, there exists $\tilde{\lambda}'_2 \in \tilde{\mathcal{L}}_{2,r}$ such that for every $0 < p < p' \leq s$, there exists a segment $\tilde{\sigma}_1 \subset \tilde{\lambda}_1$ satisfying:

- $\tilde{f}^{m}(\tilde{\sigma}_1)$ joins $T_0^{sp}(\tilde{\lambda}'_2)$ and $T_0^{sp'}(\tilde{\lambda}'_2)$;
- the interior of $\tilde{f}^{m}(\tilde{\sigma}_1)$ is included in $R(T_0^{sp}(\tilde{\lambda}'_2)) \cap R(T_0^{sp'}(\tilde{\lambda}'_2))$;
- $\tilde{f}^{m}(\tilde{\sigma}_1)$ is included in $L(\tilde{\lambda}_0)$;

Fix $T_1 \in G$ that sends $\tilde{\lambda}_1$ onto $\tilde{\lambda}_0$. Like in Sect. 3, we can prove that for every $m \geq m_5$, there exists $T_m \in G$ such that $\tilde{f}^m \circ T_m^{-1}$ has a fixed point, where T_m can be written $T_m = T_0^{sn_m} \circ T_1$. Choose a fixed point $\tilde{z}_m$ of $\tilde{f}^m \circ T_m^{-1}$. It projects onto a fixed point $z_m \in S$ of f^m. Let us prove that the period of z_m tends to $+\infty$ with m. Otherwise, there exists $r \geq 0$ and an increasing sequence $(m_l)_{l\geq 0}$ such that z_{m_l} has period r. So, there exists $S_l \in G$ such that $\tilde{f}^r(\tilde{z}_{m_l}) = S_l(\tilde{z}_{m_l})$. The map $\tilde{f}$ commutes with the covering automorphisms. We deduce on one side that $\tilde{f}^{m_l}(\tilde{z}_{m_l}) = T_{m_l}(\tilde{z}_{m_l})$ and on the other side that $\tilde{f}^{m_l}(\tilde{z}_{m_l}) = S_l^{m_l/r}(\tilde{z}_{m_l})$ and so $T_{m_l} = S_l^{m_l/r}$. Let us explain why it is impossible if l is large enough. Note that

$$S_l^{m_l/r}(\tilde{\lambda}_1) = T_{m_l}(\tilde{\lambda}_1) = T_0^{sn_m} \circ T_1(\tilde{\lambda}_1) = T_0^{sn_m}(\tilde{\lambda}_0) = \tilde{\lambda}_0.$$

This implies that

$$S_l^{m_l/r}(\alpha(\tilde{\lambda}_1)) = \alpha(\tilde{\lambda}_0) \quad \text{and} \quad S_l^{m_l/r}(\omega(\tilde{\lambda}_1)) = \omega(\tilde{\lambda}_0).$$

One can find two disjoint segments $\tilde{\sigma}_\alpha$ and $\tilde{\sigma}_\omega$ of $\partial\tilde{S}$, the first one joining $\alpha(\tilde{\lambda}_1)$ to $\alpha(\tilde{\lambda}_0)$, the second one joining $\omega(\tilde{\lambda}_1)$ to $\omega(\tilde{\lambda}_0)$. This implies that for every $0 \leq k \leq m_l/r$ it holds that $S_l^k(\alpha(\tilde{\lambda}_1)) \in \sigma_\alpha$ and $S_l^k(\omega(\tilde{\lambda}_1)) \in \sigma_\omega$. Let $\tilde{\sigma}$ be a segment of $\tilde{S}$ that joins $\tilde{\lambda}_1$ to $\tilde{\lambda}_2$. The set of lifts of $\hat{\lambda}$ that meet $\tilde{\sigma}$ is finite. Let N' be its cardinal. Suppose that $m_l/r \geq 2N + N'$. There exists $0 < k < m_l/r$ such that $S_l^k(\tilde{\lambda}_1)$ does not meet $\tilde{\lambda}_1 \cup \tilde{\lambda}_0 \cup \tilde{\sigma}$. Nevertheless one of the end of $S_l^k(\tilde{\lambda}_1)$ is in the interior of $\tilde{\sigma}_\alpha$ and the other one in the interior of $\tilde{\sigma}_\omega$. We have a contradiction.

Case where $\rho_1 = 0$. Here there is no boundary twist condition. The twist condition is given by the existence of a periodic point which has a non zero rotation number. The proof is inspired by arguments of Lellouch appearing in his thesis [23].

Lemma 4.2. *There exists an increasing sequence* $(n_p)_{p\geq 0}$ *such that* $\tilde{f}^n(\tilde{\delta}_1)\cap T_0^{sp}(\tilde{\lambda}_2)\neq\emptyset$ *if* $n\geq n_p$.

Proof. Every lift of $\hat{\lambda}$ being a Brouwer line, it is sufficient to prove that there exists n_p such that $\tilde{f}^{n_p}(\tilde{\delta}_1)\cap T_0^{sp}(\tilde{\lambda}_2)\neq\emptyset$. Indeed, if $n>n_p$, then $\tilde{f}^n(\tilde{\delta}_1)\cap L(T_0^{sp}(\tilde{\lambda}_2))\neq\emptyset$ because

$$\tilde{f}^{n-n_p}(\tilde{f}^{n_p}(\tilde{\delta}_1)\cap T_0^{sp}(\tilde{\lambda}_2))=\tilde{f}^n(\tilde{\delta}_1)\cap\tilde{f}^{n-n_p}(T_0^{sp}(\tilde{\lambda}_2))\subset\tilde{f}^n(\tilde{\delta}_1)\cap L(T_0^{sp}(\tilde{\lambda}_2)),$$

and it implies that $\tilde{f}^{n_p}(\tilde{\delta}_1)\cap T_0^{sp}(\tilde{\lambda}_2)\neq\emptyset$. By induction, it is sufficient to prove that there exists $n_1\geq n_0$ such that $\tilde{f}^{n_1}(\tilde{\delta}_1)\cap T_0^{s}(\tilde{\lambda}_2)\neq\emptyset$.

The fact that $\tilde{f}^{n_0}(\tilde{\delta}_1)\cap\tilde{\lambda}_2\neq\emptyset$ implies that there exists a half-line $\tilde{l}_1\subset\tilde{\delta}_1$ and a half-line $\tilde{l}_2\subset\tilde{\lambda}_2$ such that $\tilde{f}^{n_0}(\tilde{l}_1)$ and $\tilde{l}_2$ intersect in a unique point and such that $\tilde{f}^{n_0}(\tilde{l}_1)\cup\tilde{l}_2$ is a line $\tilde{l}$ of $\tilde{U}_1$. For the same reason, there exists a half-line $\tilde{l}'_1\subset\tilde{\delta}_1$ and a half-line $\tilde{l}'_2\subset\tilde{\lambda}_2$ such that $\tilde{l}'_1$ and $\tilde{f}^{-n_0}(\tilde{l}'_2)$ intersect in a unique point and such that $\tilde{l}'_1\cup\tilde{f}^{-n_0}(\tilde{l}'_2)$ is a line $\tilde{l}'$ of $\tilde{U}_1$. We can also make a choice such that $\tilde{l}_1\cap\tilde{l}'_1$ is a half-line of $\tilde{\delta}_1$ and $\tilde{l}_2\cap\tilde{l}'_2$ a half-line of $\tilde{\lambda}_2$

If $m\geq 0$, then $\tilde{f}^m(\tilde{l}_2)$ and $\tilde{f}^{m+n_0}(\tilde{l}_2)$ are contained in $\overline{L(\tilde{\lambda}_2)}$ and so are disjoint from $T_0^s(\tilde{l}'_1)$ and $T_0^s(\tilde{l}'_2)$. We deduce that $\tilde{f}^m(\tilde{l}_2)\cap T_0^s(\tilde{l}')=\emptyset$. Moreover $\tilde{f}^m(\tilde{l}_1)\cap T_0^s(\tilde{l}'_1)=\emptyset$. So, to get Lemma 4.2, it is sufficient to prove that there exists $m>0$ such that $\tilde{f}^m(\tilde{l})\cap T_0^s(\tilde{l}')\neq\emptyset$.

We will argue by contradiction and suppose that $\tilde{f}^m(\tilde{l})\cap T_0^s(\tilde{l}')=\emptyset$ for every $m\geq 0$. We can orient $\tilde{l}$ and $T_0^s(\tilde{l}')$ such that $L(\tilde{l})\subset L(T_0^s(\tilde{l}'))$. The ends of $\tilde{l}$ and $\tilde{l}'$ (on $\tilde{S}_1$ and $\tilde{J}_2$) are the same. The fact that $\rho_1=0$ implies that for every $m>0$, the ends of $\tilde{f}^m(\tilde{l})$, which are the images by $\tilde{f}^m$ of the ends of $\tilde{l}$, stay smaller than the ends of $T_0^s(\tilde{l}')$, which are the images by $\tilde{T}_0^s$ of the ends of $\tilde{l}$. So we have $L(\tilde{f}^m(\tilde{l}))\subset L(T_0^s(\tilde{l}'))$. To get the contradiction, just notice that if m is large enough, then $\tilde{f}^{-m}(\tilde{z}_0)\in L(\tilde{l})$ and $\tilde{f}^m(\tilde{z}_0)\in R(T_0^s(\tilde{l}'))$. But we should have

$$\tilde{f}^m(\tilde{z}_0)=\tilde{f}^{2m}(\tilde{f}^{-m}(\tilde{z}_0))\in L(\tilde{f}^{2m}(\tilde{l}))\subset L(T_0^s(\tilde{l}')).$$

□

Denote $\tilde{\lambda}_1$ the element of $\mathcal{L}_{1,l}$ that contains $\tilde{\delta}_1$. We deduce from Lemma 4.2 that, for every $q\geq 2$, for every $m\geq n_q$ and every $0<p<p'\leq q$, there exists a segment $\tilde{\sigma}_1\subset\tilde{\lambda}_1$ satisfying:

- $\tilde{f}^m(\tilde{\sigma}_1)$ joins $T_0^{sp}(\tilde{\lambda}_2)$ and $T_0^{sp'}(\tilde{\lambda}_2)$;
- the interior of $\tilde{f}^m(\tilde{\sigma}_1)$ is included in $R(T_0^{sp}(\tilde{\lambda}_2))\cap R(T_0^{sp'}(\tilde{\lambda}_2))$;
- $\tilde{f}^m(\tilde{\sigma}_1)$ is included in $\bigcup_{k\leq 0}\tilde{f}^{-k}(L(\tilde{\lambda}_0))$.

Like in the proof of Proposition 3.5, and using the the fact that the $T_0^{sp}(\tilde{\lambda}_2)$ are Brouwer lines (or equivalently that $[f]$ is the identity map) we deduce that, for every $q\geq 2$, there exists $m_q\geq n_q$ such that for every $m\geq m_q$ and every $0<p<p'\leq q$, there exists a segment $\tilde{\sigma}_1\subset\tilde{\lambda}_1$ satisfying:

- $\tilde{f}^m(\tilde{\sigma}_1)$ joins $T_0^{sp}(\tilde{\lambda}_2)$ and $T_0^{sp'}(\tilde{\lambda}_2)$;
- the interior of $\tilde{f}^m(\tilde{\sigma}_1)$ is included in $R(T_0^{sp}(\tilde{\lambda}_2))$ and in $R(T_0^{sp'}(\tilde{\lambda}_2))$;
- $\tilde{f}^m(\tilde{\sigma}_1)$ is included in $L(\tilde{\lambda}_0)$.

Finally, like in the proof of Proposition 3.6, we show that if T_1 is the unique covering automorphism such that

$$T_1(\tilde{\lambda}_1) = \tilde{\lambda}_0 \quad \text{and} \quad T_0^{3s}(\tilde{\lambda}_2) \prec T_1(T_0^s(\tilde{\lambda}_1)) \prec T_0^{4s}(\tilde{\lambda}_2),$$

then, for every $m \geq m_5$, the map $\tilde{f}^m \circ T_1^{-1}$ has a fixed point.

To conclude, choose a fixed point $\tilde{z}_m$ of $\tilde{f}^m \circ T_1^{-1}$. It projects onto a fixed point $z_m \in S$ of f^m. Let us prove that the period of z_m tends to $+\infty$ with m. Otherwise, there exist $r \geq 0$ and an increasing sequence $(m_l)_{l\geq 0}$ such that z_{m_l} has period r. So, there exists $S_l \in G$ such that $\tilde{f}^r(\tilde{z}_{m_l}) = S_l(\tilde{z}_{m_l})$. One deduces that $\tilde{f}^{m_l}(\tilde{z}_{m_l}) = S_l^{m_l/r}(\tilde{z}_{m_l})$ and so $T_1 = S_l^{m_l/r}$. In particular S_l belongs to the centralizer of T. But it is well known that the centralizer of T_1 is a cyclic group. We have got a contradiction. □

5. The case of the torus

The goal of this section is to prove Theorem 1.2. The arguments that follow are the ones appearing in [2] but we need to verify that, up to slight modifications, they are still valid when the area-preserving condition is replaced with the non wandering condition.

Let us consider $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$. There are three possibilities:

- M is hyperbolic, meaning that its eigenvalues have a modulus different from 1;
- M is conjugate in $\mathrm{SL}(2,\mathbb{Z})$ to $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ or to $\begin{pmatrix} -1 & k \\ 0 & -1 \end{pmatrix}$, where $k \in \mathbb{Z}\backslash\{0\}$;
- M has finite order (and in that case its order is 1, 2, 3, 4 or 6).

The matrix M induces an orientation preserving automorphism $[M]$ of $\mathbb{T}^2$ by the formula $[M](x,y) = (ax+by, cx+dy)$. We will denote $\mathrm{Aut}(\mathbb{T}^2)$ the group of such automorphisms. Every orientation preserving homeomorphism f of $\mathbb{T}^2$ is isotopic to a unique automorphism, that will be denoted $[f]$, as its associated matrix.

Setting $D = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, we have the following classification for an orientation homeomorphism f of $\mathbb{T}^2$:

- $[f]$ is hyperbolic;
- there exists $q \in \{1,2\}$ and $k \in \mathbb{Z}\backslash\{0\}$ such that f^q is conjugate to a homeomorphism isotopic to $[D]^k$;
- there exists $q \in \{1,2,3,4,6\}$ such that f^q is isotopic to the identity.

Let us recall now the definition of the rotation set of a homeomorphism of $\mathbb{T}^2$ isotopic to the identity (see [25] or [26]). For every homeomorphism f of $\mathbb{T}^2$, denote $\mathcal{M}_f$ the set of Borel probability measures invariant by f.

Suppose that f is isotopic to the identity. Every lift $\tilde{f}$ of f to $\mathbb{R}^2$ commutes with the integer translations $\tilde{z} \mapsto z + k$, $k \in \mathbb{Z}^2$. So, the map $\tilde{f} - \mathrm{Id}$ lifts a continuous function $\psi_{\tilde{f}} : \mathbb{T}^2 \to \mathbb{R}^2$. One can define the *rotation vector* $\mathrm{rot}_{\tilde{f}}(\mu) = \int_{\mathbb{T}^2} \psi_{\tilde{f}}\, d\mu \in \mathbb{R}^2$ of $\mu \in \mathcal{M}_f$, that measures the mean displacement of $\tilde{f}$. The *rotation set* $\mathrm{rot}(\tilde{f}) = \{\mathrm{rot}_{\tilde{f}}(\mu) \,|\, \mu \in \mathcal{M}_f\}$ is a non empty compact convex subset of $\mathbb{R}^2$. Of course, it depends on the lift $\tilde{f}$ but if $\tilde{f}' = \tilde{f} + k$, $k \in \mathbb{Z}^2$, is another lift, it holds that $\mathrm{rot}(\tilde{f}') = \mathrm{rot}(\tilde{f}) + k$ because $\psi_{\tilde{f}'} = \psi_{\tilde{f}} + k$.

The following properties are very classical (the two first ones are due to Franks [12], [13], the last one is an easy consequence of the characterization of the ergodic measures as extremal points of $\mathcal{M}_f$).

(1) If p/q belongs to the interior of $\mathrm{rot}(\tilde{f})$, where $p \in \mathbb{Z}^2$ and $q \in \mathbb{N}\backslash\{0\}$, then there exists $\tilde{z} \in \mathbb{R}^2$ such that $\tilde{f}^q(\tilde{z}) = \tilde{z} + p$.
(2) If $p \in \mathbb{Z}^2$ and $q \in \mathbb{N}\backslash\{0\}$ are such that $p/q = \mathrm{rot}_{\tilde{f}}(\mu)$, where $\mu \in \mathcal{M}_f$ is ergodic, then there exists $\tilde{z} \in \mathbb{R}^2$ such that $\tilde{f}^q(\tilde{z}) = \tilde{z} + p$.
(3) Every extremal point of $\mathrm{rot}(\tilde{f})$ is the rotation vector of an ergodic measure.

Let us recall now the definition of the vertical rotation set of a homeomorphism of $\mathbb{T}^2$ isotopic to a Dehn twist (see [1] or [8]). Suppose that f is isotopic to $[D]^k$, where $k \neq 0$ and that $\hat{f}$ is a lift of f to $\mathbb{T} \times \mathbb{R}$. It commutes with the vertical translation $V : \hat{z} \mapsto \hat{z} + (0, 1)$. So, the map $p_2 \circ \hat{f} - p_2$ lifts a continuous function $\psi_{\hat{f}} : \mathbb{T}^2 \to \mathbb{R}$. One can define the *vertical rotation number* $\mathrm{vrot}_{\hat{f}}(\mu) = \int_{\mathbb{T}^2} \psi_{\hat{f}}\, d\mu \in \mathbb{R}$ of a measure $\mu \in \mathcal{M}_f$ and the *vertical rotation set* $\mathrm{vrot}(\hat{f}) = \{\rho(\mu) \,|\, \mu \in \mathcal{M}_f\}$, which is a non empty segment of $\mathbb{R}$. Here again, it depends on the lift $\hat{f}$ but if $\hat{f}' = V^k \circ \hat{f}$, $k \in \mathbb{Z}$, is another lift, we have $\mathrm{vrot}(\hat{f}') = \mathrm{vrot}(\hat{f}) + k$.

We will need the following properties, the first one being proved in [1] and [8], the second one in [3]:

(1) if p/q belongs to the interior of $\mathrm{vrot}(\hat{f})$, where $p \in \mathbb{Z}$ and $q \in \mathbb{N}\backslash\{0\}$, there exists $\hat{z} \in \mathbb{T} \times \mathbb{R}$ such that $\hat{f}^q(\hat{z}) = V^p(\hat{z})$;
(2) if $\mathrm{vrot}(\hat{f}) = \{p/q\}$, where $p \in \mathbb{Z}$ and $q \in \mathbb{N}\backslash\{0\}$, there exists a compact connected essential set $\hat{K} \subset \mathbb{T} \times \mathbb{R}$ invariant by $\hat{f}^q \circ V^{-p}$.

Let f be an orientation preserving homeomorphism of $\mathbb{T}^2$. It is well known that if $[f]$ is hyperbolic, then f has periodic points of period arbitrarily large. So, Theorem 1.2 can be deduced from the two following results:

Proposition 5.1. *Let f a non wandering homeomorphism of $\mathbb{T}^2$ isotopic to $[D]^k$, $k \neq 0$. Then:*

- *either f has periodic points of period arbitrarily large;*
- *or f has no periodic orbit and there exists $\delta \in \mathbb{T}\backslash\mathbb{Q}/\mathbb{Z}$ such that for every lift $\hat{f}$ of f to $\mathbb{T} \times \mathbb{R}$, there exists $\hat{\delta} \in \mathbb{R}$ such that $\hat{\delta} + \mathbb{Z} = \delta$ and $\mathrm{vrot}(\hat{f}) = \{\hat{\delta}\}$.*

Proposition 5.2. *Let f a non wandering homeomorphism of $\mathbb{T}^2$ isotopic to the identity. Exactly one of the following assertions holds:*

(1) *f has periodic points of period arbitrarily large;*
(2) *If $\tilde{f}$ is a lift of f to $\mathbb{R}^2$, then $\mathrm{rot}(\tilde{f})$ is a point or a segment that does not meet $\mathbb{Q}^2/\mathbb{Z}^2$. In this case f has no periodic point.*
(3) *There exists an integer $q \geq 1$ such that:*
    - *the periodic points of f^q are fixed;*
    - *the fixed point set of f^q is non empty and f^q is isotopic to the identity relative to it;*
    - *the rotation set of the lift of f^q that has fixed points is reduced to 0 or is a segment with irrational slope that has zero as an end point.*

Proof of Proposition 5.1. Fix a lift $\hat{f}$ of f to $\mathbb{T}\times\mathbb{R}$. If $\mathrm{vrot}(\hat{f})$ is not reduced to a point, then for every rational number $p/q \in \mathrm{int}(\mathrm{vrot}(\hat{f}))$ there exists $\hat{z} \in \mathbb{T}\times\mathbb{R}$ such that $\hat{f}^q(\hat{z}) = V^p(\hat{z})$. If p and q are chosen relatively prime, then $\hat{z}$ projects onto a periodic point of f of period q. Consequently, f has periodic points of period arbitrarily large.

Suppose now that $\mathrm{vrot}(\hat{f})$ is reduced to a rational number p/q. Replacing f with f^q and $\hat{f}$ with $V^{-p}\circ\hat{f}^q$, one can suppose that $\mathrm{vrot}(\hat{f}) = \{0\}$. There exists a compact connected essential set $\hat{K} \subset \mathbb{T}\times\mathbb{R}$ that is invariant by $\hat{f}$. Let μ be a Borel probability measure supported on $\hat{K}$ and invariant by $\hat{f}$. For every $m \in \mathbb{Z}$, the measure $V^m_*(\mu)$ is supported on $V^m(\hat{K})$ and invariant by $\hat{f}$. Fix a lift $\tilde{f}$ of $\hat{f}$ to $\mathbb{R}^2$. For every $m \in \mathbb{Z}$, one has $\mathrm{rot}_{\tilde{f}}(V^m_*(\mu)) = \mathrm{rot}_{\tilde{f}}(\mu)+mk$. So, by Theorem 2.4 it holds that

- either, for every rational number $p/q \in \mathbb{R}$, there exists a periodic point $\hat{z}$ of $\hat{f}$ of period q and rotation number p/q for $\tilde{f}$;
- or there exists an essential simple loop $\hat{\lambda} \in \mathbb{T}\times\mathbb{R}$ such that $\hat{f}(\hat{\lambda})\cap\hat{\lambda} = \emptyset$.

Here again, in the first situation, if p and q are chosen relatively prime, $\hat{z}$ projects onto a periodic point of f of period q and so f has periodic points of period arbitrarily large. Let us prove now that the second situation never holds. Suppose that there exists an essential simple loop $\hat{\lambda} \in \mathbb{T}\times\mathbb{R}$ such that $\hat{f}(\hat{\lambda})\cap\hat{\lambda} = \emptyset$. Then one can find a relatively compact wandering disk $\hat{U}$. The fact that every $V^m(\hat{K})$, $m \in \mathbb{Z}$, is compact, essential and $\hat{f}$-invariant implies that $\bigcup_{k\in\mathbb{Z}} \hat{f}^k(\hat{U})$ is relatively compact. This contradicts Proposition 2.3. □

Proof of Proposition 5.2. Fix a lift $\tilde{f}$ of f to $\mathbb{R}^2$. If $\mathrm{rot}(\tilde{f})$ has non empty interior, then f has periodic points of period arbitrarily large. Indeed, if $(p_1/q, p_2/q)$ belongs to the interior of $\mathrm{rot}(\tilde{f})$, then there exists $\tilde{z} \in \mathbb{R}^2$ such that $\tilde{f}^q(\tilde{z}) = \tilde{z}+(p_1, p_2)$. Moreover, if p_1, p_2 and q are chosen with no common divisor, then $\tilde{z}$ projects onto a periodic point of f of period q.

Suppose now that $\mathrm{rot}(\tilde{f})$ is a point or a segment that does not meet $\mathbb{Q}^2/\mathbb{Z}^2$. Then f has no periodic point and (2) holds.

Suppose now that $\mathrm{rot}(\tilde{f})$ meets $\mathbb{Q}^2/\mathbb{Z}^2$ in a unique point p/q. Either $\mathrm{rot}(\tilde{f})$ is reduced to p/q or is a segment with irrational slope. It has been proven in [21] that p/q is an end point of $\mathrm{rot}(\tilde{f})$ in this last case. In particular in both cases, $\tilde{f}^q - p$ has at least one fixed point, because p/q is the rotation vector of an ergodic measure. Note that every periodic point of f^q is lifted to a periodic point of $\tilde{f}^q - p$, meaning it is contractible. Using [19], one deduces that:

- either f has periodic points of arbitrarily large period;
- or the periodic points of f^q are all fixed and lifted to fixed points of $\tilde{f}^q - p$, moreover f^q is isotopic to the identity relative to its fixed point set.

It remains to study the case where $\mathrm{rot}(\tilde{f})$ is a non trivial segment with rational slope that intersects $\mathbb{Q}^2/\mathbb{Z}^2$. Replacing f with f^q and $\tilde{f}$ with $\tilde{f}^q - p$, where $p/q \in \mathrm{rot}(\tilde{f})$, one can suppose that $0 \in \mathrm{rot}(\tilde{f})$. The linear line containing $\mathrm{rot}(\tilde{f})$ is generated by $p' \in \mathbb{Z}^2 \backslash \{0\}$ and invariant by the translation $T : z \mapsto z + p'$. Let $\hat{f}$ be the homeomorphism of the annulus $A_{p'} = \mathbb{R}^2/T$ lifted by $\tilde{f}$. A result of Dávalos [7] tells us that the orbits of $\hat{f}$ are uniformly bounded, or equivalently that there exists a compact connected essential set $\hat{K} \subset \mathbb{A}_{p'}$ invariant by $\hat{f}$. The rotation set of $\tilde{f}$ being non trivial, one can find two compactly supported ergodic measures of $\hat{f}$ with different rotation numbers (for $\tilde{f}$). Like in the proof of Proposition 5.1, we can prove that for every essential simple loop $\hat{\lambda} \in \mathbb{A}_{p'}$ it holds that $\hat{f}(\hat{\lambda}) \cap \hat{\lambda} \neq \emptyset$ and then that f has periodic point of arbitrarily large periods. □

References

[1] Addas Zanata, S.: On properties of the vertical rotation interval for twist mappings. II. Qual. Theory Dyn. Syst. **4**(2), 125–137 (2003)

[2] Addas-Zanata, S., Tal, F.A.: On periodic points of area preserving torus homeomorphisms. Far East J. Dyn. Syst. **9**(3), 371–378 (2007)

[3] Addas-Zanata, S., Garcia, B., Tal, F.A.: Dynamics of homeomorphisms of the torus homotopic to Dehn twists. Ergod. Theory Dyn. Syst. **34**(2), 409–422 (2014)

[4] Anosov, D.V., Katok, A.B.: New examples in smooth ergodic theory. Ergodic diffeomorphisms. Trans. Moscow Math. Soc. **23**, 1–35 (1970)

[5] Birkhoff, G.-D.: An extension of Poincaré's last geometric theorem. Acta Math. **47**, 297–311 (1926)

[6] Casson, A., Bleiler, S.: Automorphisms of surfaces after Nielsen and Thurston. London Mathematical Society Student Texts, vol. 9. Cambridge University Press, Cambridge (1988). iv+105 pp. ISBN: 0-521-34203-1

[7] Dávalos, P.: On annular maps of the torus and sublinear diffusion. J. Inst. Math. Jussieu. **17**(4), 913–978 (2018)

[8] Doeff, E.: Rotation measures for homeomorphisms of the torus homotopic to a Dehn twist. Ergod. Theory Dyn. Syst. **17**(3), 575–591 (1997)

[9] Fathi, A., Laudenbach, F., Poénaru, V.: Travaux de Thurston sur les surfaces. Astérisque, 66–67. Société Mathématique de France, Paris (1979)

[10] Fayad, B., Saprykina, M.: Weak mixing disc and annulus diffeomorphisms with arbitrary Liouville rotation number on the boundary. Ann. Sci. École Norm. Sup. (4) **38**(3), 339–364 (2005)

[11] Franks, J.: Generalizations of the Poincaré–Birkhoff theorem. Ann. Math. (2) **128**(1), 139–151 (1988)

[12] Franks, J.: Recurrence and fixed points of surface homeomorphisms. Ergod. Theory Dyn. Syst. **8***, 9–107 (1988)

[13] Franks, J.: Realizing rotation vectors for torus homeomorphisms. Trans. Am. Math. Soc. **311**(1), 107–115 (1989)

[14] Franks, J.: Area preserving homeomorphisms of open surfaces of genus zero. N. Y. J. Math. **2**, 1–19 (1996)

[15] Franks, J., Handel, M.: Periodic points of Hamiltonian surface diffeomorphisms. Geom. Topol. **7**, 713–756 (2003)

[16] Handel, M.: Global shadowing of pseudo-Anosov homeomorphisms. Ergod. Theory Dyn. Syst. **5**(3), 373–377 (1985)

[17] Homma, T.: An extension of the Jordan curve theorem. Yokohama Math. J. **1**, 125–129 (1953)

[18] Le Calvez, P.: Une version feuilletée équivariante du théorème de translation de Brouwer. Publ. Math. Inst. Hautes Études Sci. **102**, 1–98 (2005)

[19] Le Calvez, P.: Periodic orbits of Hamiltonian homeomorphisms of surfaces. Duke Math. J. **133**(1), 125–184 (2006)

[20] Le Calvez, P., Sambarino, M.: Homoclinic orbits for area preserving diffeomorphisms of surfaces. Ergod. Theory Dyn. Syst. **42**(3), 1122–1165 (2022)

[21] Le Calvez, P., Tal, F.A.: Forcing theory for transverse trajectories of surface homeomorphisms. Invent. Math. **212**(2), 619–729 (2018)

[22] Le Calvez, P., Tal, F.A.: Topological horseshoe for surface homeomorphisms. arXiv:1803.04557 (2016)

[23] Lellouch, G.: Sur les ensembles de rotation des homéomorphismes de surface en genre ≥ 2. PhD thesis, Sorbonne Université (2019)

[24] Mather, J.: Topological proofs of some purely topological consequences of Caratheodory's Theory of prime ends. In: Rassias, T.H.M., Rassias, G.M. (eds.) Selected Studies, vol. 168, pp. 225–255. North Holland, Amsterdam (1982)

[25] Misiurewicz, M., Ziemian, K.: Rotation sets for maps of tori. J. Lond. Math. Soc. **40**, 490–506 (1989)

[26] Schwartzman, S.: Asymptotic cycles. Ann. Math. **68**, 270–284 (1957)

[27] Thurston, W.: On the geometry and dynamics of diffeomorphisms of surfaces. Bull. Am. Math. Soc. (N.S.) **19**(2), 417–431 (1988)

[28] Wang, J., Zhang, Z.: The rigidity of pseudo-rotations on the two-torus and a question of Norton–Sullivan. Geom. Funct. Anal. **28**(5), 1487–1516 (2018)

Patrice Le Calvez
Institut de Mathématiques de Jussieu-Paris Rive Gauche, IMJ-PRG
Sorbonne Université, Université Paris Cité, CNRS
4, place Jussieu
75005 Paris
France

and

Institut Universitaire de France
Villeurbanne, 103, boulevard Saint-Michel,
75005 Paris
France
e-mail: patrice.le-calvez@imj-prg.fr

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:36
https://doi.org/10.1007/s11784-022-00955-8
Published online April 12, 2022

Journal of Fixed Point Theory and Applications

The Anosov–Katok method and pseudo-rotations in symplectic dynamics

Frédéric Le Roux and Sobhan Seyfaddini

Dedicated to Claude Viterbo. Joyeux anniversaire, Claude !

Abstract. We prove that toric symplectic manifolds admit Hamiltonian pseudo-rotations with a finite, and in a sense minimal, number of ergodic measures. The set of ergodic measures of these pseudo-rotations consists of the measure induced by the symplectic volume form and the Dirac measures supported at the fixed points of the torus action. Our construction relies on the conjugation method of Anosov and Katok.

Mathematics Subject Classification. 37C05, 53D05.

Keywords. Anosov–Katok method, pseudo rotations, unique ergodicity, symplectic dynamics.

Contents

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

1. Introduction

The main goal of this paper is to exhibit examples of Hamiltonian diffeomorphisms on symplectic manifolds of dimension greater than two, which, on the one hand, have a finite number of periodic points, and on the other hand, have interesting and complicated dynamics. We will refer to Hamiltonian diffeomorphisms with a finite number of periodic points as Hamiltonian **pseudo-rotations**.[1] Such diffeomorphisms have been of great interest in dynamical systems and symplectic topology; see, for example, [1,4–10,14,15,19,20,27,32,33].

The only closed surface admitting Hamiltonian pseudo-rotations is the sphere with the simplest examples being irrational rotations.[2] More interesting examples, with only three ergodic[3] measures, were constructed by Fayad and Katok [15], using the so-called conjugation method of Anosov–Katok [1]. Note that three is the minimal possible number of ergodic measures for a Hamiltonian diffeomorphism of the sphere because any such diffeomorphism has at least two fixed points and so preserves, in addition to the area, the Dirac delta measures supported at the fixed points.

Consider a closed symplectic manifold (M, ω) carrying a Hamiltonian circle action S such that the fixed point set of the action $\mathrm{Fix}(S)$ is finite and the action is locally free on $M \backslash \mathrm{Fix}(S)$. Any such manifold, like the sphere and the projective spaces $\mathbb{C}P^n$, admits pseudo-rotations obtained from the irrational elements of the circle. However, these example have very simple dynamics. As we explain in Sect. 2.3, in this context, a straight-forward adaptation of the Anosov–Katok method yields pseudo-rotations which are

[1] There exist several working definitions of Hamiltonian pseudo-rotations in the literature; see [9, Def. 1.1] and the discussion therein. The examples we construct in this article do satisfy the requirements in all definitions known to us.

[2] Closed surfaces of positive genus fall under the class of symplectic manifolds which satisfy the Conley conjecture [18,24,26] asserting that any Hamiltonian diffeomorphism has infinitely many simple periodic points. Clearly, such symplectic manifolds do not admit pseudo-rotations.

[3] Recall that a Borel probability measure μ is said to be ergodic for $f : M \to M$ if it is preserved by f and, furthermore, satisfies the following condition: for any Borel subset $A \subset M$ such that $f^{-1}(A) \subset A$ either $\mu(A) = 0$ or $\mu(A) = 1$.

transitive, i.e. have dense orbits. The presence of a circle action as above is an indispensable ingredient of the Anosov–Katok method and, as far as we know, there are no known examples of pseudo-rotations on manifolds not carrying circle actions.

Producing pseudo-rotations with dynamics more complicated than transitivity is more involved and is the main goal of this article. Our main result, whose proof relies on the Anosov–Katok method, guarantees the existence of pseudo-rotations with a finite, and in a sense optimal, number of ergodic measures on **toric** symplectic manifolds. Recall that a $2n$-dimensional symplectic manifold is called toric if it admits an effective Hamiltonian action of the torus $\mathbb{T}^n$; we review this definition and relevant facts in Sect. 3.1.

Theorem 1. *Let (M, ω) be a closed toric symplectic manifold, and denote by ℓ the number of fixed points of the corresponding torus action. Then (M, ω) admits a Hamiltonian pseudo-rotation f with exactly $\ell + 1$ ergodic measures. The set of ergodic measures of f consists of the measure induced by the symplectic volume form ω^n and the ℓ Dirac measures supported at the fixed points of the torus action.*

Applying the above theorem to $(\mathbb{C}P^n, \omega_{FS})$, which admits a toric action with n fixed points, yields pseudo-rotations with $n + 1$ ergodic measures.

The number $\ell + 1$ appears to be optimal because the rank of the singular homology of M is ℓ; see, for example, [11, Theorem 3.3.1]. Hence, by the Arnold conjecture any non-degenerate Hamiltonian diffeomorphism of (M, ω) has at least ℓ fixed points and thus must have at least $\ell + 1$ ergodic measures. Non-degeneracy of the pseudo-rotations we produce can be guaranteed by appropriately applying the Anosov–Katok method, see Remark 2; moreover, some authors incorporate non-degeneracy into the definition of pseudo-rotations, see [9, Def. 1.1].

Denote by B_r the standard closed Euclidean ball of radius r, and let $\{B_i\}, \{B'_i\}, i = 1, \ldots, k'$, be two collections of pair-wise disjoint subsets of (M, ω) all of which are images of B_r under symplectic embeddings. In the case of the 2-sphere, a key component of the argument is the fact that one can find a symplectomorphism ψ such that $\psi(B_i) = B'_i$ for $i = 1, \ldots, k$. However, the existence of such ψ cannot be guaranteed on higher dimensional symplectic manifolds even if $k = 1$. Indeed, under certain assumptions, there exist obstructions to the existence of such ψ which are generally referred to as symplectic camel obstructions; see [28] for further details. The camel-type obstructions do disappear for sufficiently small values of r. Hence, by Katok's Basic Lemma [25], one can try to surmount these obstacles by breaking B_i, B'_i into balls of sufficiently small radius to obtain a symplectomorphism ψ such that $\psi(B_i)\Delta B'_i$ is of nearly, but not exactly, zero measure; here Δ stands for the symmetric difference of sets. It appears to us that this approach could potentially yield ergodic pseudo-rotations, but we could not utilize it to construct pseudo-rotations with a finite number of ergodic measures.[4] Instead, we overcome the camel-type difficulties by taking advantage of the

[4]As explained in [15], to get ergodicity one needs to control *almost all* orbits; but to prove Theorem 1 one needs to control *all* orbits.

existence of a "nearly global" system of action-angle coordinates on toric symplectic manifolds; see Section 3.1. It is not clear to us if the assumption of (M, ω) being toric is necessary for the existence of pseudo-rotations with few ergodic measures.

Finally, we should mention that several authors have used the Anosov–Katok method to produce interesting examples, other than Hamiltonian pseudo-rotations, in higher dimensional symplectic manifolds. In [25], Katok constructs an autonomous Hamiltonian with numerous properties including ergodicity of the restriction of the Hamiltonian flow to its energy levels. In [30], Polterovich uses Katok's lemma to prove, among other results, that every closed symplectic manifold admits contractible Hamiltonian loops which are strictly ergodic; see Theorem 1.2.A therein. In [23], Hernàndez–Corbato and Presas, construct examples of (non-Hamiltonian) minimal symplectomorphisms and strictly ergodic contactomorphisms.

Organization of the paper

In Sects. 2.1 and 2.2, we present the general scheme of the Anosov–Katok method. In Sect. 2.3, which may be viewed as a warm-up for the proof of the main result, we explain how to construct transitive pseudo-rotations. In Sect. 2.4, we state Proposition 4, which is the key technical proposition of the paper, and we use it to prove Theorem 1.

The rest of the paper is dedicated to the proof of Proposition 4. In Sect. 3, we review the relevant aspects of toric symplectic geometry and prove preliminary symplectic lemmas which will be used in the following section. Section 4 is the technical heart of the paper and contains the proof of Proposition 4.

2. The conjugation method of Anosov and Katok

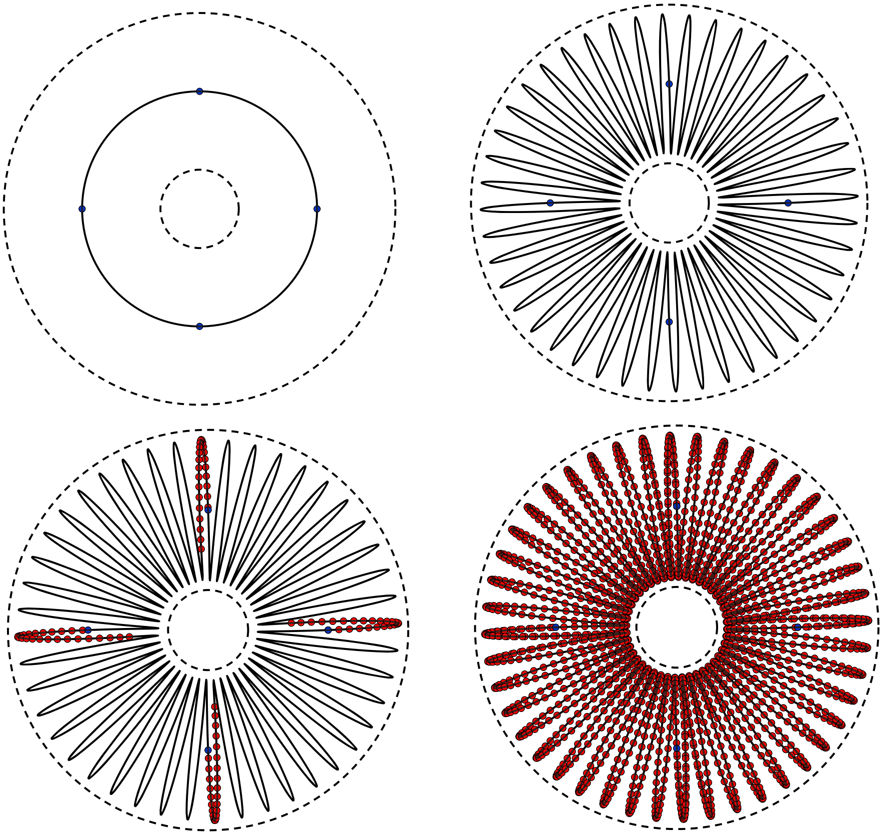
Let (M, ω) be a closed symplectic manifold which admits a smooth Hamiltonian action of the circle, denoted by $(S_\alpha)_{\alpha \in \mathbb{S}^1}$. We identify the circle $\mathbb{S}^1$ with $\mathbb{R}/\mathbb{Z}$. We denote the fixed point set of the action by $\mathrm{Fix}(S)$ and we suppose that the action is locally free outside of $\mathrm{Fix}(S)$. This means that there exists a neighborhood V of $0 \in \mathbb{S}^1$ such that for every non-zero $\alpha \in V$ the homeomorphism S_α has no fixed points in $M \backslash \mathrm{Fix}(S)$.

2.1. General scheme

We outline here the general scheme of constructing a pseudo-rotation, say f, via the conjugation method of Anosov–Katok (see also Fig. 1). The pseudo-rotation f will be obtained as the C^∞ limit of a sequence of Hamiltonian diffeomorphisms (f_n) which are of the form

$$f_n = H_n S_{\alpha_{n+1}} H_n^{-1},$$

where H_n is a symplectic diffeomorphism of M and $\alpha_n = \frac{p_n}{q_n} \in \mathbb{Q}/\mathbb{Z}$ (all fractions are implicitly supposed to be irreducible). We can start with any choice for H_0 and α_1, e.g. $H_0 = \mathrm{Id}$ and $\alpha_1 = \frac{1}{4}$. The Hamiltonian diffeomorphism f_n is obtained from f_{n-1} in the following manner: We construct a



1. One orbit of the circle action $(S_\alpha)_{\alpha\in\mathbb{S}^1}$ with one discrete orbit for S_{α_1};
2. The same orbit for $(h_1 S_\alpha h_1^{-1})_{\alpha\in\mathbb{S}^1}$, the discrete orbit is unchanged;
3. First 100 iterates for $h_1 S_{\alpha_2} h_1^{-1}$, with α_2 very close to α_1;
4. The discrete orbit is ε_1-dense for some small ε_1.

Figure 1. The first two steps in Anosov–Katok method to construct a transitive diffeomorphism (here $\alpha_1 = \frac{1}{4}$), see also Sect. 2.3

symplectic diffeomorphism h_n which coincides with the identity near $\mathrm{Fix}(S)$ and commutes with S_{α_n}; let $H_n := H_{n-1} \circ h_n$. Observe that

$$H_n S_{\alpha_n} H_n^{-1} = H_{n-1} S_{\alpha_n} H_{n-1}^{-1} = f_{n-1}.$$

Hence, by choosing α_{n+1} to be sufficiently close to α_n, we can ensure that f_{n+1}, defined by the above formula, is as close as desired to f_n (in the C^∞ topology). On the one hand, this ensures the C^∞ convergence of the sequence (f_n) and on the other hand it will allow us to prove that the limit map f inherits the approximate dynamical and ergodic properties of the f_n's. Note that if the convergence is fast enough, the sequence (α_n) will converge to some irrational number α, and the fixed point set of f will coincide with the fixed point set of the initial circle action (see Sect. 2.2). We should point out that the map f, being a C^∞ limit of Hamiltonian diffeomorphisms, is indeed a Hamiltonian diffeomorphism of M; this non-trivial fact is a consequence of the C^∞ Flux conjecture which was settled by Ono [29].[5]

To ensure that the Hamiltonian diffeomorphism h_n commutes with S_{α_n}, we will carry out the construction of h_n on the quotient of $M\backslash\mathrm{Fix}(S)$ by the action of the rotation $S_{\frac{1}{q_n}}$. To make sure that this quotient is a smooth manifold we must pick q_n such that the action of $S_{\frac{1}{q_n}}$ on $M\backslash\mathrm{Fix}(S)$ is free which is not automatically guaranteed because the circle action is assumed to be only locally free on $M\backslash\mathrm{Fix}(S)$. To overcome this technical difficulty, the numbers α_n will be picked as follows. Consider the subgroup of $\mathbb{S}^1$ generated by the union of the stabilizers of all points in $M\backslash\mathrm{Fix}(S)$. Since the action is locally free on $M\backslash\mathrm{Fix}(S)$, this subgroup is generated by some $\frac{1}{q_0}$. We denote

$$\mathcal{Q} = \{q \in \mathbb{N} : q \text{ is relatively prime to } q_0\}. \tag{1}$$

This set is closed under multiplication. The set $\mathcal{Q}$ is significant to our construction because of the following property: for every rational number $\frac{1}{q}$ with q in $\mathcal{Q}$, the action of $\mathbb{Z}/q\mathbb{Z}$ generated by $S_{\frac{1}{q}}$ on $M\backslash\mathrm{Fix}(S)$ is free. Hence, the quotient

$$\frac{M\backslash\mathrm{Fix}(S)}{S_{\frac{1}{q}}}$$

is a smooth manifold. Moreover, it naturally inherits the symplectic structure of M. Lastly, note that the set of rationals $\frac{p}{q}$ with $q \in \mathcal{Q}$ is dense in $\mathbb{S}^1$.

Remark 2. We do not know if pseudo-rotations are in general necessarily non-degenerate. However, as we now explain, it is possible to ensure that the pseudo-rotation f from Theorem 1 is non-degenerate. Recall that being non-degenerate means that the derivative of f at any of its fixed points does not have 1 as an eigenvalue.

The sequence of rational numbers α_n has a limit which we denote by α; as noted before, fast convergence of the sequence guarantees that α is irrational. The derivative of f at any of its fixed points has the same eigenvalues as the derivative of S_α at the same fixed point. This is because the diffeomorphisms H_n coincide with the identity in a neighborhood of $\mathrm{Fix}(S)$. Hence, the diffeomorphism f is non-degenerate if and only if S_α is non-degenerate.

In the case of the circle action used in the construction of the pseudo-rotations of Theorem 1, α being irrational guarantees that S_α is indeed non-degenerate. This fact, which can be verified via the equivariant version of the Darboux theorem (see [11, Theorem 3.1.2]), is a consequence of the conditions imposed by Eq. (3).

Let us add that pseudo-rotations of the 2-sphere are necessarily non-degenerate (this is a well-known consequence of the rotation vectors theory,

[5]It can easily be seen that f is isotopic to the identity in $\mathrm{Symp}(M,\omega)$. Hence, f is automatically Hamiltonian if $H^1(M)$ is trivial, which is the case for all toric symplectic manifolds.

see [16,17]). It would be interesting to know if pseudo-rotations are always non-degenerate in higher dimensions.

2.2. Why is f a pseudo-rotation?

Let f be a Hamiltonian diffeomorphism obtained via the conjugation method as described in the previous section. As mentioned earlier, our goal is to construct f so that it is a pseudo-rotation (i.e. no periodic points outside of $\mathrm{Fix}(S)$) and it displays complicated dynamical behaviour (transitivity, ergodicity, unique ergodicity). Being a pseudo-rotation is automatic for the f satisfying the conclusion of Theorem 1 because every periodic orbit supports an ergodic measure. In this section, we will explain how the conjugation method can easily provide a transitive pseudo-rotation; this can be considered as a warm-up for the more technical proof of Theorem 1. As we will now explain, if the convergence of the sequence (f_n) is fast enough, which can be arranged by choosing the α_n's appropriately, then the limit map f will also be a pseudo-rotation. The general idea, which will be useful in many places, e.g. to get transitivity, is that an open property which holds at some step n is automatically transmitted to the limit map "by fast convergence".

Here are the details. Prior to the construction, we fix an increasing sequence (K_n) of compact subsets of $M\backslash\mathrm{Fix}(S)$ whose union is $M\backslash\mathrm{Fix}(S)$. At the beginning of step n of the construction, we define the set $U(K_n, q_n) \subset \mathrm{Ham}(M,\omega)$ consisting of all Hamiltonian diffeomorphisms which have no periodic points of period less than q_n inside K_n. (Recall that q_n is determined in step $n-1$.) This set contains f_{n-1} and is open in the C^0-topology. We choose a C^∞-neighborhood V_n of f_{n-1} whose C^∞-closure is contained in $U(K_n, q_n)$. At each subsequent step $i \geqslant n$ we simply choose the number α_i close enough to α_{i-1} so that the Hamiltonian diffeomorphism f_i belongs to V_n. This implies that the limit map f belongs to $U(K_n, q_n)$, that is, f will have no periodic points of period less than q_n inside K_n. The map f will have this property for every n, and the numbers q_n will be picked such that $q_n \to \infty$, which ensures that it will have no periodic points in $M\backslash\mathrm{Fix}(S)$.

2.3. Transitivity

In this section, we explain how to construct a transitive pseudo-rotation with the previous general scheme. This construction is illustrated by Fig. 1. Every Anosov–Katok construction in Hamiltonian dynamics requires certain information on abundance of symplectic diffeomorphisms. For transitivity the required information is very light: we need the fact that the symplectic group acts transitively on p-tuples of distinct points, as expressed by Lemma 3 below.

We say that a subset Z of M is ε-*dense* in M if the open balls of radius ε around each point of Z cover M. Note that when ε-density holds for Z it also holds for some finite subset Z' of Z, by compactness, and for any subset Z'' close enough to Z'.

Let (ε_n) be any fixed sequence of positive numbers that converges to 0. Assume inductively that α_n, H_{n-1} have been constructed as above, with $f_{n-1} = H_{n-1} S_{\alpha_n} H_{n-1}^{-1}$ which is ε_{n-1}-*transitive*: there exists some point whose orbit is ε_{n-1}-dense, and more precisely there exists some $x_{n-1} \in M$ and $N_{n-1} > 0$ such that the balls

$$B_{\varepsilon_{n-1}}(x_{n-1}), \ldots, B_{\varepsilon_{n-1}}(f_{n-1}^{N_{n-1}}(x_{n-1})),$$

cover M. We now explain how to construct h_n, α_{n+1} so that f_n is ε_n-transitive and arbitrarily close to f_{n-1} (see also Fig. 1).

Let $U = M \backslash \mathrm{Fix}(S)$, and consider the quotient map

$$\pi : U \to M' = U/S_{\alpha_n}.$$

The space M' is a smooth manifold, on which ω induces a symplectic structure. Symplectic diffeomorphisms of M' corresponds to symplectic diffeomorphisms of U that commutes with S_{α_n}. The Hamiltonian circle action S induces a Hamiltonian circle action on M'. First choose some finite subset F_2 of M' such that $\pi^{-1}(F_2)$ is η_n-dense in M, where η_n is provided by the continuity of H_{n-1}: any two points that are η_n-close in M have their images ε_n-close. Then choose some finite set F_1 of M' included in a single orbit of the circle action, and which has the same cardinality as F_2. By the next lemma, there is a compactly supported symplectic diffeomorphism of M' that sends F_1 to F_2. We lift this diffeomorphism to a symplectic diffeomorphism h_n of M which is the identity near $\mathrm{Fix}(S)$ and sends $\pi^{-1}(F_1)$ to $\pi^{-1}(F_2)$.

Lemma 3. *Given a symplectic manifold M, and an integer $p > 0$, the group of (compactly supported) symplectic diffeomorphisms of M acts transitively on p-tuples of distinct points: for every $(x_1, \ldots, x_p), (y_1, \ldots, y_p) \in M^p$ with $x_i \neq x_j, y_i \neq y_j$ for every $i \neq j$, there is some $\Phi \in \mathrm{Symp}_0(M)$ such that $\Phi(x_i) = y_i$ for every $i = 1, \ldots, p$.*

Now denote by C the orbit of the circle action S which contains $\pi^{-1}(F_1)$. Note that $h_n(C)$ is η_n-dense in M, and thus $H_{n-1}(h_n(C))$ is ε_n-dense in M. Choose α_{n+1} to be a rational number very close to α_n, so that there is some point in C whose (discrete) orbit C' under $S_{\alpha_{n+1}}$ is η_n-dense in C. Then, $H_{n-1}(h_n(C'))$ is still ε_n-dense in M. Let $H_n = H_{n-1} \circ h_n$ and $f_n = H_n S_{\alpha_{n+1}} H_n^{-1}$ as in the general scheme, then $H_n(C')$ is an orbit of f_n which is ε_n-dense, as wanted. (Of course, we also make sure that f_n satisfies the constraints ensuring that the limit map f is a pseudo-rotation, by taking α_{n+1} even closer to α_n if needed, as described in the previous section.)

It remains to check that, provided the convergence is fast enough, the limit map f will have a dense orbit. At step n the map f_n has an ε_n dense orbit. The set of maps having an ε_n-dense orbit is open in the C^0-topology. Thus this property will be transmitted to the limit map f by fast convergence. It is a classical Baire category argument that a map that has ε-dense orbits for arbitrarily small ε's has a dense orbit. Thus, f is transitive.

2.4. Pseudo-rotations with minimal number of ergodic measures

In this section, we prove Theorem 1 relying on Proposition 4 below whose proof takes up the remainder of the paper. Our standing assumption, while proving Theorem 1 and Proposition 4, is that (M, ω) is a toric symplectic manifold and that the locally free circle action S is compatible with the torus action in the sense that it is obtained by composing the torus action with a group morphism from the circle to the n-torus. We explain why S as described here exists in Sect. 3.1; see the discussion around Equation (3).

Denote by $\mathcal{P}(M)$ the space of Borel probability measures on M. This space is endowed with the weak topology: a sequence (μ_n) in $\mathcal{P}(M)$ converges to μ if and only if the sequence $(\int f d\mu_n)$ converges to $\int f d\mu$ for all continuous functions f. Recall that if μ is a Borel probability measure on a topological space N and $h : N \to M$ is a continuous mapping, then the pushforward of μ by h is defined by the formula

$$h_*\mu(E) := \mu(h^{-1}(E)).$$

Let $\mathcal{E} \subset \mathcal{P}(M)$ be the set consisting of the volume measure and the ℓ Dirac measures supported at the fixed points of the circle action. We denote by $\mathrm{Conv}(\mathcal{E})$ the convex hull of $\mathcal{E}$ in $\mathcal{P}(M)$. Lastly, recall (1), the definition of the set $\mathcal{Q} \subset \mathbb{N}$ of permitted denominators for the α_ns.

Proposition 4. *Let $q \in \mathcal{Q}$ be a positive integer and $\mathcal{U} \subset \mathcal{P}(M)$ an open neighborhood of* $\mathrm{Conv}(\mathcal{E})$*. There exists $h \in \mathrm{Symp}_0(M, \omega)$ with the following properties:*

1. *h coincides with the identity near* $\mathrm{Fix}(S)$*,*
2. *$hS_{\frac{1}{q}} = S_{\frac{1}{q}}h$,*
3. *For every $x \in M$, the push-forward of the Lebesgue measure on the circle by the map $t \mapsto hS_th^{-1}(x)$ belongs to $\mathcal{U}$.*

We will now show that the above proposition implies the existence of a pseudo-rotation whose set of invariant ergodic measures is exactly $\mathcal{E}$.

Proof of Theorem 1. Let $\mathcal{U}_n$ be a sequence of open neighborhoods of $\mathrm{Conv}(\mathcal{E})$, such that

$$\bigcap \mathcal{U}_n = \mathrm{Conv}(\mathcal{E}).$$

As explained above, we start step n with the rational number $\alpha_n = \frac{p_n}{q_n}$ and the maps

$$H_{n-1}, \quad f_{n-1} = H_{n-1}S_{\alpha_n}H_{n-1}^{-1}.$$

Since H_{n-1} is symplectic and fixes $\mathrm{Fix}(S)$, its action $(H_{n-1})_*$ on $\mathcal{P}(M)$ fixes every element of $\mathrm{Conv}(\mathcal{E})$. Thus, $(H_{n-1})_*^{-1}(\mathcal{U}_n)$ is an open neighborhood of $\mathrm{Conv}(\mathcal{E})$.

We denote the Lebesgue measure on the circle by $\mathrm{Leb}_{\mathbb{S}^1}$. Applying Proposition 4 to the integer q_n and the set $(H_{n-1})_*^{-1}(\mathcal{U}_n)$, we obtain $h_n \in \mathrm{Symp}(M, \omega)$ such that $S_{\alpha_n}h_n = h_nS_{\alpha_n}$ and

$$\left(t \mapsto h_nS_th_n^{-1}x\right)_* \mathrm{Leb}_{\mathbb{S}^1} \in (H_{n-1})_*^{-1}(\mathcal{U}_n), \ \forall x \in M. \tag{2}$$

Eq. (2) may equivalently be restated as

$$(H_{n-1}h_n)_*(t \mapsto S_ty)_*\mathrm{Leb}_{\mathbb{S}^1} \in \mathcal{U}_n, \ \forall y \in M.$$

We let $H_n = H_{n-1}h_n$. Consider the map

$$\begin{aligned} \Theta : \mathcal{P}(\mathbb{S}^1) \times M &\longrightarrow \mathcal{P}(M) \\ (\mu, x) &\longmapsto (H_n)_*(t \mapsto S_tx)_*\mu. \end{aligned}$$

This mapping is continuous, and since $\mathcal{P}(\mathbb{S}^1)$ and M are both compact, it is uniformly continuous. Now by Eq. (2), $\Theta(\mathrm{Leb}_{\mathbb{S}^1}, x) \in \mathcal{U}_n$ for every $x \in M$. Thus, there exists a neighborhood $\mathcal{V}$ of $\mathrm{Leb}_{\mathbb{S}^1}$ such that $\Theta(\mathcal{V} \times M) \subset \mathcal{U}_n$.

Let $\alpha_{n+1} = \frac{p_{n+1}}{q_{n+1}}$ be a rational written in irreducible form, and denote by $\mu_n \in \mathcal{P}(\mathbb{S}^1)$ the measure given by the average of the Dirac measures on the orbit of $0 \in \mathbb{S}^1$ under the circle rotation by α_{n+1}. If q_{n+1} is large enough then the measure μ_n will be in $\mathcal{V}$ (by convergence of the Riemann sums). Hence we get $\Theta(\mu_n, x) \in \mathcal{U}_n$ for all $x \in M$. Note that $\Theta(\mu_n, x)$ is the average of the Dirac measures along the orbit of $H_n(x)$ under the map $f_n = H_n S_{\alpha_{n+1}} H_n^{-1}$. Lastly, we additionally impose that $q_{n+1} \in \mathcal{Q}$.

As explained in Sect. 2.1, this construction provides a sequence of Hamiltonian diffeomorphisms $f_n = H_n S_{\alpha_{n+1}} H_n^{-1}$ that converges, in C^∞ topology, to a Hamiltonian diffeomorphism f. We will now prove that, by appropriately choosing the sequence $(\alpha_n) = (\frac{p_n}{q_n})$, we can ensure that the limit map f has the desired property: its set of invariant ergodic measures is exactly $\mathcal{E}$.

At the beginning of step n of the construction, we define the set $U(q_n) \subset \mathrm{Ham}(M, \omega)$ consisting of all Hamiltonian diffeomorphisms g which have the following property: for every $x \in M$, the probability measure $\frac{1}{q_n} \sum_{k=0}^{q_n - 1} \delta_{g^k x}$ belongs to the set $\mathcal{U}_{n-1}$. This set contains f_{n-1} and is open in the C^0-topology. The numbers α_{i+1}, for $i \geqslant n$, will be chosen such that the Hamiltonian diffeomorphisms f_i, for $i \geqslant n$, are all contained in a (C^∞) neighborhood of f_n whose (C^∞) closure is contained in $U(q_n)$. This implies that the limit map f will also be contained in $U(q_n)$. Of course, the α_n's may be picked such that the map f will have this property for every n. In other words, for every n and every $x \in M$, the probability measure

$$\nu_n := \frac{1}{q_n} \sum_{k=0}^{q_n - 1} \delta_{f^k x}$$

is contained in $\mathcal{U}_n$.

We claim this implies that the set of invariant ergodic probability measures of f is exactly $\mathcal{E}$. Indeed, to obtain a contradiction, assume that this is not the case: f has an invariant ergodic probability measure μ which is not in $\mathcal{E}$. Since the ergodic measures are extremal points in the set of invariant probability measures, the measure μ does not belong to $\mathrm{Conv}(\mathcal{E})$. This entails the existence of a continuous function $\varphi : M \to \mathbb{R}$ which vanishes on $\mathrm{Fix}(S)$ and has the property that $\int \varphi \, d\mathrm{Vol} = 0$ but $\int \varphi d\mu \neq 0$. To see that such φ indeed exists, observe that a probability measure γ belongs to $\mathrm{Conv}(\mathcal{E})$ if and only if it satisfies the following criterion: for every pair of open sets O, O' which are disjoint from $\mathrm{Fix}(S)$ and have the same volume, we must have $\gamma(O) = \gamma(O')$.

By Birkhoff's Ergodic Theorem and the ergodicity of μ, there exists $x \in M$ such that the Birkhoff means of φ converge to $\int \varphi d\mu$. This, in particular, means that the sequence

$$\int \varphi \, d\nu_n = \frac{1}{q_n} \sum_{k=0}^{q_n - 1} \varphi(f^k x)$$

converges to the non-zero number $\int \varphi d\mu$. The contradiction is that, up to passing to a subsequence, the sequence $\int \varphi d\nu_n$ converges to zero because $\nu_n \in \mathcal{U}_n$ and thus must have a weak limit $\nu \in \mathrm{Conv}(\mathcal{E})$ and φ was picked such that $\int \varphi \, d\nu = 0$ for all $\nu \in \mathrm{Conv}(\mathcal{E})$. □

3. Preliminaries

The goal of this section is to recall some basic notions of symplectic & differential geometry as well as proving certain preliminary results which will be used in the proof of Proposition 4.

3.1. Preliminaries on symplectic geometry

Throughout the section (M, ω) denotes a symplectic manifold. Recall that a symplectomorphism is a diffeomorphism $\varphi : M \to M$ such that $\varphi^*\omega = \omega$. The set of all symplectic diffeomorphisms of M is denoted by $\mathrm{Symp}(M, \omega)$. We will let $\mathrm{Symp}_0(M, \omega)$ denote those elements of $\mathrm{Symp}(M, \omega)$ which are isotopic to the identity via a compactly supported isotopy. Note that the assumption on compactness of the support of the isotopy is not common.

Hamiltonian diffeomorphisms constitute an important class of examples of symplectic diffeomorphisms. These are defined as follows. A smooth, compactly supported, Hamiltonian $H \in C_c^\infty([0,1] \times M)$ gives rise to a time-dependent vector field X_H which is defined via the equation: $\omega(X_H(t), \cdot) = -dH_t$. The Hamiltonian flow of H, denoted by ϕ_H^t, is by definition the flow of X_H. A compactly supported Hamiltonian diffeomorphism is a diffeomorphism which arises as the time-one map of a Hamiltonian flow generated by a compactly supported Hamiltonian. The set of all compactly supported Hamiltonian diffeomorphisms is denoted by $\mathrm{Ham}(M, \omega)$.

Toric symplectic manifolds. Recall that a **toric** symplectic manifold is a closed and connected symplectic manifold (M^{2n}, ω) equipped with an effective Hamiltonian action of a torus $\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1$ whose dimension is half that of M, i.e. a one-to-one smooth morphism from $\mathbb{T}^n$ to $\mathrm{Ham}(M, \omega)$. Throughout the paper, we identify $\mathbb{S}^1$ with $\mathbb{R}/\mathbb{Z}$ and $\mathbb{T}^n$ with $\mathbb{R}^n/\mathbb{Z}^n$.

Each of the circle factors in the torus $\mathbb{T}^n$ yields a Hamiltonian circle action, or equivalently a periodic Hamiltonian flow with period 1; we will denote by $\mu_1, \ldots, \mu_n \in C^\infty(M)$ the Hamiltonians corresponding to these circle actions. The **moment map** $\mu : M \to \mathbb{R}^n$ is defined by $x \mapsto (\mu_1(x), \ldots, \mu_n(x))$. Its image $\Delta := \mu(M)$ is referred to as the **moment polytope** of the action. According to the convexity theorem of Atiyah and Guillemin–Sternberg [2, 22] Δ is a convex polytope whose vertices are the images of the fixed points of the action. Furthermore, Δ is simple (there are n edges meeting at each vertex), rational (the edges meeting at a vertex v are of the form $v + tu_i$ where $t \geqslant 0$ and $u_i \in \mathbb{Z}^n$), and smooth (at each vertex v of Δ the corresponding $u_1, \ldots, u_n$ may be chosen to be a $\mathbb{Z}$-basis of $\mathbb{Z}^n$).

The points of M with non trivial stabiliser are exactly the points which are mapped under μ to the boundary of Δ. Furthermore, all the points in the inverse image of a given face of Δ have the same stabiliser. Thus, we see that there are finitely many stabilisers, each of which is a subgroup of $\mathbb{T}^n$

isomorphic to $\mathbb{T}^p$, where p is the codimension of the corresponding face of Δ. In particular, we may pick a one-to-one morphism

$$\begin{aligned} \Phi : \mathbb{S}^1 &\to \mathbb{T}^n \\ t &\mapsto tZ, \end{aligned} \tag{3}$$

for some primitive element $Z \in \mathbb{Z}^n$, such that the image of Φ is not included in the union of the stabilisers. Then, the composition of Φ with the torus action yields a circle action S which is locally free on $M \backslash \mathrm{Fix}(S)$. Such a circle action will be at the basis of our Anosov–Katok construction.

It turns out that one can construct a section of the moment map, that is, a continuous map $\sigma : \Delta \to M$ such that $\mu \circ \sigma = \mathrm{Id}$. The fact that the section σ exists can be deduced from the proof of Delzant's theorem on classification of toric symplectic manifolds; see for example the construction of the Delzant space X_Δ, pages 9 – 13 in [21]: in the case of the model X_Δ, one can see directly from the construction that the moment map admits a section; existence of the section σ for arbitrary toric symplectic manifolds then follows from Delzant's theorem that such manifolds are classified by their moment polytopes. Furthermore, the section σ is smooth on $\mathrm{Int}(\Delta)$.

Given such a section $\sigma : \Delta \to M$ of the moment map, define the mapping $\Xi : \Delta \times \mathbb{T}^n \to M$ by

$$(s, t) \mapsto \mathbb{T}^n_t(\sigma(s)), \tag{4}$$

where $\mathbb{T}^n_t(\sigma(s))$ denotes the image of the point $\sigma(s)$ under the action of $t \in \mathbb{T}^n$. Endow $\mathrm{Int}(\Delta) \times \mathbb{T}^n$ with the symplectic form

$$\frac{1}{2\pi} \Sigma_{i=1}^n d\mu_i \wedge d\theta_i.$$

Then, the mapping Ξ is a symplectomorphism between $\mathrm{Int}(\Delta) \times \mathbb{T}^n$ and $\mu^{-1}(\mathrm{Int}(\Delta))$. This yields a global system of symplectic action-angle coordinates on $\mu^{-1}(\Delta)$ (this is the content of Remark IV.4.19 on global action-angle coordinates in [3]). Note however that Ξ is not one-to-one on $\Delta \times \mathbb{T}^n$, since points on $\mu^{-1}(\partial\Delta)$ have non-trivial stabilisers. One can also describe the action in the neighborhood of any degenerate orbit. In particular, there is a local normal form near the orbit of any point x that depends only on the dimension of the face of the moment polytope the interior of which x belongs. In the sequel, we will just use the fact that for every face F, $\mu^{-1}(F)$ is a submanifold of M (see, for example, [3], Proposition IV.4.16). For further details on toric symplectic manifolds, we refer the reader to the books [3, 21].

3.2. Preliminary lemmas

We will be using the following lemmas in the course of proving Proposition 4. In many cases, the lemmas of this section will be applied to a quotient symplectic manifold of the form $M' = (M \backslash \mathrm{Fix}(S))/S_{\frac{1}{q}}$ where $q \in \mathcal{Q}$; recall that M' is indeed a manifold because picking $q \in \mathcal{Q}$ guarantees that the action of $S_{\frac{1}{q}}$ on $M \backslash \mathrm{Fix}(S)$ is free.

We will be using the following notations throughout the remainder of the paper. As before, we consider a circle action $S : (x, t) \mapsto S_t(x)$ on a symplectic manifold M, where $t \in \mathbb{S}^1$ and $x \in M$ and we denote the Lebesgue probability measure on the circle by Let $\mathrm{Leb}_{\mathbb{S}^1}$. Given a measurable subset E of M and a point x in M, the *time spent by the orbit of x in E* is the number

$$\mathrm{Leb}_{\mathbb{S}^1}(\{t \in \mathbb{S}^1 : S_t(x) \in E\}).$$

The *vector field tangent to the circle action S* is defined as

$$\vec{V}(x) = \frac{d}{dt} S_t(x)_{|t=0}.$$

The orbits of the action are the integral curves of $\vec{V}$. Given a diffeomorphism Φ, the action given by $S'_t(x) = \Phi S_t(\Phi^{-1}x)$ is called the conjugated action. The vector field tangent to the conjugated action is the push-forward vector field $\Phi_*\vec{V}$.

(a) The transversality lemma. Given a vector field $\vec{V}$ on some manifold M', and a submanifold X, we wish to remove the tangency points between $\vec{V}$ and X by performing a small perturbation of X, thus obtaining a submanifold which is transverse to every integral curve of the vector field $\vec{V}$. Achieving this form of transversality is not in general possible: consider the case where $\vec{V}$ is a horizontal vector field in $\mathbb{R}^3$, with X being the graph of a function on $\mathbb{R}^2$; we cannot remove the points on X corresponding to local maxima or minima. However, the following lemma provides a partial "fix" for this situation: it is possible to perform a small perturbation of X, or equivalently $\vec{V}$, so that X becomes "as transverse as possible" to the integral curves of $\vec{V}$.

We suppose that $X \subset M'$ is a submanifold without boundary (but not necessarily compact[6]). Let K be a compact subset of X. We will say that X is *almost transverse* to a vector field $\vec{V}$ on K if for every integral curve γ of $\vec{V}$, every point of $\gamma \cap K$ is isolated in $\gamma \cap X$.

We will say that X is *stably almost transverse* to $\vec{V}$ on K if this property holds not only for $\vec{V}$ but also for $\Phi_*\vec{V}$ for every C^∞ diffeomorphism Φ of M' in some neighborhood of the identity.

Lemma 5. *Consider a symplectic manifold M' with a non-vanishing vector field $\vec{V}$. Let X be a submanifold of M' of codimension at least* 1*, without boundary, and K a compact subset of X. For any C^∞ neighborhood of the identity $\mathcal{W} \subset \mathrm{Symp}(M,\omega)$ and any open neighborhood O of X, there exists $\Phi \in \mathcal{W}$, whose support is contained in O, such that X is stably almost transverse to $\Phi_*\vec{V}$ on K.*

(b) The thickening lemma.

Lemma 6. *Consider a metric space M' with a continuous circle action S. Let X be a compact subset of M', and assume that no orbit of the action spends a positive amount of time in X.*

Then, for every $\varepsilon > 0$, there exists $\delta > 0$ such that no orbit of the action spends more time than ε in the δ-neighborhood of X.

[6]In the applications of the lemma, the manifold M' will be a quotient of the complement of the fixed points set of our circle action.

(c) The stability lemma.

Lemma 7. *Consider a symplectic manifold M' with a locally free circle action S. Let K_1, K_2 be compact subsets of M' with $K_1 \subset \mathrm{Int}(K_2)$.*

Let $c \in (0,1)$ and assume that every orbit of the action S spends more time than c in K_1. Then, there exists a C^0 neighborhood $\mathcal{W}$ of $\mathrm{Id} \in \mathrm{Diff}(M)$ such that every orbit of the conjugated action $\Phi S \Phi^{-1}$ spends more time than c in K_2.

(d) The transportation lemmas. Consider $\mathbb{R}^{2n}$ equipped with the coordinates $x_1, y_1, \ldots, x_n, y_n$. A standard polydisc in $\mathbb{R}^{2n}$ is a subset of the form

$$\prod_{i=1}^{n}[a_i, b_i] \times [c_i, d_i] := \{(x_i, y_i) : x_i \in [a_i, b_i], y_i \in [c_i, d_i]\}.$$

By a *polydisc* in M we mean the image of a symplectic embedding of a standard polydisc. Note that our polydiscs are all closed.

In the next two lemmas, P denotes the polydisc $[-a, a]^{2n} \subset \mathbb{R}^{2n}$ for some $a > 0$. We assume that P is equipped with the standard symplectic structure it inherits from $\mathbb{R}^{2n}$.

Lemma 8. *Let $\phi_1, \phi_2 : P \to M''$ be two symplectic embeddings of P into a symplectic manifold (M'', ω). There exists $\delta_0 > 0$ and a compactly supported $\Psi \in \mathrm{Symp}_0(M'')$ such that*

$$\Psi \circ \phi_1 = \phi_2 \text{ on } [-\delta_0, \delta_0]^{2n}.$$

Furthermore, if $\phi_1(0) \neq \phi_2(0)$, then we may require, in addition to the above, that

$$\Psi \circ \phi_2 = \phi_1 \text{ on } [-\delta_0, \delta_0]^{2n}.$$

In the next lemma, $P_1, P_2 \subset \mathbb{R}^{2n}$ denote the polydiscs $p_1 + [-b, b]^{2n}$, $p_2 + [-b, b]^{2n}$, respectively, where p_1, p_2 are points in $\mathbb{R}^{2n}$. The polydisc P is as in the above lemma. The following statement is well known in the two dimensional setting when $n = 1$; we leave it to the reader to check that the proof can be reduced to the case where $n = 1$.

Lemma 9. *Suppose that P_1, P_2 are disjoint and are contained in the interior of P. There exists a symplectomorphism Ψ, whose support is compactly contained in the interior of P, such that*

$$\Psi(P_1) = P_2 \ \& \ \Psi(P_2) = P_1.$$

3.3. Proofs of the lemmas

In this section, we will provide proofs for Lemmas 5, 6, 7, 8 of the previous section, leaving the proof of Lemma 9 to the reader.

(a) Proof of lemma 5: Transversality lemma. We will need the following notion in the course of the proof. Let M' be as in the statement of the lemma and X be a hypersurface without boundary in M'. Given a vector field $\vec{V}$ and a positive integer k, we say that X *has a contact of order* k *with* $\vec{V}$ at some point $x \in X$ if there exist an integral curve $\alpha : (-\varepsilon, \varepsilon) \to M'$ of $\vec{V}$ and a smooth curve $\beta : (-\varepsilon, \varepsilon) \to X$ such that $\alpha(0) = \beta(0) = x$ and the derivatives of α and β coincide at 0 up to order k.

Remark 10. Here are some useful observations about the above notion.

1. If Φ is a diffeomorphism, then $\Phi(X)$ has a contact of order k with $\vec{V}$ if and only if X has a contact of order k with $\Phi_*^{-1}\vec{V}$.
2. If $\alpha : I \to M'$ is any integral curve of $\vec{V}$ and if the intersection set
$$\alpha(I) \cap X$$
has an accumulation point at $z = \alpha(t)$, then X has a contact of every positive order with $\vec{V}$ at z.
3. If K is a compact subset of X, then the set of diffeomorphisms Φ such that X has no contact of order k with $\Phi_*\vec{V}$ at points of K is open in the space of diffeomorphisms equipped with the C^∞ topology.
4. If X has no contact of order k with $\vec{V}$ on K, for some positive integer k and some compact subset K of X, then X is stably almost transverse to $\vec{V}$ on K. This is a direct consequence of the previous two points.

Our proof of Lemma 5 requires the claim below. We assume the hypotheses of Lemma 5.

Claim 11. *Consider a symplectic manifold M', a submanifold $X \subset M'$ without boundary and of codimension at least* 1, *and some neighborhood O of X. Let $k \geqslant \dim(M')$.*

Then, every point $z_0 \in X$ admits a pair of open neighborhoods (W, U) with $W \subset \overline{W} \subset U$ in M', which are contained in O and have the following property. For every non-vanishing vector field $\vec{V}$, and every neighbourhood $\mathcal{N}$ of the identity in $\mathrm{Symp}(M')$, *there exists $\Phi \in \mathcal{N}$, compactly supported in U, such that X has no contact of order k with $\Phi_*\vec{V}$ in $\overline{W}$.*

Before proving Claim 11, we will show that it implies Lemma 5.

Proof of the transversality lemma. Observe that it is sufficient to find $\Phi \in \mathcal{W}$, with support in O, such that $\Phi(X)$ has no contact of order $k = \dim(M')$ with $\vec{V}$ at any point of K; see Remark 10.

For each point $z \in K$, let W_z, U_z be open neighborhoods of z in M' as provided by Claim 11. The collections W_z, U_z are two coverings of K by open subsets of M'. By passing to subcovers we may assume these coverings are finite and we may denote them by W_i, U_i, where $i \in \{1, \ldots, N\}$

A first application of the claim provides $\Phi_1 \in \mathcal{W}$, with support in $U_1 \subset O$, such that X has no contact of order k with $\Phi_{1*}\vec{V}$ on $\overline{W}_1$. A second application provides $\Phi_2 \in \mathrm{Symp}(M')$ such that X has no contact of order k with $(\Phi_2 \circ \Phi_1)_*\vec{V}$ on $\overline{W}_2$. Furthermore, since Φ_2 may be chosen to be arbitrarily C^∞ close to the identity, by point 3 of Remark 10 we can ensure that $\Phi_2 \circ \Phi_1 \in \mathcal{W}$ and that X has no contact of order k with $(\Phi_2 \circ \Phi_1)_* \vec{V}$ on $\overline{W}_1$ either. Note also that $\Phi_2 \circ \Phi_1$ is supported in $U_1 \cup U_2 \subset O$.

We proceed to build a sequence $(\Phi_i)_{i=1,\dots,N}$ in $\mathrm{Symp}(M')$ with analogous properties. In particular, the map $\Phi = \Phi_N \circ \cdots \circ \Phi_1$ belongs to $\mathcal{W}$, is supported in $U_1 \cup \cdots \cup U_N \subset O$, and X has no contact of order k with $\Phi_* \vec{V}$ on $\overline{W}_1 \cup \cdots \cup \overline{W}_N$, which contains K. By point 4 of Remark 10, X is stably almost transverse to $\Phi_* \vec{V}$ on K, as wanted. $\square$

Proof of Claim 11. Since every submanifold is contained in an open hypersurface without boundary, we will restrict ourselves to the case where X is a hyper surface. Take $(U, x_1, y_1, x_2, y_2, \dots, x_n, y_n)$ to be a Darboux chart centered at z_0, and contained in O, in which X is given by the equation $x_n = 0$. By working in these local coordinates we may assume that $M' = \mathbb{R}^{2n}$, X is the hyperplane given by $x_n = 0$, and the point $z_0 = 0$, where 0 is the origin in $\mathbb{R}^{2n}$. We pick W to be any open neighborhood of the origin such that $\overline{W}$ is contained in U.

Denote $\dim(M') = 2n$. To prove the claim, we will translate the property of having contact of order k to an equivalent property in the space $J^k(\mathbb{R}^{2n}, \mathbb{R})$ of k-jets of maps from $\mathbb{R}^{2n}$ to $\mathbb{R}$. Recall that $J^k(\mathbb{R}^{2n}, \mathbb{R})$ identifies with $\mathbb{R}^{2n} \times \mathrm{Pol}_k$, where Pol_k is the linear space of polynomials of degree at most k on $\mathbb{R}^{2n}$; see for example Section 1.1 of [13].

We now consider a non vanishing vector field $\vec{V}$ on $\mathbb{R}^{2n}$. Let $\Sigma_{\vec{V}}$ be the set of couples $(z, P) \in J^k(\mathbb{R}^{2n}, \mathbb{R})$ such that there exists some integral curve $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^{2n}$ of the vector field $\vec{V}$ with $\gamma(0) = z$ and $\lim_{t \to 0} \frac{P(\gamma(t))}{t^k} = 0$. Note that $\Sigma_{\vec{V}}$ is a submanifold of $J^k(\mathbb{R}^{2n}, \mathbb{R})$ of codimension $k+1$. Indeed, to verify this particular point we can work in a small neighborhood of the point $z \in \mathbb{R}^{2n}$ where we may assume that $\vec{V}$ coincides with the vector field $\frac{\partial}{\partial x_1}$. Then $\Sigma_{\vec{V}}$ may be described locally as

$$\left\{ (z, P) : P(0) = 0, \frac{\partial P}{\partial x_1}(0) = 0, \dots, \frac{\partial^k P}{\partial x_1}(0) = 0 \right\},$$

which is a linear subspace of $J^k(\mathbb{R}^{2n}, \mathbb{R})$ of codimension $k+1$.

Finally, here is the promised translation. Let $f : \mathbb{R}^{2n} \to \mathbb{R}$ be a smooth function and suppose that $f^{-1}(0)$ is a hypersurface. Then, $j^k(f)_z$, the k-jet of f at z, belongs to $\Sigma_{\vec{V}}$ if and only if $f^{-1}(0)$ has a contact of order k with $\vec{V}$ at z.

For the rest of the proof we set $f = x_n$ so that $X = f^{-1}(0)$. We claim that we can find a smooth function $H : \mathbb{R}^{2n} \to \mathbb{R}$, arbitrarily C^∞ close to 0, such that under the map $\mathbb{R}^{2n} \to J^k(\mathbb{R}^{2n}, \mathbb{R})$,

$$z \mapsto j^k(f \circ \phi_H^1)_z,$$

the image of U is disjoint from $\Sigma_{\vec{V}}$.

Let us first suppose that such H exists and explain how this leads to the conclusion. Note that $(f \circ \phi_H^1)^{-1}(0) = \phi_H^{-1}(X)$ and so we have that $\phi_H^{-1}(X)$ has no contact of order k with $\vec{V}$ inside the set U. Since H can be picked to be arbitrarily C^∞ close to 0, we may perform a cut-off to obtain a function

G, which is compactly supported in U and is still C^∞ close to zero, such that $\phi_G^{-1}(X)$ has no contact of order k with $\vec{V}$ inside $\overline{W}$.

It remains to explain how to find H. Consider the map

$$\begin{aligned}\Psi : \overline{U} \times \mathrm{Pol}_{k+1} &\to J^k(\overline{U}, \mathbb{R})\\ (z, H) &\mapsto j^k(f \circ \phi_H^1)_z.\end{aligned}$$

We will check below that there exists some neighborhood $\mathcal{U}$ of 0 in Pol_{k+1} such that Ψ is a submersion at every point $(z, H) \in \overline{U} \times \mathcal{U}$.

This entails that the restriction of Ψ to $\overline{U} \times \mathcal{U}$ is transverse to any submanifold; in particular it is transverse to $\Sigma_{\vec{V}}$. Next, applying the parametric transversality theorem, we obtain a dense subset of polynomials $\mathcal{H} \subset \mathcal{U}$ such that for any fixed $H \in \mathcal{H}$ the mapping

$$\begin{aligned}\Psi(\cdot, H) : \overline{U} &\to J^k(\overline{U}, \mathbb{R})\\ z &\mapsto j^k(f \circ \phi_H^1)_z,\end{aligned}$$

is transverse to $\Sigma_{\vec{V}}$. Since $k \geqslant 2n$, we get $\dim(\mathbb{R}^{2n}) < k + 1 = \mathrm{codim}(\Sigma_{\vec{V}})$. Then, transversality implies that the image of $\overline{U}$ under $z \mapsto j^k(f \circ \phi_H^1)_z$ is actually disjoint from $\Sigma_{\vec{V}}$, which is the desired property.

It remains to check that Ψ is a submersion for all H in some neighborhood of 0 in Pol_{k+1}. Since being a submersion is an open property, and $\overline{U}$ is compact, it is enough to check that the mapping Ψ is a submersion when $H = 0$. In the computations below we will identify $J^k(\overline{U}, \mathbb{R}) = \overline{U} \times \mathrm{Pol}_k$. Furthermore, we will also identify the tangent spaces to points of $\overline{U} \times \mathrm{Pol}_k$ and $\overline{U} \times \mathrm{Pol}_{k+1}$, with $\mathbb{R}^{2n} \times \mathrm{Pol}_k$ and $\mathbb{R}^{2n} \times \mathrm{Pol}_{k+1}$, respectively.

We would like to compute $D\Psi_{(z,0)}(u, G)$ where u is a vector in $\overline{U}$ and $G \in \mathrm{Pol}_{k+1}$. It is easy to see that $D\Psi_{(z,0)}(u, 0) = (u, 0)$. Hence, we will only consider the case where $u = 0$. Now, we compute:

$$\begin{aligned}D\Psi_{(z,0)}(0, G) &= \frac{\partial}{\partial t}|_{t=0}\ \Psi(z, tG) = \frac{\partial}{\partial t}|_{t=0}\ j^k(f \circ \phi_G^t)_z\\ &= j^k(\frac{\partial}{\partial t}|_{t=0}\ f \circ \phi_G^t)_z = j^k(\{f, G\})_z,\end{aligned}$$

where we have used the fact that $\frac{\partial}{\partial t}|_{t=0}\ f \circ \phi_G^t = \{f, G\}$. Here, $\{f, G\}$ denotes the Poisson bracket of f and G and it is given by $\{f, G\} = \sum_{i=1}^{2n} \frac{\partial f}{\partial x_i}\frac{\partial G}{\partial y_i} - \frac{\partial f}{\partial y_i}\frac{\partial G}{\partial x_i}$. Recall that we picked $f = x_n$ and so $\{f, G\} = \frac{\partial G}{\partial y_n}$. Hence, we have established

$$D\Psi_{(z,0)}(0, G) = j^k\left(\frac{\partial G}{\partial y_n}\right)_z.$$

Now, for any given polynomial $P \in \mathrm{Pol}_k$ let $G = y_n P$ and note that $j^k\left(\frac{\partial G}{\partial y_n}\right)_z = P$. We conclude that $D\Psi_{(z,0)}$ is surjective, for any $z \in \overline{U}$, and so Ψ is a submersion at $(z, 0)$ for all $z \in \overline{U}$. $\square$

(b) Proof of lemma 6: Thickening lemma. As before, we will denote by $t \mapsto S_t$ the action of the circle by homeomorphisms on M. Fix $\varepsilon > 0$ and let $x \in X$. Since $\mathrm{Leb}_{\mathbb{S}^1}(\{t \in \mathbb{S}^1 : S_t(x) \in X\}) = 0$, we can find a compact subset $I_x \subset \mathbb{S}^1$, whose Lebesgue measure is more than $1 - \varepsilon$, and such that for every $t \in I_x$, we have $S_t(x) \notin X$. Compactness of I_x and X implies that there is some $\delta_x > 0$ such that $d(S_t(x), X) > \delta_x$ for every $t \in I_x$. Observe that for every $x \in X$, we can find a neighborhood V_x of x in M' such that we still have $d(S_t(y), X) > \delta_x$ for $t \in I_x$ and $y \in V_x$. Since X is compact, we can find $V_{x_1}, \ldots, V_{x_m}$ such that the finite union $V := \cup_i V_{x_i}$ covers X. Pick $0 < \delta$, less than each δ_{x_i}, such that $V_\delta(X)$, the δ-neighborhood of X, is contained in V. Observe that we have proven the following statement: for every $y \in V$

$$\mathrm{Leb}(\{t \in \mathbb{S}^1 : S_t(y) \in V_\delta(X)\}) < \varepsilon.$$

The above inequality also holds for any point y whose orbit under the action meets V, since its orbit coincides with the orbit of a point in V. On the other hand, if y is a point in M' whose orbit does not meet V, then

$$\mathrm{Leb}(\{t \in \mathbb{S}^1 : S_t(y) \in V_\delta(X)) = 0.$$

This completes the proof.

(c) Proof of lemma 7: Stability lemma. If Φ is uniformly close to the identity then the orbits of the conjugated action are uniformly close to the orbits of the action S. On the other hand the hypotheses entail that the distance from K_1 to the complement of K_2 is positive. The lemma follows.

(d) Proof of lemma 8: Transportation lemma. We begin by proving the first statement of the lemma, that is there exist $\Psi \in \mathrm{Symp}_0(M'', \omega)$ and $\delta_0 > 0$ such that $\Psi\phi_1 = \phi_2$ on $[-\delta_0, \delta_0]^{2n}$.

Note that since $\mathrm{Symp}_0(M'')$ acts transitively on M'' we can suppose that $\phi_1(0) = \phi_2(0)$. Let U be a Darboux chart centered at $\phi_1(0) = \phi_2(0)$. By replacing P with a smaller polytope, we may suppose that the image of P under ϕ_1, ϕ_2 is contained in the Darboux chart U. This allows us to reduce the problem to the following setting: $M'' = \mathbb{R}^{2n}$ and ϕ_1, ϕ_2 are symplectic embeddings of a polydisc $P = [-a, a]^{2n}$ such that $\phi_1(0) = \phi_2(0) = 0$. We must find $\delta_0 > 0$ and a symplectic isotopy Ψ^t, which is compactly supported in U, with the property that $\Psi^0 = \mathrm{Id}$ and $\Psi^1 = \phi_2\phi_1^{-1}$ on $[-\delta_0, \delta_0]^{2n}$.

Pick any $0 < b < a$ and let $P' = [-b, b]^{2n}$. By the well-known "extension after restriction principle" [12], we can find a compactly supported symplectomorphism ψ of $\mathbb{R}^{2n}$ such that $\psi(0) = 0$ and $\psi|_{P'} = \phi_2\phi_1^{-1}|_{P'}$. There exists a symplectic isotopy $(\Psi^t)_{t\in[0,1]}$, compactly supported in $\mathbb{R}^{2n}$, such that $\Psi^0 = \mathrm{Id}$, $\Psi^t(0) = 0$ for every $t \in [0, 1]$, and $\Psi^1 = \psi$; for an explanation see [31], proof of Proposition 1.7. There exists $\delta_0 < b$ such that

$$K := \bigcup_{t\in[0,1]} \Psi^t([-\delta_0, \delta_0]^{2n}) \subset U.$$

Let H be the Hamiltonian whose flow is Ψ^t. By cutting off the Hamiltonian H in a neighborhood of the set K, we obtain a new Hamiltonian G

which is supported in U and has the property that $\phi^1_G = \psi$ on $[-\delta_0, \delta_0]^{2n}$. We set $\Psi = \phi^1_G$. This completes the proof of the first part of the lemma.

For the second part of Lemma 8, note that we can find a compactly supported symplectomorphism $\theta \in \mathrm{Symp}_0(M'')$ which exchanges $\phi_1(0)$ and $\phi_2(0)$. Now, if we apply the above construction, which is local, independently once near $\phi_1(0)$ and another time near $\phi_2(0)$ we will obtain Ψ_1, Ψ_2 supported near $\phi_1(0), \phi_2(0)$, respectively, such that $\Psi_1\theta\phi_2 = \phi_1$ and $\Psi_2\theta\phi_1 = \phi_2$ on $[-\delta_0, \delta_0]^{2n}$ for some $\delta_0 > 0$. We let $\Psi = \Psi_1\Psi_2\theta$.

4. Proof of proposition 4

For the rest of the paper (M, ω) will denote a closed toric symplectic manifold with moment map $\mu : M \to \Delta$. We fix a locally free Hamiltonian circle action S obtained as described via Equation (3). We consider an integer $q \in \mathcal{Q}$ and a small positive $\varepsilon \in \mathbb{R}$. We fix a Riemannian metric on M and denote

$$B_\varepsilon(\mathrm{Fix}(S)) = \bigcup_{x \in \mathrm{Fix}(S)} B_\varepsilon(x),$$

where $B_\varepsilon(x)$ is the ball of radius ε around x.

Before giving the proof of Proposition 4 in detail, we outline here the main ideas of the proof. Our goal is to construct a Hamiltonian diffeomorphism h satisfying the three requirements of Proposition 4. To ensure that the first two properties are satisfied, we will carry out the construction on a compact subset of the quotient $(M \backslash \mathrm{Fix}(S))/S_{\frac{1}{q}}$. Recall that q being in $\mathcal{Q}$ guarantees that this quotient is a smooth symplectic manifold; see Equation (1).

The difficult part of our task is to ensure that the third property in Proposition 4 is satisfied. To achieve this, we will first construct, in Lemma 12, a collection of closed sets denoted by $A_1, \ldots, A_N$ satisfying the following equidistribution property:

The sets A_i have diameters less than ε, their interiors are disjoint, their volumes are equal, their boundaries are of zero volume and

$$\mathrm{Fix}(S) \subset W := \left(M \backslash \bigcup_i A_i\right) \subset B_\varepsilon(\mathrm{Fix}(S)).$$

We will refer to the A_i's as the *equidistribution boxes*. Then, we will construct h such that for each $x \in M$, the orbit $hS_th^{-1}(x)$ under the conjugated action spends roughly the same amount of time in each of the A_i's: for all i, j we will have

$$\mathrm{Leb}(\{t \in \mathbb{S}^1 : hS_th^{-1}(x) \in A_i\}) \approx \mathrm{Leb}(\{t \in \mathbb{S}^1 : hS_th^{-1}(x) \in A_j\}).$$

A more precise version of the above statement, along with the proof of the fact that it implies the 3rd item in Proposition 4, is given in Proposition 20.

To construct h we fix $\varepsilon' > 0$ and we will build a collection of disjoint, symplectomorphic polydiscs c_k with the following two critical properties:

i. There exists a C^∞-small symplectomorphism Ψ, commuting with $S_{\frac{1}{q}}$, such that for each $x \in M$ its orbit $\Psi S_t \Psi^{-1}(x)$, under the conjugated action will spend more time than $1-\varepsilon'$ in $W \cup_k c_k$:

$$\mathrm{Leb}(\{t \in \mathbb{S}^1 : \Psi S_t \Psi^{-1}(x) \in W \cup_k c_k\}) > 1-\varepsilon'.$$

ii. There exists a symplectomorphism Θ, commuting with $S_{\frac{1}{q}}$, which equidistributes the small boxes c_k among the equidistribution boxes A_i.

The symplectomorphism h will then be the composition $\Theta \circ \Psi$. The construction of the small boxes and the symplectomorphism Ψ is carried out in Lemma 16. The symplectomorphism Θ is constructed in Claim 21 in the course of the proof of Proposition 20.

We should mention that, in general, there exist symplectic obstructions to finding a symplectomorphism equidistributing a given collection of polydiscs; see for example the symplectic camel problem in Sect. 1.2 of [28]. To avoid such obstruction, the small boxes c_k must be picked to be sufficiently small. The details of the construction requires the introduction of a third collection of polydiscs B_j of intermediate size which will be referred to as the transportation boxes; see Lemma 13.

4.1. Equidistribution boxes

The goal of this section is to construct the equidistribution boxes mentioned above and prove that they satisfy the properties stated in the lemma below. The mapping $\Xi : \Delta \times \mathbb{T}^n \to M$ in the statement below is as in Eq. (4). By saying that a map *s acts as a permutation by k-cycles* on a set E we mean that s permutes the elements of E, and every orbit of this permutation has cardinality k.

Lemma 12 (Equidistribution Boxes). *Let $q \in \mathcal{Q}$ and $\varepsilon > 0$. There exist $N \in \mathcal{Q}$ and closed subsets $A_1, \dots, A_N$ of $M\backslash \mathrm{Fix}(S)$ such that:*

1. *(Equidistribution Property) The sets A_i have diameters less than ε, their interiors are disjoint, their volumes are equal, their boundaries are of zero volume, and*

$$W := \left(M\backslash \bigcup_i A_i\right) \subset B_\varepsilon(\mathrm{Fix}(S)).$$

2. *$\bigcup_i A_i$ is invariant under the circle action S_t.*
3. *$S_{\frac{1}{q}}$ acts on the set $\{A_1, \dots, A_N\}$ as a permutation by q-cycles.*
4. *(Action-Angle Coordinates) For each A_i, we have*

$$A_i = \Xi(P \times T),$$

where Ξ is the map defined in Eq. (4), P is some polytope included in the moment polytope Δ, and T is a cube in the torus $\mathbb{T}^n$ obtained from a subdivision of $\mathbb{T}^n$ into equal cubes. The map Ξ defines a symplectomorphism from $\mathrm{Int}(P \times T)$ to $\mathrm{Int}(A_i)$.

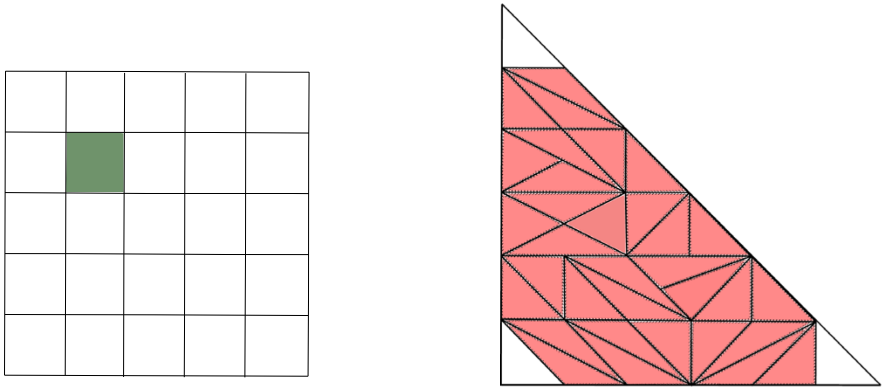


Figure 2. Depiction of the equidistribution boxes A_i in the case of $\mathbb{C}P^2$: On the left, a subdivision of $\mathbb{T}^2$ into small equal cubes T with one such cube colored in green. On the right, a sample subdivision of the moment polytope of $\mathbb{C}P^2$ into small polytopes of equal volume. The collection $\mathcal{P} = \{P\}$ consists of the polytopes colored in pink. The sets A_i have equal volumes

Proof. To construct the A_i's, we begin by subdividing the moment polytope Δ into a collection of polytopes of small diameter and equal volume. By subdividing we mean that two distinct polytopes can only intersect at lower dimensional faces and their union covers Δ. Now, let $\mathcal{P} = \{P\}$ be the set consisting of those polytopes from the subdivision which do not contain any of the vertices of Δ. Let N'' denote the total number of the polytopes in $\mathcal{P}$. The subdivision may be carried in such a way that N'' is relatively prime to q_0 and so $N'' \in \mathcal{Q}$ (see Eq. 1).

We consider a decomposition of the torus, which is invariant under translation by $S_{\frac{1}{q}}$, into equal cubes T. More precisely, we obtain these cubes by subdividing each $\mathbb{S}^1$ factor of $\mathbb{T}^n$ into $N'q$ subintervals of equal length, where N' is picked to be relatively prime to q_0. Hence, the cubes T are all cubes in $\mathbb{T}^n$ of the form

$$v + \left[0, \frac{1}{N'q}\right]^n,$$

where $v \in \frac{1}{N'q}\mathbb{Z}^n$. Note that the total number of cubes T is $(N'q)^n$. The fact that this decomposition of $\mathbb{T}^n$ into the cubes T is invariant under translation by $S_{\frac{1}{q}}$ follows from Eq. (3) (Fig. 2).

Finally, we obtain the A_i's by considering the images under Ξ of all the products $P \times T$ of the polytopes $P \in \mathcal{P}$ and the cubes T. Observe that the total number of the A_i's is $N = N''(N'q)^n$ which belongs to $\mathcal{Q}$. Also note that, since Ξ is a symplectomorphism, the volume of each A_i equals the product of the Haar volume of T with the standard volume of P, and thus they are all equal.

By picking the subdivisions of Δ and $\mathbb{T}^n$ into polytopes and cubes, respectively, to be sufficiently fine we can ensure that the A_i's are of diameter less than ε and that W is contained $B_\varepsilon(\mathrm{Fix}(S))$ (for this it is crucial that our global section σ is defined on Δ and not only on $\mathrm{Int}(\Delta)$). It is not difficult to check, from the construction, that the A_is satisfy all the remaining properties. □

The following observation will be used below. As we recalled in Sect. 3, the inverse images of the faces of the moment polytope are submanifolds of M. Thus, by point 4 of the lemma, each face of each A_i is compactly included in an open submanifold of M.

4.2. Transportation boxes

Let $B' \subset B$ be two polydiscs in M. We say B' is a *sub-polydisc* of B, of the form

$$\prod_{i=1}^{n}[a_i', b_i'] \times [c_i', d_i'] \subset \prod_{i=1}^{n}[a_i, b_i] \times [c_i, d_i],$$

if there exists a symplectic embedding which maps $\Pi_{i=1}^n[a_i, b_i] \times [c_i, d_i]$ to B and $\Pi_{i=1}^n[a_i', b_i'] \times [c_i', d_i']$ to B'. In the statement below, $A_1, \ldots, A_N$ are the equidistribution boxes given by Lemma 12 and $W = M \setminus \cup A_i$; in particular N denotes the number of equidistribution boxes (these data depend on $q \in \mathcal{Q}$ and $\varepsilon > 0$).

Roughly speaking, the aim of the following lemma is to cover most of the interior of the A_i's by disjoint polydiscs B_j' so that

1. The B_j''s are symplectomorphic, and each A_j contains the same number of B_j''s,
2. The collection $\{B_j'\}$ is invariant under $S_{\frac{1}{Nq}}$,
3. There exists a slightly perturbed circle action each of whose orbits spends most of its time inside the union of these polydiscs.

The number ε was used to ensure that the boxes A_i are well distributed in M. We will use a different number, denoted by ε', to ensure that every orbit is well distributed among the A_i's (this will become clear in Proposition 20).

Lemma 13 (Transportation Boxes). *Let $q \in \mathcal{Q}$ and $\varepsilon > 0$ be as in Lemma 12, and let $\varepsilon' > 0$. There exist two families of polydiscs $B_1' \subset B_1, \ldots, B_{N_1}' \subset B_{N_1}$, and a map $\Psi_1 \in \mathrm{Symp}_0(M, \omega)$ such that the following properties are satisfied:*

1. *The polydiscs B_i are all symplectomorphic to the standard polydisc $[-r, r]^{2n}$ for some $r > 0$, have disjoint interiors, each B_i is included in the interior of some A_j, and each A_j contains the same number of B_i's,*
2. *B_i' is a sub-polydisc of B_i of the form $[-r', r']^{2n} \subset [-r, r]^{2n}$, where $r' < r$,*
3. *$S_{\frac{1}{Nq}}$ acts on the B_i's and on the B_i''s as a permutation by Nq-cycles,*
4. *Ψ_1 is C^∞-close to the identity, its support is disjoint from $\mathrm{Fix}(S)$, and it commutes with $S_{\frac{1}{Nq}}$,*

5. *There exists a compact subset K of* $\mathrm{Int}\,(\cup_i B_i')$ *such that every orbit of the conjugated circle action* $\Psi_1 S \Psi_1^{-1}$ *spends more time than* $1-\varepsilon'$ *in* $K \cup W$.

The families of polydiscs B_i, B_i' will be referred to as the *transportation boxes.*

Proof. The construction of Ψ_1 and B_i's will be mostly carried out in the quotient

$$M' = (M \backslash \mathrm{Fix}(S))/S_{\frac{1}{Nq}}.$$

Recall, from the explanation after Eq. (1), that M' is a manifold and, moreover, it naturally inherits the symplectic structure and the Hamiltonian circle action of M. We will denote the symplectic form and the circle action on M' by, respectively, ω' and S_t' where now t belongs to the circle $\mathbb{R} \mod \frac{1}{Nq}$. When dealing with this new circle action, the "time spent" in some set will be with respect to the (non normalised) Lebesgue measure on $\mathbb{R} \mod \frac{1}{Nq}$, with total mass $\frac{1}{Nq}$.

The quotient map $\pi : M\backslash\mathrm{Fix}(S) \to M'$ is a covering map; hence, an element $\psi \in \mathrm{Symp}_0(M', \omega')$ lifts to a symplectic diffeomorphism $\Psi \in \mathrm{Symp}_0(M, \omega)$ which commutes with $S_{\frac{1}{Nq}}$. Furthermore, if ψ is close to the identity then so is Ψ.

Recall that one of the steps in the construction of the equidistribution boxes A_i requires a decomposition of the torus $\mathbb{T}^n$, into cubes T, which is invariant under translation by $S_{\frac{1}{q}}$. By taking a refinement of that decomposition, which is invariant under translation by $S_{\frac{1}{Nq}}$, we obtain smaller boxes $a_1, \dots, a_{N'}$[7] which have the following list of properties:

1. (Action-Angle Coordinates) Each a_j is contained in some A_i. Moreover, a_j is the subset of A_i obtained by restricting the map from the third item in Lemma 12,
$$\begin{aligned} \Xi : P \times T &\to M \\ (s,t) &\mapsto \mathbb{T}^n_t(\sigma(s)), \end{aligned}$$
to $P \times T'$, where T' is one of the cubes arising from the subdivision of the cube T into smaller equal cubes. The map Ξ defines a symplectomorphism from $\mathrm{Int}(P \times T')$ to $\mathrm{Int}(a_i)$.

   Note that each A_i contains the same number of a_j's. Therefore, it is sufficient to prove the statement of Lemma 13 with the A_i's replaced by a_j's.
2. Each box a_j contains at most one point of each orbit under $S_{\frac{1}{Nq}}$, so the projection $\pi : M \to M'$ restricts to a symplectomorphism between a_j and $\pi(a_j)$, for all j. This implies, in particular, that $\mathrm{Int}(\pi(a_j))$ is also symplectomorphic to $\mathrm{Int}(P \times T')$.
3. Note that since W and $\cup_i A_i$ are disjoint and invariant under the circle action, the sets $\pi(W)$ and $\cup_j \pi(a_j)$ are disjoint.

[7]The N' here is not the same as the N' used in the proof of Lemma 12.

We leave it to the reader to check that Lemma 13 follows from the discussion above by setting B_i''s, B_i's, Ψ_1, and K to be, respectively, the lifts of the b_i''s, b_i's, ψ_1, and $\mathcal{K}$ from the claim below.

Claim 14. *There exist two finite collections of polydiscs $b_i' \subset b_i \subset M'$ and $\psi_1 \in \mathrm{Symp}_0(M', \omega')$ such that the following properties are satisfied:*

1. *The polydiscs b_i are all symplectomorphic to the standard polydisc $[-r, r]^{2n}$ for some $r > 0$, have disjoint interiors, each b_i is included in the interior of some $\pi(a_j)$, and each $\pi(a_j)$ contains the same number of b_i's,*
2. *b_i' is a sub-polydisc of b_i of the form $[-r', r']^{2n} \subset [-r, r]^{2n}$, where $r' < r$,*
3. *ψ_1 is C^∞–close to the identity and is compactly supported,*
4. *Every orbit of the conjugated circle action $\psi_1 S' \psi_1^{-1}$ spends more time than $\frac{1}{Nq} - \frac{\varepsilon'}{Nq}$ in $\mathcal{K} \cup \pi(W)$, where $\mathcal{K}$ is a compact subset of $\mathrm{Int}\,(\cup_i b_i')$.*

We will prove the above claim in two steps. We construct the b_i's in the first step, and the b_i''s in the second step. At each step we have to perturb the action, so that after Step 1 no orbit spends too much time near the boundary of the $\pi(a_j)$'s, and after Step 2 no orbit spend too much time near the boundary of the $\pi(b_i)$'s. Figures 3 and 4 represent the two steps of our construction.

Step 1: In the first step of the proof of Claim 14, depicted in Fig. 3, we will construct the polydiscs b_i and a symplectomorphism $\theta_1 \in \mathrm{Symp}_0(M', \omega')$ such that the following properties are satisfied:

1. The polydiscs b_i are all symplectomorphic to the standard polydisc $[-r, r]^{2n}$ for some $r > 0$, have disjoint interiors, each b_i is included in the interior of some $\pi(a_j)$, and each $\pi(a_j)$ contains the same number of b_is,

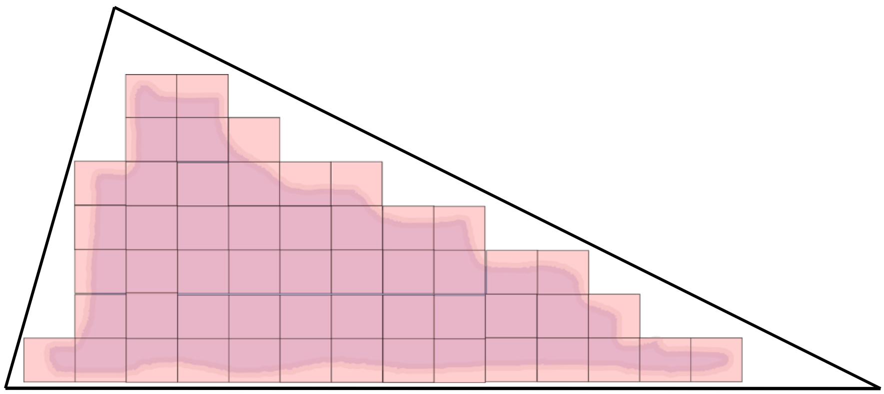


FIGURE 3. Depiction of Step 1 in the construction of transportation boxes: The triangle represents $\pi(a_j)$ for a fixed j. The boxes b_i are in pink and the region $\mathcal{K}_j'$ is shaded purple. The orbits of $\theta_1 S' \theta_1^{-1}$ spend most of their time in $\pi(W) \cup_j \mathcal{K}_j'$

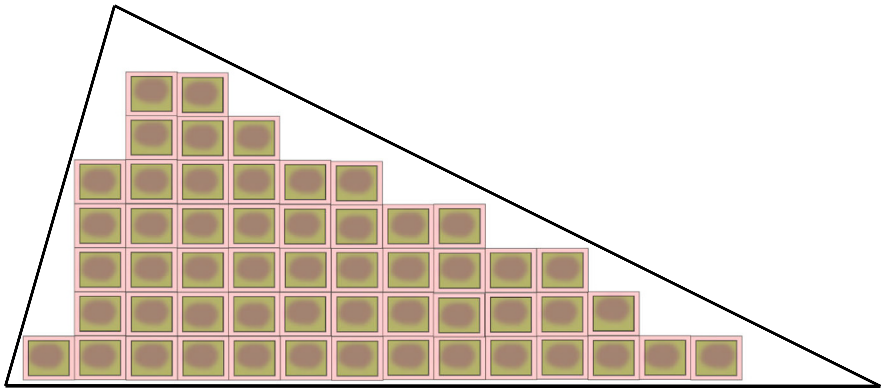


Figure 4. Depiction of Step 2 in the construction of transportation boxes: The triangle represents $\pi(a_j)$ for a fixed j. The boxes $b_i' \subset b_i$ are in green and pink, receptively, and the regions comprising $\mathcal{K}$ are shaded purple. The orbits of $\psi_1 S \psi_1^{-1}$ spend most of their time in $\pi(W) \cup \mathcal{K}$

2. θ_1 is C^∞-close to the identity and is compactly supported,
3. Every orbit of the conjugated circle action $\theta_1 S' \theta_1^{-1}$ spends more time than $\frac{1}{Nq} - \frac{\varepsilon'}{2Nq}$ in $\pi(W) \cup_i b_i$.

Let $\vec{V}$ be the vector field on M' tangent to the circle action. Let Y denote the union of the boundaries of the $\pi(a_j)$'s. We will now make a perturbation of the vector field so that each of its orbits spends little time near Y.

The boundary of each $\pi(a_j)$ is a union of submanifolds with boundary. Let $(Y_k)_{k=0,\ldots,m}$ be a numbering of all these submanifolds, and let $(X_k)_{k=0,\ldots,m}$ denote open hypersurfaces in M such that for each k, X_k contains Y_k. Applying our transversality lemma, Lemma 5, we obtain a C^∞ small and compactly supported symplectomorphism φ_0 of M' such that X_0 is stably almost transverse to $\varphi_{0*}\vec{V}$ on Y_0. We repeatedly apply Lemma 5 to obtain C^∞ small symplectomorphisms $(\varphi_k)_{k=0,\ldots,m}$ of M' such that for each k and each $j \leqslant k$, X_j is stably almost transverse to $(\varphi_k \circ \cdots \circ \varphi_0)_*\vec{V}$ on Y_j. Now, let $\theta_1 = \varphi_m \circ \cdots \circ \varphi_0$. We have that X_k is stably almost transverse to $\theta_{1*}\vec{V}$ on Y_k for each $0 \leqslant k \leqslant m$. Observe that this implies that every orbit of the conjugated action $\theta_1 S' \theta_1^{-1}$ meets $Y = \cup Y_k$ at most finitely many times. According to the Thickening Lemma 6, there exists $\delta > 0$ so that every orbit of $\theta_1 S' \theta_1^{-1}$ spends less time than $\frac{\varepsilon'}{2Nq}$ in the δ-neighborhood O of the set Y.

For each j, let $\mathcal{K}_j' = \pi(a_j) \backslash O$; this is a compact subset of the interior of $\pi(a_j)$. Let $\mathcal{K}' = \cup \mathcal{K}_j'$ and observe that every orbit of $\theta_1 S' \theta_1^{-1}$ spends more time than $\frac{1}{Nq} - \frac{\varepsilon'}{2Nq}$ in $\mathcal{K}' \cup \pi(W)$.

To complete the first step of the proof, it remains to show that we can find symplectomorphic polydiscs b_i with disjoint interiors such that each b_i is contained in some $\pi(a_j)$, each $\pi(a_j)$ contains the same number of b_i's, and $\mathcal{K}' \subset \operatorname{Int}(\cup_i b_i)$.

As mentioned above, there exists a symplectic identification of $\mathrm{Int}(\pi(a_j))$ with the interior of a product of the form $P_j \times T$, where P_j is some polytope in $\mathbb{R}^n$ and T is a cube in the torus $\mathbb{T}^n$. Hence, we may suppose that $\mathcal{K}'_j \subset \mathrm{Int}(P_j \times T)$. We will take care of P_j and T separately. Recall that the polytopes P_j were picked to have equal volumes.

Claim 15. *Let k'_j denote the image of $\mathcal{K}'_j$ under the canonical projection $P_j \times T \to P_j \subset \Delta \subset \mathbb{R}^n$. Then, k'_j may be covered by cubes $e_1, \ldots, e_l$ such that*

1. *Each of the cubes $e_1, \ldots, e_l$ is a translation of the cube $[0, \eta]^n$, for some η,*
2. *The cubes $e_1, \ldots, e_l$ are all contained in the interior of P_j,*
3. *The cubes $e_1, \ldots, e_l$ have disjoint interiors,*
4. *The number of cubes $e_1, \ldots, e_l$ used to cover k'_j does not depend on j.*

To prove the first three items in the above claim one can simply place a grid on $\mathbb{R}^n$ whose edge-size is η, for a sufficiently small η, and select the closed cubes from the grid that are included in the interior of P_j. To get the last item one must use the fact that the P_j's are polytopes of equal volume in $\mathbb{R}^n$. Note that if we make the size of the grid converge to zero, then the volume v_j of the union of the selected cubes converges to the volume of P_j. Thus we may choose a common size η such that all the v_j's are very close to the common volume of the P_j's. If η is small enough then many cubes near the boundary ∂P_j will actually not intersect k'_j. Then we may discard some of those unnecessary cubes, in each P_j, to adjust for the number and get item 4.

Now, we will complete Step 1 of the proof of Claim 14. All the cubes T in $\mathbb{T}^n$ have the same size, so we may identify them with some $T_0 = [0, \eta]^n$. We divide T_0 into very small equal cubes such that the cubes which are contained in the interior of T cover the projection of $\mathcal{K}'_j$ onto T, for every j. The b_i's are obtained by simply taking the products of these cubes with those from Claim 15. As a consequence of Claim 15, each $\pi(a_j)$ contains the same number of the b_i's. Furthermore, it is not difficult to see that there exists some r such that the b_i's are all symplectomorphic to $[-r, r]^{2n}$.

Step 2: In the second step of the proof of Claim 14, depicted in Fig. 4, we construct polydiscs b'_i and $\theta_2 \in \mathrm{Symp}_0(M', \omega')$ such that

1. Each b'_i is a sub-polydisc of b_i of the form $[-r', r']^{2n} \subset [-r, r]^{2n}$, where $r' < r$,
2. θ_2 is C^∞–close to the identity and is compactly supported,
3. Every orbit of the conjugated circle action $\psi_1 S \psi_1^{-1}$, where $\psi_1 = \theta_2 \circ \theta_1$, spends more time than $\frac{1}{Nq} - \frac{\varepsilon'}{Nq}$ in $\mathcal{K} \cup \pi(W)$, where $\mathcal{K}$ is a compact subset of $\mathrm{Int}\,(\cup_i b'_i)$.

It is clear that the proof of Claim 14 will be completed once b'_i's and θ_2, satisfying the above properties, are constructed.

Recall that every orbit of $\theta_1 S' \theta_1^{-1}$ spends more time than $\frac{1}{Nq} - \frac{\varepsilon'}{2Nq}$ in $\pi(W) \cup_i b_i$. We can find a C^∞-small symplectomorphism θ_2 such that every orbit of the conjugated action $\theta_2 \theta_1 S' \theta_1^{-1} \theta_2^{-1}$ spends

- more time than $1 - \frac{\varepsilon'}{Nq}$ in $\pi(W) \cup_i b_i$ and,
- less time than $\frac{\varepsilon'}{Nq}$ in some small open neighborhood, say O', of the union of the boundaries of the polydisc b_i.

The construction of θ_2 is very similar to that of θ_1 from Step 1 and so it will be omitted. We let $\psi_1 = \theta_2\theta_1$. For each i, let $\mathcal{K}_i = b_i \backslash O'$; this is a compact subset of the interior of b_i. One can check that every orbit of $\psi_1 S \psi_1^{-1}$ spends more time than $\frac{1}{Nq} - \frac{\varepsilon'}{Nq}$ in $\pi(W) \cup_i \mathcal{K}_i$. Let $\mathcal{K} = \cup_i \mathcal{K}_i$.

Recall that each b_i is symplectomorphic to $[-r, r]^{2n}$. Since the $\mathcal{K}_i$'s are compact subsets of the b_i's we can find $r' < r$ such that $\mathcal{K}_i \subset \operatorname{Int}(b_i')$, where b_i' is the sub-polydisc of b_i of the form $[-r', r']^{2n} \subset [-r, r]^{2n}$. Hence, we have established that $\mathcal{K}$ is a compact subset of $\operatorname{Int}(\cup_i b_i')$ and every orbit of $\psi_1 S \psi_1^{-1}$ spends more time than $\frac{1}{Nq} - \frac{\varepsilon'}{Nq}$ in $\pi(W) \cup \mathcal{K}$. This completes Step 2, and hence the entirety of the proof of Claim 14. □

4.3. Small boxes

We continue to work in the settings of Sects. 4.1 and 4.2: We have the equidistribution boxes A_i, provided by Lemma 12, the transportation boxes B_i, B_i', from Lemma 13, and lastly $\Psi_1 \in \operatorname{Symp}_0(M)$ as described in Lemma 13.

Lemma 16 [Small Boxes]. *Let $q \in \mathcal{Q}$ and $\varepsilon, \varepsilon' > 0$ be as in the previous lemmas. There exist a finite collection of polydiscs $\{c_k\}$ in M and $\Psi \in \operatorname{Symp}_0(M, \omega)$ such that the following properties are satisfied:*

1. *The c_k's are disjoint, each c_k is included in some transportation box $B_i' = B'(c_k)$, and each transportation box contains the same number of c_k's.*
2. *$S_{\frac{1}{Nq}}$ acts on the c_ks as a permutation by Nq-cycles.*
3. *Ψ is C^∞-close to the identity, its support is disjoint from $\operatorname{Fix}(S)$, and it commutes with $S_{\frac{1}{Nq}}$.*
4. *Every orbit of the circle action $\Psi S \Psi^{-1}$ spends more time than $1 - \varepsilon'$ in $W \cup (\cup_k c_k)$.*
5. *(Transport) For any small box c_k, let*
$$\mathcal{O}(c_k) = \{S_{\frac{j}{q}}(c_k) \mid j \in \{0, \ldots, q-1\}\}$$
be the orbit of c_k under the action of $S_{\frac{1}{q}}$. For any two small boxes c_{k_1}, c_{k_2}, there exists $\Phi_{k_1k_2}$ a compactly supported symplectomorphism of $M \backslash W$ which commutes with $S_{\frac{1}{q}}$ and has the following properties:
   (a) *$\Phi_{k_1k_2}$ acts as a permutation on the set of all c_k's,*
   (b) *$\Phi_{k_1k_2}(\mathcal{O}(c_{k_1})) = \mathcal{O}(c_{k_2})$ and $\Phi_{k_1k_2}(\mathcal{O}(c_{k_2})) = \mathcal{O}(c_{k_1})$,*
   (c) *$\Phi_{k_1k_2}(c_k) \subset B'(c_k)$ for each small box $c_k \notin \mathcal{O}(c_{k_1}) \cup \mathcal{O}(c_{k_2})$.*

Proof. **Preparation for the construction of c_k's:**
Recall that B_i' is a sub-polydisc of B_i of the form $[-r', r']^{2n} \subset [-r, r]^{2n}$. Let V_i be the polydisc in $B_i \backslash \operatorname{Int}(B_i')$ of the form $[r', r' + 2\eta] \times [-\eta, \eta]^{2n-1}$, where $2\eta < r - r'$. Note that the collection of V_i's is invariant under the action of $S_{\frac{1}{Nq}}$.

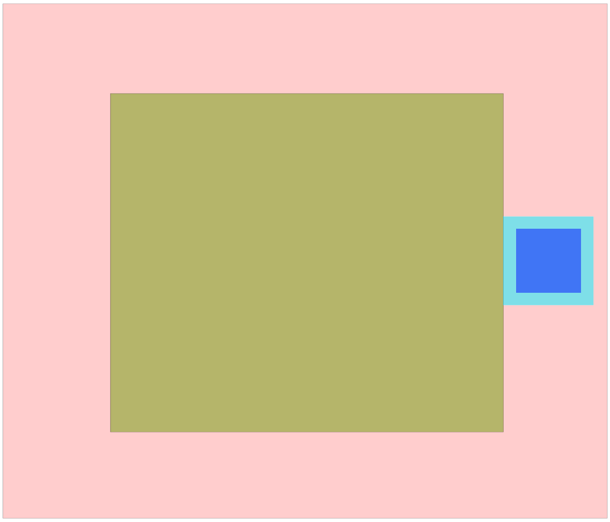


FIGURE 5. As before, $B_i' \subset B_i$ are in pink and green, respectively. The set V_i is in light blue and $U_i(\delta)$ in dark blue

For each $0 < \delta < \eta$, let $U_i(\delta)$ be the sub-polydisc in V_i corresponding to $[r' + \eta - \delta, r' + \eta + \delta] \times [-\delta, \delta]^{2n-1}$; see Fig. 5. Note that the collection of $U_i(\delta)$'s is also invariant under the action of $S_{\frac{1}{Nq}}$.

Claim 17. *There exists $\delta_0 > 0$ with the following property: Take any $\delta \leqslant \delta_0$ and consider any two V_{i_1}, V_{i_2} which have disjoint orbits under the action of $S_{\frac{1}{q}}$, i.e. $S_{\frac{j}{q}}(V_{i_1}) \neq V_{i_2}$ for any $j \in \{0, \dots, q-1\}$. Then, there exists a compactly supported symplectomorphism $\Upsilon_{i_1 i_2} \in \mathrm{Symp}_0(M \backslash (W \cup (\cup_i B_i')))$ such that*

1. $\Upsilon_{i_1 i_2}(U_{i_1}(\delta)) = U_{i_2}(\delta)$ *and* $\Upsilon_{i_1 i_2}(U_{i_2}(\delta)) = U_{i_1}(\delta)$,
2. $\Upsilon_{i_1 i_2} S_{\frac{1}{q}} = S_{\frac{1}{q}} \Upsilon_{i_1 i_2}$.

Proof of Claim 17. Observe that $S_{\frac{1}{q}}$ acts on $M \backslash (W \cup (\cup_i B_i'))$; this is because W is invariant under the circle action and $\cup_i B_i'$s is invariant under the action of $S_{\frac{1}{Nq}}$. Furthermore, this action is free. Therefore, we may consider the quotient symplectic manifold $M'' = M \backslash (W \cup (\cup_i B_i'))/S_{\frac{1}{q}}$. Let $\pi : M \backslash (W \cup (\cup_i B_i')) \to M''$ denote the quotient map.

We leave it to the reader to check that proving Claim 17 may be reduced to proving the following statement on M''. There exists $\delta_0 > 0$ with the property that for every $\delta < \delta_0$ and any two i_1, i_2, such that $\pi(\mathrm{Int}(V_{i_1})) \neq \pi(\mathrm{Int}(V_{i_2}))$, we can find a compactly supported symplectomorphism $\Upsilon_{i_1 i_2}$ of M'' such that $\Upsilon_{i_1 i_2}(\pi(U_{i_1}(\delta))) = \pi(U_{i_2}(\delta))$ and $\Upsilon_{i_1 i_2}(\pi(U_{i_2}(\delta))) = \pi(U_{i_1}(\delta))$.

The fact $\Upsilon_{i_1 i_2}$ of the previous paragraph exists for small enough values of δ is a consequence of Lemma 8 applied in the symplectic manifold M''.

The map $\Upsilon_{i_1 i_2}$ is depicted by the dotted line in Fig. 7. ☐

Construction of c_k's: Recall that B'_i is symplectomorphic to$[-r', r']^{2n}$. We subdivide each B'_i into polydiscs all of which are symplectomorphic to $[-\delta', \delta']^{2n}$, for some δ' smaller than the δ_0 given by the previous claim; we will denote the collection of these polydiscs by $\{c'_k\}$. Each B'_i contains the same number of c'_ks and the c'_ks have disjoint interiors. Since the collection of B'_is is invariant under the action of $S_{\frac{1}{Nq}}$, we can ensure that the collection of c'_k's is also invariant under the action of $S_{\frac{1}{Nq}}$.

Recall from Lemma 13 that every orbit of the conjugated action $\Psi_1 S \Psi_1^{-1}$ spends more time than $1-\varepsilon'$ inside $K \cup W$, where K is a compact subset of Int $(\cup_i B'_i)$. As was done in the proof of of Lemma 13, using transversality Lemma 5 and Thickening Lemma 6, we can find Ψ_2 such that every orbit of the conjugated action $\Psi_2\Psi_1 S \Psi_1^{-1}\Psi_2^{-1}$ spends more time than $1-\varepsilon'$ in $W \cup_k c_k$ where c_k is the sub-polydisc of c'_k of the form $[-\delta, \delta]^{2n} \subset [-\delta', \delta']^{2n}$, with δ

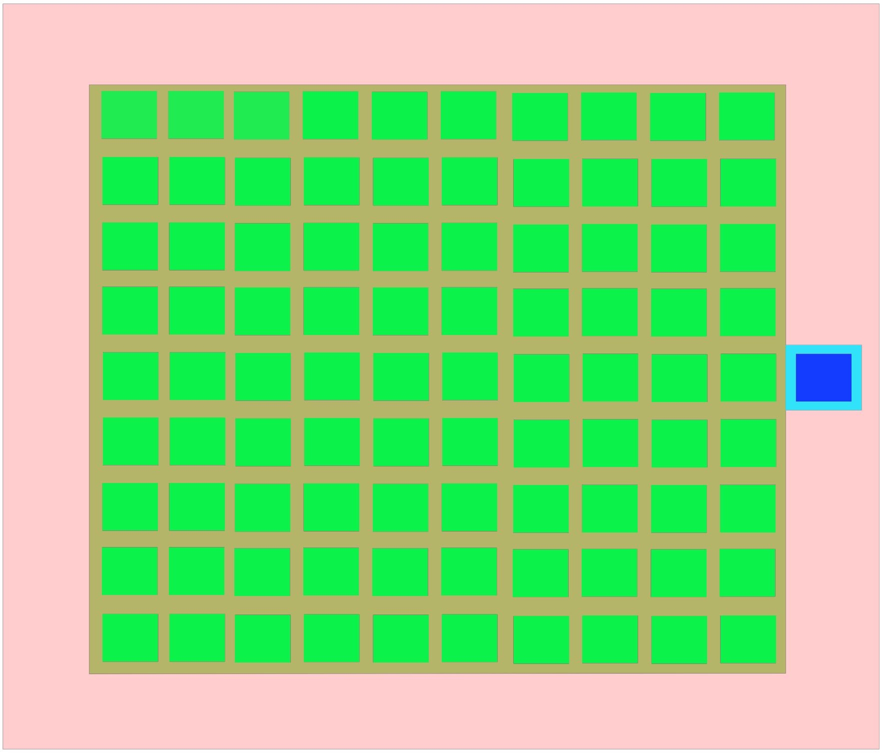


FIGURE 6. The small boxes c_k, in bright green, are added to the previous picture; they are symplectomorphic to $U(c_k)$ depicted in dark blue. Every orbit of $\Psi S \Psi^{-1}$ spends more time than $1-\varepsilon'$ in $W \cup_k c_k$.
Any two small boxes c_{k_1}, c_{k_2}, in bright green, can be exchanged via swapping several adjacent boxes. Similarly, $U(c_k)$ can be swapped with the small box adjacent to it

being slightly smaller than δ'. Furthermore, the map Ψ_2 can be picked such that it commutes with $S_{\frac{1}{Nq}}$. We will not describe the construction of Ψ_2 as it is very similar to that of the maps Ψ_1 and ψ_1 from Lemma 13; see in particular the first paragraph of the proof of Claim 14. The symplectomorphism Ψ is the composition $\Psi_2 \circ \Psi_1$. Note that Ψ and the c_k's satisfy items 1–4 of the lemma. It remains to construct the maps $\Phi_{k_1k_2}$ required by the last item. See Fig. 6 for a depiction of the small boxes c_k.

Proof of the last item (Transport):

We will need to introduce a bit of notation for the remainder of the proof. For any small box c_k, there exists (unique) i_k such that $c_k \subset B'_{i_k} \subset B_{i_k}$. We will denote B'_{i_k}, B_{i_k} by $B'(c_k), B(c_k)$, respectively. Similarly, we will denote $U_{i_k}(\delta)$ and V_{i_k} by $U(c_k)$ and $V(c_k)$, respectively. Observe that $U(c_k)$ and c_k are both polydiscs symplectomorphic to $[-\delta, \delta]^{2n}$.

Roughly speaking, our goal here is to find a symplectomorphism $\Phi_{k_1k_2}$ which exchanges two given small boxes c_{k_1} and c_{k_2} while leaving the remaining small boxes more or less untouched. Of course, the remaining boxes cannot be left entirely untouched because $\Phi_{k_1k_2}$ must commute with $S_{\frac{1}{q}}$ and so we are automatically forced to swap $\mathcal{O}(c_{k_1})$ and $\mathcal{O}(c_{k_2})$. However, the exchange may be achieved such that the remaining c_k's do not leave their transportation boxes. The rough idea of the construction is as follows: first, any two small boxes c_{k_1}, c_{k_2} which are in the same transportation box $B'(c_k)$ can be swapped; the remaining small boxes in $B'(c_k)$ might be affected, however, they will remain in $B'(c_k)$; this is the content of Claim 18 and its proof is rather evident in dimension two, see Fig. 7, and the argument generalizes to higher dimensions. Second, a similar reasoning can be used to obtain a symplectomorphism which swaps $U(c_k)$ with the small box adjacent to it; see Fig. 7 and Claim 19. Third, we can exchange $U(c_{k_1})$ and $U(c_{k_2})$; see Fig. 7 and Claim 17. Finally, we obtain $\Phi_{k_1k_2}$ by combining the above facts.

We will be needing the following two claims.

Claim 18. *Fix a small box c_k and let c_{k_1}, c_{k_2} be any two small boxes which are contained in $B(c_k)$. Then, there exists $\Phi_{k_1k_2} \in \mathrm{Symp}_0(M, \omega)$ with the following property:*

1. $\Phi_{k_1k_2}$ *is supported in* $\cup_j S_{\frac{j}{q}}(B'(c_k)) \subset \cup_j S_{\frac{j}{q}}(B(c_k)) \subset M \backslash W$,
2. $\Phi_{k_1k_2}$ *commutes with* $S_{\frac{1}{q}}$,
3. $\Phi_{k_1k_2}(c_{k_1}) = c_{k_2}$ *and* $\Phi_{k_1k_2}(c_{k_2}) = c_{k_1}$,
4. $\Phi_{k_1k_2}$ *acts as a permutation on the set* $\{c_{k'} : c_{k'} \subset B(c_k)\}$.

Proof of Claim 18. Since the set of B'_i's is invariant under the action of $S_{\frac{1}{q}}$, it is enough to prove the following: There exists a symplectomorphism $\phi_{k_1k_2}$ which is compactly supported in $B'(c_k)$, is isotopic to the identity, and has the following properties:

1. $\phi_{k_1k_2}(c_{k_1}) = c_{k_2}$ and $\phi_{k_1k_2}(c_{k_2}) = c_{k_1}$,
2. $\phi_{k_1k_2}$ acts as a permutation on the set $\{c_{k'} : c_{k'} \subset B(c_k)\}$.

We leave it to the reader to check that the existence of $\phi_{k_1k_2}$ can be deduced from Lemma 9. The map $\phi_{k_1k_2}$ corresponds to a symplectomorphism

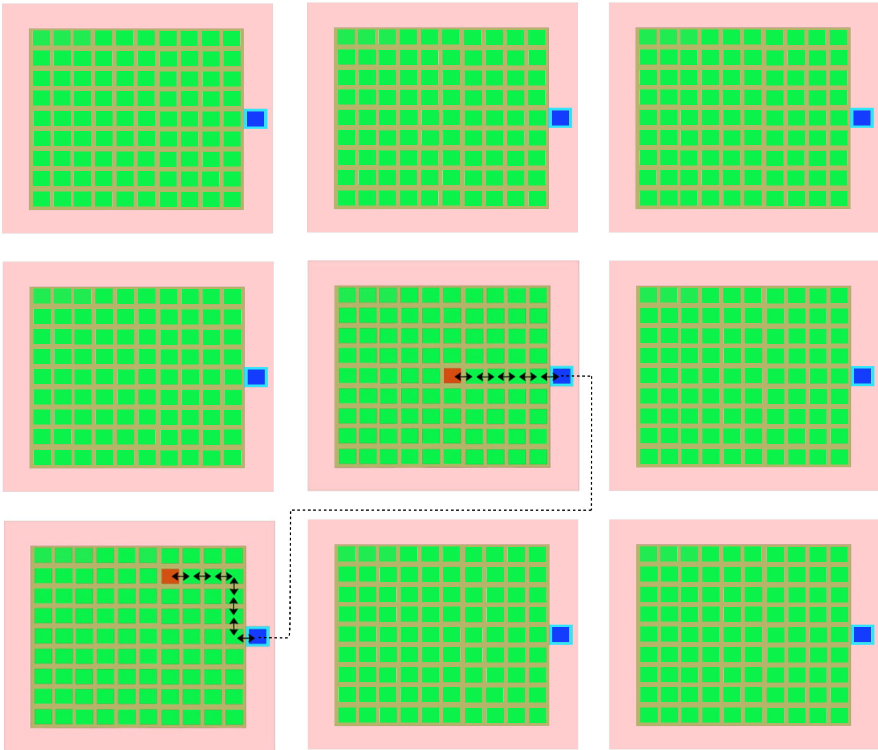


FIGURE 7. Depiction of $\Phi_{k_1k_2}$ exchanging c_{k_1} and c_{k_2} which are colored in red. First, via swapping adjacent small boxes we map c_{k_1} and c_{k_2} to $U(c_{k_1})$ and $U(c_{k_2})$, respectively. The swaps are depicted by the two-headed arrows. We then exchange $U(c_{k_1})$ and $U(c_{k_2})$ using the space in between the transportation boxes; the dotted line depicts a path taken by a symplectic isotopy swapping $U(c_{k_1})$ and $U(c_{k_2})$. The existence of such isotopy is guaranteed by Claim 17

which, within the same transportation box $B(c_k)$, exchanges two bright green squares in either of Figs. 6 or 7. This can be achieved via a composition of symplectomorphisms which swap adjacent squares as depicted in Fig. 7. □

Claim 19. *For each small box c_k, there exists $\Theta_k \in \mathrm{Symp}_0(M,\omega)$ with the following properties:*

1. Θ_k *is supported in* $\cup_j S_{\frac{j}{q}}(B(c_k)) \subset M\backslash W$,
2. Θ_k *commutes with* $S_{\frac{1}{q}}$,
3. $\Theta_k(c_k) = U(c_k)$,
4. Θ_k *acts as a permutation on the collection of polydiscs* $\{U(c_k)\} \cup \{c_{k'} : c_{k'} \subset B(c_k)\}$.

Proof of Claim 19. Since the set of B_is is invariant under the action of $S_{\frac{1}{q}}$, it is enough to prove the following: For each small box c_k, there exists a symplectomorphism θ_k which is compactly supported in $B(c_k)$, is isotopic to the identity, with the following properties:

1. $\theta_k(c_k) = U(c_k)$,
2. θ_k acts as a permutation on the collection of polydiscs $\{U(c_k)\} \cup \{c_{k'} : c_{k'} \subset B(c_k)\}$.

We leave it to the reader to check that the existence of θ_k can be deduced from Lemma 9. In Fig. 7, θ_k corresponds to a symplectomorphism which exchanges the dark blue square with the red square in the adjacent transportation box; it is obtained as a composition of maps that permute adjacent squares, as indicated by the two-headed arrows. □

We now prove the last item (Transport) in the statement of Lemma 16 using Claims 17, 18, 19.

Note that it is sufficient to prove the statement up to replacing c_{k_1} or c_{k_2} with any element of $\mathcal{O}(c_{k_1})$ or $\mathcal{O}(c_{k_2})$, respectively. First, we consider the simpler case where there exists j such that $S_{\frac{j}{q}}(c_{k_2}) \subset B(c_{k_1})$. Then, up to replacing c_{k_2} with $S_{\frac{j}{q}}(c_{k_2})$, we may assume that $B(c_{k_1}) = B(c_{k_2})$. In this case, $\Phi_{k_1k_2}$ is given by Claim 18.

Next, we treat the case where $B(c_{k_1}) \neq B(c_{k_2})$ even up to replacing c_{k_1}, c_{k_2} with elements of $\mathcal{O}(c_{k_1}), \mathcal{O}(c_{k_2})$; Fig. 7 depicts $\Phi_{k_1k_2}$ in this scenario. Let $\Theta_{k_1}, \Theta_{k_2}$ be as given by Claim 19; note that these two maps have disjoint supports. Then, define

$$\Phi_{k_1k_2} = \Theta_{k_1}^{-1}\Theta_{k_2}^{-1}\Upsilon\Theta_{k_2}\Theta_{k_1},$$

where $\Upsilon \in \mathrm{Symp}_0(M\backslash W)$ commutes with $S_{\frac{1}{q}}$ and satisfies $\Upsilon(U(c_{k_1})) = U(c_{k_2})$ and $\Upsilon(U(c_{k_2})) = U(c_{k_1})$; the existence of Υ is guaranteed by Claim 17 and is depicted by the dotted line in Fig. 7. We leave it to the reader to check that $\Phi_{k_1k_2}$ satisfies all the requirements of the last item of Lemma 16. □

4.4. From boxes to proposition 4

Having proven the lemmas of Sections 4.1, 4.2, and 4.3, we are now well positioned to prove the following proposition which in turn will entail Proposition 4.

Proposition 20. *For any $q \in \mathcal{Q}$ and any $\varepsilon, \varepsilon' > 0$, there exist $h \in \mathrm{Symp}_0(M, \omega)$ and $A_1, \ldots, A_N$ closed subsets of M such that:*

1. *The sets A_i satisfy the equidistribution Property from Lemma 12,*
2. *The support of h is disjoint from $\mathrm{Fix}(S)$ and $hS_{\frac{1}{q}} = S_{\frac{1}{q}}h$,*
3. *Every orbit of the conjugated action hSh^{-1} is almost equidistributed among the sets A_i in the following sense: There exists $E \subset M$ such that for each $x \in M$ we have $\mathrm{Leb}_{\mathbb{S}^1}(\{t \in \mathbb{S}^1 : hS_th^{-1}(x) \in E\}) < \varepsilon'$ and the following properties are satisfied:*
   (a) *$E \subset \cup_i A_i$ and $\partial A_i \subset E$ for every i,*
   (b) *For each $x \in M$, let $I_i(x) := \{t \in \mathbb{S}^1 : hS_th^{-1}(x) \in A_i\backslash E\}$, Then, $\mathrm{Leb}(I_i(x)) = \mathrm{Leb}(I_j(x))$ for all i, j.*

Proof of proposition 20. We will be applying the lemmas of the previous sections with the given q, ε and ε'. Lemma 12 gives us the equidistribution boxes $A_1, \ldots, A_N$.

Applying Lemma 16 we obtain small boxes $\{c_k\}$ and $\Psi \in \mathrm{Symp}_0(M, \omega)$ satisfying, among others, the following properties:

1. Each A_i contains the same number of small boxes c_k,
2. $S_{\frac{1}{Nq}}$ acts by a cyclic permutation of order Nq on the c_k's.
3. Ψ is C^∞-close to the identity, its support is disjoint from $\mathrm{Fix}(S)$, and it commutes with $S_{\frac{1}{Nq}}$,
4. Every orbit of the circle action $\Psi S \Psi^{-1}$ spends more time than $1 - \varepsilon'$ in $W \cup_k c_k$.

Given a small box c, we will denote by

$$\mathcal{O}_q(c) = \{S_{\frac{j}{q}}(c) \mid j = 0, \ldots, q-1\}, \quad \mathcal{O}_{Nq}(c) = \{S_{\frac{j}{Nq}}(c) \mid j = 0, \ldots, Nq-1\}$$

the orbits of c respectively under the actions of $S_{\frac{1}{q}}$ and $S_{\frac{1}{Nq}}$. Using Lemma 16, we can prove the following claim.

Claim 21. *There exists $\Theta \in \mathrm{Symp}_0(M)$ which is compactly supported in $M \backslash W$ such that*

1. $\Theta S_{\frac{1}{q}} = S_{\frac{1}{q}} \Theta$,
2. *For any small box c, the interior of each equidistribution box A contains exactly q of the elements of the set*

$$\Theta(\mathcal{O}_{Nq}(c)) = \{\Theta(c), \Theta(S_{\frac{1}{Nq}}(c)), \ldots, \Theta(S_{\frac{Nq-1}{Nq}}(c))\}.$$

Proof of Claim 21. We begin by explaining the main idea of the proof of this before proceeding to the give the details of the proof.

Given a small box c, and any symplectomorphism Θ, we will say $\Theta(\mathcal{O}_{Nq}(c))$ is equidistributed if each equidistribution box A contains exactly q of its elements. Note that if $\mathcal{O}_{Nq}(c))$ is equidistributed for every c, then we are done with $\Theta = \mathrm{Id}$. If not we can find transportation boxes say A_1, A_2, and a small box c_{k_1} contained in A_1 such that A_1 contains more than q of the elements of $\mathcal{O}_{Nq}(c_{k_1})$ and A_2 contains less than q of them. Since each A_i contains exactly the same number of small boxes, we can find a small box, which we denote by c_{k_2}, such that A_2 contains more than q of the elements of $\mathcal{O}_{Nq}(c_{k_2})$. By the transport item of Lemma 16, there exists $\Phi_{k_1 k_2}$, a compactly supported symplectomorphism of $M \backslash W$, which commutes with $S_{\frac{1}{q}}$ and has the following property: it exchanges the orbits $\mathcal{O}_q(c_{k_1})$ and $\mathcal{O}_q(c_{k_2})$. As for the other small boxes, $\Phi_{k_1 k_2}$ leaves them nearly unchanged in the sense that c and $\Phi_{k_1 k_2}(c)$ remain in the same equidistribution box. We see that after applying $\Phi_{k_1 k_2}$, A_1 will contain one less of the elements of $\mathcal{O}_{Nq}(c_{k_1})$ and A_2 will contain one more. Repeating this process will allow us to construct the map Θ as the compositions of all such $\Phi_{k_1 k_2}$'s.

We will now proceed to give more details of the proof. As will be explained below, we will successively construct, for $k = 1, \ldots,$ symplectomorphisms Θ_k which are compactly supported in $M \backslash W$, commute with $S_{\frac{1}{q}}$, act as a permutation on the collection of small boxes and such that

$\Theta_k \circ \cdots \circ \Theta_1(\mathcal{O}_{Nq}(c_k))$ is equidistributed. Furthermore, each Θ_k will have the following additional property: For each small box c denote by $A(c)$ the equidistribution box which contains it. Then, for all $1 \leqslant i \leqslant k-1$ and every $c \in \mathcal{O}_{Nq}(c_i)$ we have

$$A(\Theta_k \circ \Theta_i \cdots \circ \Theta_1(c)) = A(\Theta_i \cdots \circ \Theta_1(c)).$$

This implies, in particular, that $\Theta_k \circ \cdots \circ \Theta_1(\mathcal{O}_{Nq}(c_i))$ is equidistributed for all $1 \leqslant i \leqslant k$. Once Θ_k with such properties is constructed we can simply set Θ to be the composition of all the Θ_k's.

Leaving the case where $k = 1$ to the reader, we will now describe the construction of Θ_k, assuming $\Theta_1, \ldots, \Theta_{k-1}$ have been constructed. If $\Theta_{k-1} \circ \cdots \circ \Theta_1(\mathcal{O}_{Nq}(c_k))$ is equidistributed we set $\Theta_k = \mathrm{Id}$. If not, we can find two equidistribution boxes say A_1, A_2, such that A_1 contains more than q of the elements of $\Theta_{k-1} \circ \cdots \circ \Theta_1(\mathcal{O}_{Nq}(c_k))$ and A_2 contains less than q of them. By induction we know that A_2 contains exactly q of the elements of $\Theta_{k-1} \circ \cdots \circ \Theta_1(\mathcal{O}_{Nq}(c_{k'}))$ for $k' < k$, and moreover each A_i contains exactly the same number of small boxes. Thus, there exists $k' > k$ such that A_2 contains more than q of the elements of $\Theta_{k-1} \circ \cdots \circ \Theta_1(\mathcal{O}_{Nq}(c_{k'}))$ (in the sequel we will just need one of these elements).

Let c_{k_1}, c_{k_2} denote $\Theta_{k-1} \circ \cdots \circ \Theta_1(c_k)$ and $\Theta_{k-1} \circ \cdots \circ \Theta_1(c_{k'})$, respectively. By the transport item of Lemma 16, there exists $\Phi_{k_1 k_2}$ a compactly supported symplectomorphism of $M \backslash W$ which commutes with $S_{\frac{1}{q}}$ and satisfies properties (a), (b), (c) of Lemma 16.

We leave it to the reader to check that property (c) has the following consequence: for all $1 \leqslant i \leqslant k-1$ and every $c \in \mathcal{O}_{Nq}(c_i)$ we have

$$A(\Phi_{k_1 k_2} \circ \Theta_i \cdots \circ \Theta_1(c)) = A(\Theta_i \circ \cdots \circ \Theta_1(c)).$$

This implies, in particular, that $\Phi_{k_1 k_2} \circ \cdots \circ \Theta_1(\mathcal{O}_{Nq}(c_i))$ is equidistributed for all $1 \leqslant i \leqslant k-1$.

By property (b), the number of elements of $\Phi_{k_1 k_2} \circ \Theta_{k-1} \circ \cdots \circ \Theta_1(\mathcal{O}_{Nq}(c))$ which are contained in A_1 is one less than the number of elements of $\Theta_{k-1} \circ \cdots \circ \Theta_1(\mathcal{O}_{Nq}(c))$ which are contained in A_1. It follows that by repeatedly applying the transport item of Lemma 16, we can continue the above process to obtain other $\Phi_{k_i k_j}$'s the composition of all of which gives the map Φ_k. $\square$

We will now show that Claim 21 implies Proposition 20. Indeed, let $h = \Theta\Psi$ and consider the conjugated circle action hSh^{-1}. It is clear that the first two items in the statement of the proposition hold. We must prove the third item. We define the set $E := \cup A_i \backslash \Theta(\cup_k c_k)$. It is clear that $E \subset \cup A_i$ and $\partial A_i \subset E$ for each i.

Observe that, by point 4 of Lemma 16, every orbit of the circle action hSh^{-1} spends more time than $1 - \varepsilon'$ in $\Theta(W \cup_k c_k) = W \cup \Theta(\cup_k c_k)$. Hence, we immediately obtain $\mathrm{Leb}_{\mathbb{S}^1}(\{t \in \mathbb{S}^1 : hS_t h^{-1}(x) \in E\}) < \varepsilon'$ for every x.

It remains to show that $\mathrm{Leb}(I_i(x)) = \mathrm{Leb}(I_j(x))$ for all i, j. This is equivalent to showing that the quantity

$$\mathrm{Leb}(\{t \in \mathbb{S}^1 : hS_t h^{-1}(x) \in A_i \cap \Theta(\cup_k c_k)\})$$

does not depend on i. Now, using the action of $S_{\frac{1}{Nq}}$ on the c_ks, we see that this quantity equals

$$\sum_{\mathcal{O}_{Nq}(c)} \mathrm{Leb}(\{t \in \mathbb{S}^1 : hS_th^{-1}(x) \in A_i \cap \Theta\left(\mathcal{O}_{Nq}(c)\right)\}),$$

where the sum is taken over distinct $\mathcal{O}_{Nq}(c)$s. Hence, it is sufficient to show that for any small box c the quantity

$$\mathrm{Leb}(\ \{t \in \mathbb{S}^1 : hS_th^{-1}(x) \in A_i \cap \Theta\left(\mathcal{O}_{Nq}(c)\right)\}\ )$$

does not depend on i. By Claim 21, there exists q elements, say $c_1, \ldots, c_q \in \mathcal{O}_{Nq}(c)$ such that $A_i \cap \Theta\left(\mathcal{O}_{Nq}(c)\right) = \Theta(c_1) \cup \ldots \cup \Theta(c_q)$. Thus,

$$\begin{aligned}&\mathrm{Leb}(\{t \in \mathbb{S}^1 : hS_th^{-1}(x) \in A_i \cap \Theta\left(\mathcal{O}_{Nq}(c)\right)\})\\ &= \sum_{j=1}^{q} \mathrm{Leb}(\{t \in \mathbb{S}^1 : hS_th^{-1}(x) \in \Theta\left(c_j\right)\}).\end{aligned}$$

Now, recall that $h = \Theta\Psi$ and so $\mathrm{Leb}(\{t \in \mathbb{S}^1 : hS_th^{-1}(x) \in \Theta\left(c_j\right)\})$ coincides with $\mathrm{Leb}(\{t \in \mathbb{S}^1 : \Psi S_t\Psi^{-1}(z) \in c_j\})$, where $z = \Theta^{-1}(x)$. Lastly, because Ψ commutes with $S_{\frac{1}{Nq}}$ and $c_j \in \mathcal{O}_{Nq}(c)$, we have that $\mathrm{Leb}(\{t \in \mathbb{S}^1 : \Psi S_t\Psi^{-1}(z) \in c_j\}) = \mathrm{Leb}(\{t \in \mathbb{S}^1 : \Psi S_t\Psi^{-1}(z) \in c\})$. Hence,

$$\begin{aligned}&\mathrm{Leb}(\{t \in \mathbb{S}^1 : hS_th^{-1}(x) \in A_i \cap \Theta\left(\mathcal{O}_{Nq}(c)\right)\})\\ &= q\,\mathrm{Leb}(\{t \in \mathbb{S}^1 : \Psi S_t\Psi^{-1}(z) \in c\}),\end{aligned}$$

which clearly does not depend on i; the above equality follows from the second item of Claim 21. This finishes the proof of Proposition 20. □

It remains to explain why Proposition 4 follows from Proposition 20.

Proof of proposition 4. Let $\mathcal{U}$ denote an open neighborhood of $\mathrm{Conv}(\mathcal{E})$. Clearly, the symplectomorphism h, given to us by Proposition 20, satisfies the first two items of Proposition 4. It remains to prove the third item, that is, if ε and ε' are small enough, for every $x \in M$, the push-forward of $\mathrm{Leb}_{\mathbb{S}^1}$, the Lebesgue measure on the circle, under the map $t \mapsto hS_th^{-1}(x)$, belongs to $\mathcal{U}$.

Fix $x \in M$ and let μ be the push-forward of $\mathrm{Leb}_{\mathbb{S}^1}$ under the map $t \mapsto hS_th^{-1}(x)$. Fix $\varepsilon > 0$ and let $\varepsilon' = \frac{\varepsilon}{2N}$. We leave it to the reader to check that, as a consequence of the third item of Proposition 20, μ has the property that

$$\sum_{i=1}^{N} |\mu(A_i) - \frac{\alpha}{N}| < (N+1)\varepsilon' < \varepsilon, \tag{5}$$

where $\alpha = \mu(\cup_i A_i)$. We will show that any probability measure which satisfies the above property for sufficiently small $\varepsilon > 0$ belongs to $\mathcal{U}$.

Let $\nu \in \mathcal{P}(M)$ and recall that a basis of open neighborhoods of ν for the week topology is given by the collection of sets of the the form

$$U_{\delta, f_1, \ldots, f_k}(\nu) := \left\{\beta \in \mathcal{P}(M) : \left|\int f\, d\nu - \int f\, d\beta\right| < \delta,\ \forall f \in \{f_1, \ldots, f_k\}\right\},$$

where $\delta > 0$ is a real number and $f_1, \ldots, f_k$ are continuous functions on M.

Claim 22. *There exists $\delta > 0$ and continuous functions $f_1, \dots, f_k : M \to \mathbb{R}$, satisfying $\|f_i\|_\infty \leqslant 1$ for each i, such that for every $\nu \in \mathrm{Conv}(\mathcal{E})$ we have*

$$U_{\delta, f_1, \dots, f_k}(\nu) \subset \mathcal{U}.$$

Proof. The space of probability measures on M is metrizable; more precisely, there exists a metric d such that for every radius $r > 0$, there exists $\delta > 0$ and continuous functions $f_1, \dots, f_k : M \to \mathbb{R}$ such that for every probability measure ν, the set $U_{\delta, f_1, \dots, f_k}(\nu)$ is contained in the ball of radius r around ν (see [34], Theorem 6.4). Now take r equals to the distance between $\mathrm{Conv}(\mathcal{E})$ and the complement of $\mathcal{U}$. By compactness of $\mathrm{Conv}(\mathcal{E})$, r is positive, and the claim follows. □

Recall that $W = (M \backslash \bigcup_i A_i) \subset B_\varepsilon(\mathrm{Fix}(S))$. For each $x \in \mathrm{Fix}(S)$, write $W_x = W \cap B_\varepsilon(x)$. For small enough ε, the set W_x is the connected component of W which contains x. Consider the probability measure $\nu_\varepsilon \in \mathrm{Conv}(\mathcal{E})$ defined by

$$\nu_\varepsilon := \sum_{x \in \mathrm{Fix}(S)} \mu(W_x)\delta_x + \alpha \mathrm{Vol},$$

where δ_x denotes the Dirac measure at x. Note that $\nu_\varepsilon \in \mathrm{Conv}(\mathcal{E})$ because $\sum_{x \in \mathrm{Fix}(S)} \mu(W_x) + \alpha = \mu(W) + \mu(\cup A_i) = \mu(M) = 1$.

Proposition 4 follows immediately from Claim 22 and the next claim.

Claim 23. *Let $\delta, f_1, \dots, f_k$ be as in Claim 22. If $\mu \in \mathcal{P}(M)$ satisfies Eq. (5) for a sufficiently small value of ε, then $\mu \in U_{\delta, f_1, \dots, f_k}(\nu_\varepsilon)$.*

We will now provide a proof of the above claim. Since the A_i's are of diameter less than $\varepsilon > 0$, the following two inequalities hold for sufficiently small values of ε:

$$\left| \int f \, d\mu - \sum_i f(y_i)\mu(A_i) - \sum_{x \in \mathrm{Fix}(S)} f(x)\mu(W_x) \right| < \frac{\delta}{4},$$

$$\left| \int f \, d\nu_\varepsilon - \sum_i f(y_i)\alpha \mathrm{Vol}(A_i) - \sum_{x \in \mathrm{Fix}(S)} f(x)\mu(W_x) \right| < \frac{\delta}{4},$$

where $f \in \{f_1, \dots, f_k\}$ and $y_i \in A_i$. It follows from the above two inequalities that $\forall f \in \{f_1, \dots, f_k\}$, we have

$$\left| \int f \, d\nu_\varepsilon - \int f \, d\mu \right| < \sum_i |\alpha \mathrm{Vol}(A_i) - \mu(A_i)| + \frac{\delta}{2}.$$

Now, since $(M \backslash \bigcup_i A_i) \subset B_\varepsilon(\mathrm{Fix}(S))$, we have $\sum_i \left|\mathrm{Vol}(A_i) - \frac{1}{N}\right| < \varepsilon$, if ε is taken to be sufficiently small. Thus,

$$\begin{aligned} &\sum_i |\alpha \mathrm{Vol}(A_i) - \mu(A_i)| \leqslant \alpha \sum_i |\mathrm{Vol}(A_i) - \frac{1}{N}| + \sum_i |\frac{\alpha}{N} - \mu(A_i)| \\ &\leqslant \varepsilon + \varepsilon. \end{aligned}$$

Note that to obtain the last inequality we have used Inequality (5) and the fact that $\alpha < 1$. Finally, we conclude from the above that, if ε is taken to be sufficiently small, then

$$\left| \int f \, d\nu_\varepsilon - \int f \, d\mu \right| < \delta,$$

for every $f \in \{f_1, \ldots, f_k\}$. This completes the proof of Claim 23 and hence that of Proposition 4. □

Acknowledgements

This project began as an attempt at answering a question of Viktor Ginzburg and Başak Gürel; we are grateful to them for stimulating discussions and their interest. We also thank Laurent Charles, Sylvain Crovisier, Yael Karshon, Leonid Polterovich, Shu Shen, and Maxime Zavidovique for helpful comments. SS: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 851701).

References

[1] Anosov, D.V., Katok, A.B.: New examples of ergodic diffeomorphisms of smooth manifolds. Uspehi Mat. Nauk **25**(4(154)), 173–174 (1970)

[2] Atiyah, M.F.: Convexity and commuting Hamiltonians. Bull. Lond. Math. Soc. **14**(1), 1–15 (1982)

[3] Audin, M.: Torus actions on symplectic manifolds, volume 93 of Progress in Mathematics. Birkhäuser Verlag, Basel, revised edition (2004)

[4] Avila, A., Fayad, B., Le Calvez, P., Disheng, X., Zhang, Z.: On mixing diffeomorphisms of the disc. Invent. Math. **220**(3), 673–714 (2020)

[5] Béguin, F., Crovisier, S., Le Roux, F., Patou, A.: Pseudo-rotations of the closed annulus: variation on a theorem of J. Kwapisz. Nonlinearity **17**(4), 1427–1453 (2004)

[6] Béguin, F., Crovisier, S., Le Roux, F.: Pseudo-rotations of the open annulus. Bull. Braz. Math. Soc. (N.S.) **37**(2), 275–306 (2006)

[7] Bramham, B.: Periodic approximations of irrational pseudo-rotations using pseudoholomorphic curves. Ann. Math. (2) **181**(3), 1033–1086 (2015)

[8] Bramham, B.: Pseudo-rotations with sufficiently Liouvillean rotation number are C^0-rigid. Invent. Math. **199**(2), 561–580 (2015)

[9] Çinelli, E., Ginzburg, V. L., Gürel, B. Z.: From pseudo-rotations to holomorphic curves via quantum steenrod squares (2019). arXiv:1909.11967

[10] Çinelli, E., Ginzburg, V.L., Gürel, B.Z.: Pseudo-rotations and holomorphic curves (2019). arXiv:1905.07567

[11] da Silva, A.C.: Symplectic toric manifolds. In: Symplectic geometry of integrable Hamiltonian systems (Barcelona, 2001), Adv. Courses Math. CRM Barcelona, pages 85–173. Birkhäuser, Basel (2003)

[12] Ekeland, I., Hofer, H.: Symplectic topology and Hamiltonian dynamics. Math. Z. **200**(3), 355–378 (1989)

[13] Eliashberg, Y., Mishachev, N.: Introduction to the h-principle. Graduate Studies in Mathematics, vol. 48. American Mathematical Society, Providence, RI (2002)

[14] Fathi, A., Herman, M.R.: Existence de difféomorphismes minimaux. In: Dynamical systems, Vol. I—Warsaw, pages 37–59. Astérisque, No. 49 (1977)

[15] Fayad, B., Katok, A.: Constructions in elliptic dynamics. Ergodic Theory Dyn. Syst. **24**(5), 1477–1520 (2004)

[16] Franks, J.: Generalizations of the Poincaré–Birkhoff theorem. Ann. Math. (2) **128**(1), 139–151 (1988)

[17] Franks, J.: Rotation numbers for area preserving homeomorphisms of the open annulus. In: Dynamical systems and related topics (Nagoya, 1990), volume 9 of Adv. Ser. Dynam. Systems, pages 123–127. World Sci. Publ., River Edge, NJ (1991)

[18] Ginzburg, V.L.: The Conley conjecture. Ann. Math. (2) **172**(2), 1127–1180 (2010)

[19] Ginzburg, V.L., Gürel, B.Z.: Pseudo-rotations vs. rotations (2018). arXiv:1812.05782

[20] Ginzburg, V.L., Gürel, B.Z.: Hamiltonian pseudo-rotations of projective spaces. Invent. Math. **214**(3), 1081–1130 (2018)

[21] Guillemin, V.: Moment maps and combinatorial invariants of Hamiltonian T^n-spaces. Progress in Mathematics, vol. 122. Birkhäuser Boston Inc, Boston, MA (1994)

[22] Guillemin, V., Sternberg, S.: Convexity properties of the moment mapping. Invent. Math. **67**(3), 491–513 (1982)

[23] Hernández-Corbato, L., Presas, F.: The conjugation method in symplectic dynamics. Rev. Mat. Iberoam. **35**(1), 1–22 (2019)

[24] Hingston, N.: Subharmonic solutions of Hamiltonian equations on tori. Ann. Math. (2) **170**(2), 529–560 (2009)

[25] Katok, A.B.: Ergodic perturbations of degenerate integrable Hamiltonian systems. Izv. Akad. Nauk SSSR Ser. Mat. **37**, 539–576 (1973)

[26] Le Calvez, P.: Periodic orbits of Hamiltonian homeomorphisms of surfaces. Duke Math. J. **133**(1), 125–184 (2006)

[27] Le Calvez, P.: A finite dimensional approach to Bramham's approximation theorem. Ann. Inst. Fourier (Grenoble) **66**(5), 2169–2202 (2016)

[28] McDuff, Dusa, Salamon, Dietmar: Introduction to Symplectic Topology. Oxford Graduate Texts in Mathematics, 3rd edn. Oxford University Press, Oxford (2017)

[29] Ono, K.: Floer–Novikov cohomology and the flux conjecture. Geom. Funct. Anal. **16**(5), 981–1020 (2006)

[30] Polterovich, Leonid: Hamiltonian loops from the ergodic point of view. J. Eur. Math. Soc. (JEMS) **1**(1), 87–107 (1999)

[31] Schlenk, F.: An extension theorem in symplectic geometry. Manuscripta Math. **109**(3), 329–348 (2002)

[32] Shelukhin, E.: Pseudo-rotations and steenrod squares (2019). arXiv:1905.05108

[33] Shelukhin, E.: Pseudo-rotations and steenrod squares revisited (2019). arXiv:1909.12315

[34] Walters, P.: An introduction to ergodic theory. Graduate Texts in Mathematics, vol. 79. Springer-Verlag, New York-Berlin (1982)

Frédéric Le Roux and Sobhan Seyfaddini
Sorbonne Université and Université de Paris, CNRS, IMJ-PRG
Paris 75006
France
e-mail: frederic.le-roux@imj-prg.fr;
sobhan.seyfaddini@imj-prg.fr

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:29
https://doi.org/10.1007/s11784-022-00956-7
Published online April 7, 2022

Journal of Fixed Point Theory and Applications

Contact geometry in the restricted three-body problem: a survey

Agustin Moreno

Abstract. This survey is based on lecture notes for an online mini-course taught for postgraduate students at UDELAR, Montevideo, Uruguay, in November 2020, remotely from the Mittag–Leffler Institute in Djursholm, Sweden. Lectures were recorded and are available online.

Mathematics Subject Classification. 37-02, 70F07, 53D05, 53D10.

Keywords. Three-body problem, Contact geometry, Symplectic geometry, Hamiltonian dynamics, Symplectic dynamics.

Contents

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

1. Introduction

The current survey is an attempt to put into context a series of very recent results of the author in co-authorship with Otto van Koert [98,99], and the spin-off [100] by the author. It also serves the purpose of introducing and threading together a collection of basic and important notions, disseminated across the literature, with the main driving motivation coming from a very old and famous problem; namely, the three-body problem. We shall be, therefore, mainly interested in Hamiltonian dynamics, and the intended audience is that with a dynamical background/interest; a good deal of openness towards topological/geometric/holomorphic techniques is also recommended. We make no assumptions on previous knowledge on contact or symplectic techniques, but we move at a fast pace.

We shall start from the basics of contact and symplectic geometry, the geometries of classical mechanics, and move on to the more topological notion of open book decompositions in the context of contact topology and Giroux's correspondence. We will then make a dynamical jump to discuss the notion of global hypersurfaces of section and adapted dynamics, discussing examples along the way. After paving the road, we focus on the three-body problem (more precisely, a simplified version, the circular restricted case=CR3BP) with the main interest being the *spatial* problem where the small mass is allowed to move anywhere (SCR3BP), as opposed to the *planar* problem, which historically has been of central interest. We give a historical account of Poincaré's original approach in the planar problem, and discuss classical fixed-point theorems and perturbative results. We also provide a brief survey of the beautiful history behind the search of closed geodesics, which one may view as a spin-off of the search of closed orbits for the three-body problem; as well as how this relates to recent developments of a dynamical flavor in symplectic geometry. We further review non-perturbative modern results coming from holomorphic curve theory à-la Hofer–Wysocki–Zehnder [73]. We then introduce the main results of [98–100], which include:

- Existence of adapted open book decompositions for the SCR3BP in the low-energy range (Theorem M);
- Existence of Hamiltonian return maps reducing the dynamics to dimension 4 (Theorem N);
- A generalization of the classical Poincaré–Birkhoff theorem for Liouville domains in arbitrary even dimensions (Theorem O);

- The construction by the author of the *holomorphic shadow*, which associates with the SCR3BP (whenever the planar dynamics is convex, and energy is low) a Reeb dynamics on S^3 which is adapted to a trivial open book (Theorem R); and (perturbative) dynamical applications.

We remark that the first two results are valid for arbitrary mass ratio and are therefore non-perturbative. We also point out that the second result, while a general fixed-point theorem, has not so far seen an application to the SCR3BP, for which the generalized notion of a twist condition introduced in [99] seems, as of yet, perhaps unsuitable. The third result, while of theoretical interest, might perhaps lead to insights on the original problem coming from 3-dimensional dynamics; this is work in progress. In fact, everything in the last sections should be considered work in progress. Therefore, the reader is advised to proceed accordingly, and perhaps get excited enough to contribute to this growing body of work.

Needless to say, this account will be very biased towards the author's interests; the subject is too vast to make it proper justice. The experienced reader is encouraged to complain to the author for misinterpretations, misrepresentations, omissions, or mistakes. Disseminated across the text, we leave a series of digressions, intended for non-experts and newcomers, which the reader might choose to skip without affecting the understanding of the main body. They take up a significant part of the document, in the hope to illustrate the richness of the material.

2. Basic concepts

We start with the basic concepts underlying the general principles of classical mechanics.

2.1. Symplectic geometry

Roughly speaking, symplectic geometry is the geometry of phase space (where one keeps track of position and velocities of classical particles, and so, it is a theory in even dimensions). Formally, a *symplectic manifold* is a pair (M, ω), where M is a smooth manifold with $\dim(M) = 2n$ even, and $\omega \in \Omega^2(M)$ is a two-form (the *symplectic form*) satisfying

- (closedness) $d\omega = 0$;
- (non-degeneracy) $\omega^n = \omega \wedge \cdots \wedge \omega \in \Omega^{2n}(M)$ is nowhere-vanishing, and hence a volume form. Equivalently, the map

$$\begin{aligned} \mathfrak{X}(M) &\to \Omega^1(M) \\ X &\mapsto i_X\omega = \omega(X, \cdot) \end{aligned}$$

is a linear isomorphism.

Note that symplectic manifolds are always orientable. We assume that M is always oriented by the orientation induced by the symplectic form.

Example 2.1. (From classical mechanics)

- (Phase space) $(\mathbb{R}^{2n}, \omega_{\mathrm{std}})$, where, writing $(q,p) \in \mathbb{R}^{2n} = \mathbb{R}^n \oplus \mathbb{R}^n$ (q =position, p =momenta), we have
$$\omega_{\mathrm{std}} = -d\lambda_{\mathrm{std}} = dq \wedge dp,$$
where $\lambda_{\mathrm{std}} = pdq$ is the standard *Liouville form.* Here, we use the shorthand notation $dq \wedge dp = \sum_{i=1}^n dq_i \wedge dp_i$, and similarly, $pdq = \sum_{i=1}^n p_i dq_i$.
- (cotangent bundles) $(T^*Q, \omega_{\mathrm{std}})$, where Q is a closed n-manifold, and ω_{std} is defined invariantly as
$$\omega_{\mathrm{std}} = -d\lambda_{\mathrm{std}},$$
with
$$(\lambda_{\mathrm{std}})_{(q,p)}(\eta) = p(d_{(q,p)}\pi(\eta)),$$
also called the standard Liouville form. Here, q is a point in the base, and p a covector in T_qQ^*, and
$$\pi : T^*Q \to Q$$
is the natural projection to the base. Note that phase space corresponds to the case $Q = \mathbb{R}^n$.

A general important feature of symplectic manifolds (or, more like, the reason for their existence) is that they are locally modelled on phase space:

Theorem A. (Darboux's theorem for symplectic manifolds) *If $p \in (M, \omega)$ is an arbitrary point in a symplectic manifold, we can find local charts centered at p, so that (M, ω) is isomorphic to standard phase space $(\mathbb{R}^{2n}, \omega_{\mathrm{std}})$ in this local chart.*

The notion of isomorphism we use above is the obvious one: two symplectic manifolds (M_1, ω_1) and (M_2, ω_2) are *symplectomorphic* if there exists a diffeomorphism $f : M_1 \to M_2$ satisfying $f^*\omega_2 = \omega_1$. In particular, a symplectomorphism preserves volume, i.e., $f^*\omega_2^n = \omega_1^n$. Darboux's theorem is usually interpreted as saying that, unlike in Riemannian geometry where the curvature is a local isometry invariant, there are no local invariants for symplectic manifolds (they locally all look the same).

Hamiltonian dynamics From a dynamical perspective, symplectic manifolds are the natural geometric space where one can study Hamiltonian dynamics, via the *Hamiltonian formalism.* On a cotangent bundle T^*Q, the idea is to model the motion of a particle moving along the manifold Q, subject to the principle of minimization of energy/action associated with a given physical problem.

In general, we start with a symplectic manifold (M, ω), and a *Hamiltonian* $H : M \to \mathbb{R}$, which is simply a function (which we assume C^1, say), thought of as the *energy* function of the mechanical system. The symplectic form implicitly defines a vector field $X_H \in \mathfrak{X}(M)$ (the *Hamiltonian vector field* or *Hamiltonian gradient* of H) via the equation
$$i_{X_H}\omega = dH.$$
Note that this uniquely defines X_H due to non-degeneracy of ω. The above equation is the global, invariant version for the following:

Fundamental example: Hamilton's equation Whenever $(M,\omega) = (\mathbb{R}^{2n}, \omega_{\mathrm{std}})$, we have

$$X_H = \left(\frac{\partial H}{\partial p}, -\frac{\partial H}{\partial q}\right) = \frac{\partial H}{\partial p}\partial_q - \frac{\partial H}{\partial q}\partial_p.$$

In other words, a solution $x(t) = (q(t), p(t))$ to the ODE $\dot{x}(t) = X_H(x(t))$ is precisely a solution to the Hamilton equations

$$\begin{cases} \dot{q} = \frac{\partial H}{\partial p} \\ \dot{p} = -\frac{\partial H}{\partial q}. \end{cases}$$

By Darboux's theorem, we see that, locally, solutions to the Hamiltonian flow are solutions to the above.

More invariantly, we consider the Hamiltonian flow $\phi_t^H : M \to M$ generated by H, i.e., the unique solution to the equations

$$\phi_0^H = \mathrm{id}, \ \frac{\mathrm{d}}{\mathrm{d}t}\phi_t^H = X_H \circ \phi_t^H.$$

This flow can be thought of as a symmetry of the symplectic manifold, since it preserves the symplectic form

$$\frac{\mathrm{d}}{\mathrm{d}t}(\phi_t^H)^*\omega = \mathcal{L}_{X_H}\omega = i_{X_H}\mathrm{d}\omega + di_{X_H}\omega = 0 + d^2H = 0,$$

and so, $(\phi_t^H)^*\omega = (\phi_0^H)^*\omega = \omega$ for every t. A symplectomorphism $f : (M,\omega) \to (M,\omega)$ is called *Hamiltonian* whenever $f = \phi_H^1$ is the time-1 map of a Hamiltonian flow. Hamiltonian maps then preserve volume (which is a way of stating Liouville's theorem from classical mechanics).

Remark 2.2. The Hamiltonian usually also depends on time. We have assumed for simplicity that it does not, i.e., it is autonomous. We will see that this will hold for the simplified versions of the three-body problem we will consider.

In the above symplectic formalism, it is a fairly straightforward matter to write down the fundamental conservation of energy principle (in the autonomous case):

Theorem B. (Conservation of energy) *Assume H is autonomous. Then*

$$dH(X_H) = 0.$$

In other words, the level sets $H^{-1}(c)$ are invariant under the Hamiltonian flow.

This is also usually written down using the *Poisson bracket* as

$$\{H, H\} = 0,$$

which is another way of saying that H is preserved under the Hamiltonian flow of itself, or that H is a conserved quantity (or integral) of motion. The proof fits in one line

$$dH(X_H) = i_{X_H}\omega(X_H) = \omega(X_H, X_H) = 0,$$

since ω is skew-symmetric.

2.2. Contact geometry

Contact geometry is, roughly speaking, the odd-dimensional analogue of symplectic geometry, and arises on level sets of Hamiltonians satisfying a suitable convexity assumption (see Proposition 2.5). Formally, a *(strict) contact manifold* is a pair (X, α), where X is a smooth manifold with $\dim(X) = 2n - 1$ odd, and $\alpha \in \Omega^1(X)$ is a 1-form (the *contact form*) satisfying the *contact* condition

$$\alpha \wedge d\alpha^{n-1} \neq 0 \text{ is nowhere-vanishing, and hence a volume form.}$$

Contact manifolds are therefore orientable (see Remark 2.4 below). The codimension-1 distribution $\xi = \ker \alpha \subset TM$ (a choice of hyperplane at each tangent space, varying smoothly with the point), is called the *contact structure* or *contact distribution*, and (M, ξ) is a *contact manifold.*

Example 2.3.
- (standard) The standard contact form on $\mathbb{R}^{2n-1} = \mathbb{R} \oplus \mathbb{R}^{n-1} \oplus \mathbb{R}^{n-1} \ni (z, q, p)$ is
$$\alpha_{\mathrm{std}} = \mathrm{d}z - p\mathrm{d}q,$$
where we again use the short-hand notation $p\mathrm{d}q = \sum_{i=1}^n p_i \mathrm{d}q_i$.
- (First-jet bundles) Given a manifold Q, its first-jet bundle $J^1(Q) \to Q$, by definition, has total space the collection of all possible first derivatives of maps $f : Q \to \mathbb{R}$. The fiber over q is as all possible tuples $(q, f(q), d_q f)$, and so, $J^1(Q) \cong \mathbb{R} \times T^*Q$. It carries the natural contact form
$$\alpha = \mathrm{d}z + \lambda_{\mathrm{std}},$$
where z is the coordinate on the first factor, and λ_{std} is the standard Liouville form on T^*Q; note that the standard contact form corresponds to the case $Q = \mathbb{R}^{n-1}$.
- (contactization) More generally: if $(M, \omega = \mathrm{d}\lambda)$ is an exact symplectic manifold, then its *contactization* is
$$(\mathbb{R} \times M, \mathrm{d}z + \lambda),$$
where z is the coordinate in the first factor.

The contact condition should be thought of as a *maximally non-integrability* condition, as follows. Recall the following theorem from differential geometry:

Theorem C. (Frobenius' theorem) *If $\alpha \wedge \mathrm{d}\alpha \equiv 0$, then $\xi = \ker \alpha \subset TM$ is integrable. That is, there are codimension-1 submanifolds whose tangent space is ξ.*

The condition in Frobenius' theorem is equivalent to $d\alpha|_\xi \equiv 0$. The contact condition is the extreme opposite of the above: $\mathrm{d}\alpha|_\xi > 0$ is symplectic, i.e., non-degenerate. In fact, if $Y \subset (X, \xi)$ is a submanifold of a $(2n-1)$-dimensional contact manifold, so that $TY \subset \xi$ (i.e., Y is *isotropic*), then $\dim(Y) \leq n - 1$. The isotropic submanifolds of maximal dimension $n - 1$ are called *Legendrians.*

The analogous theorem of Darboux in the contact category is the following:

Theorem D. (Darboux's theorem for contact manifolds) *If $p \in (X, \lambda)$ is an arbitrary point in a strict contact manifold, we can find a local chart $U \cong \mathbb{R}^{2n-1}$ centered at p, so that $\lambda|_U = \alpha_{\text{std}}$.*

Reeb dynamics Whereas a contact manifold is a geometric object, a *strict* contact manifold is a dynamical one, as we shall see below. Note first that the choice of contact form for a contact structure ξ is not unique: if α is such a choice, then $\nu\alpha$ is also, for any smooth positive function $\nu > 0$. This is in fact the only ambiguity, i.e., every other contact form is of this form.

Given a contact form α, it defines an autonomous dynamical system on X, generated by the *Reeb vector field* $R_\alpha \in \mathfrak{X}(X)$. This is defined implicitly via

- $i_{R_\alpha}\mathrm{d}\alpha = 0$;
- $\alpha(R_\alpha) = 1$.

To understand the above, note that, since $\mathrm{d}\alpha|_\xi$ is symplectic, the kernel of $d\alpha$ is the 1-dimensional distribution $TX/\xi \subset TX$. This is trivialized (as a real line bundle) via a choice of contact form, which also gives it an orientation induced from the one on M. The Reeb vector field then lies in this 1-dimensional distribution; the second condition normalizes it, so that it points precisely in the positive direction with respect to the co-orientation. We emphasize that the Reeb vector field depends significantly on the contact form, and not on the contact structure; different choices give, in general, very different dynamical systems.

Remark 2.4. There are also examples of contact manifolds which are not globally co-orientable (e.g., the space of contact elements); we will not be concerned with those.

The Reeb flow φ_t has the property that it preserves the geometry in a strict way, i.e., it is a *strict contactomorphism.* This means that $\varphi_t^*\alpha = \alpha$, or in other words, the Reeb vector field generates a (strict) local symmetry of the (strict) contact manifold. This fact easily follows from the Cartan formula

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi_t^*\alpha = di_{R_\alpha}\alpha + i_{R_\alpha}\mathrm{d}\alpha = d(1) + 0 = 0,$$

and so $\varphi_t^*\alpha = \varphi_0^*\alpha = \alpha$.

More generally, a (not necessarily strict) contactomorphism is a diffeomorphism f, such that $f^*(\xi) = \xi$, or $f^*\alpha = \nu\alpha$ for some strictly positive smooth function ν.

The bridge The fundamental relationship between symplectic and contact geometry lies in the following. If the symplectic form $\omega = d\lambda$ is exact (which can only happen if the symplectic manifold is open, by Stokes' theorem), then we have a *Liouville* vector field V, defined implicitly via

$$i_V\omega = \lambda,$$

where we again use non-degeneracy of ω. To understand this vector field, consider φ_t the flow of V. The Cartan formula implies

$$\frac{\mathrm{d}}{\mathrm{d}t}\varphi_t^*\omega = di_V\omega + i_V\mathrm{d}\omega = \mathrm{d}\lambda = \omega,$$

and so, integrating, we get

$$\varphi_t^*\omega = e^t\omega.$$

Taking the top wedge power of this equation: $\varphi_t^*\omega^n = e^{nt}\omega^n$, and we see that the symplectic volume grows exponentially along the flow of V, i.e., φ_t is a *symplectic dilation*.

Assume that $X \subset (M, \omega = \mathrm{d}\lambda)$ is a co-oriented codimension-1 submanifold, and the Liouville vector field is positively transverse to X. Then, we obtain a volume form on X by contraction

$$0 < i_V\omega^n|_X = ni_V\omega \wedge \omega^{n-1}|_X = n\lambda \wedge \mathrm{d}\lambda^{n-1}|_X = n\alpha \wedge d\alpha^{n-1},$$

where $\alpha = \lambda|_X$. We have proved:

Proposition 2.5. *If $\omega = \mathrm{d}\lambda$, and the associated Liouville vector field V is positively transverse to X, then $(X, \alpha = \lambda|_X = i_V\omega|_X)$ is a strict contact manifold.*

A hypersurface X as in the above proposition is then called *contact-type*. The most relevant example to keep in mind is when $X = H^{-1}(c)$ is the level set of a Hamiltonian (in fact, locally, this is always the case). In this situation:

Proposition 2.6. *If $X = H^{-1}(c)$ is contact-type, then the Reeb dynamics on X is a positive reparametrization of the Hamiltonian dynamics of H.*

This follows from the observation that both X_H and R_α span the kernel of $d\alpha$ along X. In other words, *Reeb dynamics on contact-type Hamiltonian level sets is dynamically equivalent to Hamiltonian dynamics.* See Fig. 1 for an abstract sketch.

Example 2.7.
- (star-shaped domains) Assume that $X \subset \mathbb{R}^{2n}$ is *star-shaped*, i.e., it bounds a compact domain D containing the origin, and the radial vector field $V = q\partial_q + p\partial_p = r\partial_r$ is positively transverse to X (with the boundary orientation). Since V is precisely the Liouville vector field associated with λ_{std}, every star-shaped domain is contact-type.
- (standard contact form on S^3) As a particular case, let $S^3 = \{z \in \mathbb{R}^4 : |z| = 1\} \subset \mathbb{R}^4$ be the round 3-sphere. Then, $S^3 = H^{-1}(1/2)$, where $H : \mathbb{R}^4 \to \mathbb{R}$, $H(z) = \frac{1}{2}|z|^2$, and it is star-shaped. Writing $z = (z_1, z_2) = (x_1, y_1, x_2, y_2)$, the radial vector field

$$V = \frac{1}{2}r\partial_r = \frac{1}{2}(x_1\partial_{x_1} + y_1\partial_{y_1} + x_2\partial_{x_2} + y_2\partial_{y_2})$$

is Liouville and induces the contact form

$$\alpha = i_V\omega_{\mathrm{std}}|_{S^3} = \lambda_{\mathrm{std}}|_{S^3} = \frac{1}{2}(x_1\mathrm{d}y_1 - y_1\mathrm{d}x_1 + x_2\mathrm{d}y_2 - y_2\mathrm{d}x_2)|_{S^3}$$

on S^3 whose Reeb vector field is

$$R_\alpha = 2(x_1\partial_{y_1} - y_1\partial_{x_1} + x_2\partial_{y_2} - y_2\partial_{x_2}).$$

Its Reeb flow is, in complex coordinates, $\varphi_t(z_1, z_2) = e^{2\pi it}(z_1, z_2)$, whose orbits are precisely the fibers of the Hopf fibration $S^3 \ni (z_1, z_2) \mapsto [z_1 : z_2] \in \mathbb{C}P^1$. In particular, this flow is periodic, and all orbits have the same period.

As a side remark: the Hopf fibration $\pi : S^3 \to S^2 = \mathbb{C}P^1$ is an example of what is usually called a *prequantization bundle*, i.e., the contact form α is a connection form whose curvature form on the base is symplectic. In other words, $\mathrm{d}\alpha = i\pi^*\omega_{\mathrm{FS}}$ for a symplectic form ω_{FS} on S^2, and its Reeb orbits are the S^1-fibers (here, ω_{FS} is the Fubini–Study metric on $\mathbb{C}P^1$, and the line bundle associated with the principal S^1-bundle π is $\mathcal{O}(1) \to \mathbb{C}P^1$; see the digression on line bundles below).

- (ellipsoids) Given $a, b > 0$, define the *ellipsoid*

$$E(a,b) = \left\{ (z_1, z_2) \in \mathbb{C}^2 : \frac{\pi|z_1|^2}{a} + \frac{\pi|z_2|^2}{b} \leq 1 \right\},$$

a star-shaped domain. The restriction of the symplectic form ω_{std} is a symplectic form on $E(a,b)$, and its boundary $\partial E(a,b)$ inherits a contact form $\lambda_{\mathrm{std}}|_{\partial E(a,b)}$ whose Reeb flow is

$$\varphi_t(z_1, z_2) = (e^{2\pi iat}z_1, e^{2\pi ibt}z_2).$$

In particular, if a, b are rationally independent, then this Reeb flow has only two periodic orbits, passing through the points $z_1 = 0$, or $z_2 = 0$. If $a = b$, $E(a,a)$ is the unit ball, and we recover the Hopf flow along the standard $S^3 = \partial E(a,a)$.

- (Unit cotangent bundle and geodesic flows) Given a manifold Q, choose a Riemannian metric on TQ (which induces a metric on T^*Q), and

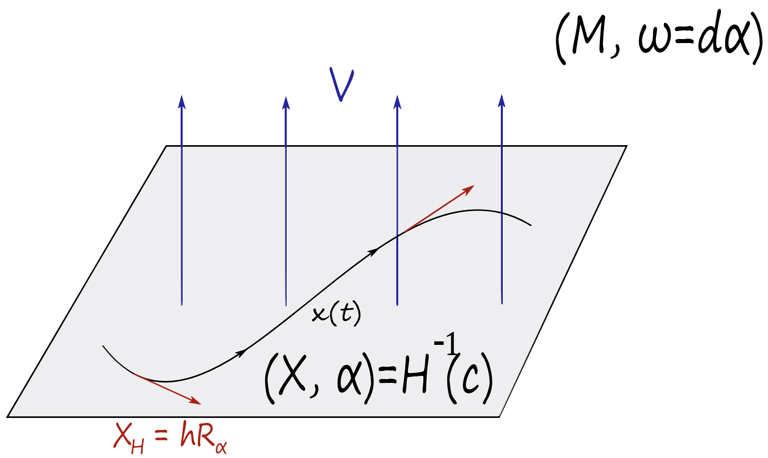


FIGURE 1. The fundamental relationship between contact and symplectic geometry is summarized here

consider its unit cotangent bundle

$$S^*Q = \{(q,p) \in T^*Q : |p| = 1\}.$$

We have $S^*Q = H^{-1}(1/2)$, where $H : T^*Q \to \mathbb{R}$, $H(q,p) = \frac{|p|^2}{2}$ is the kinetic energy Hamiltonian. The radial vector field $V = p\partial_p$ on each fiber is the Liouville vector field associated with λ_{std}, and is positively transverse to S^*Q. It follows that $\alpha_{\text{std}} := \lambda_{\text{std}}|_{S^*Q}$ is a contact form, and $(S^*Q, \xi_{\text{std}} = \ker \alpha_{\text{std}})$ is called the standard contact structure on S^*Q. Its Reeb dynamics is the (co)geodesic flow. We see that *a geodesic flow is a particular case of a Reeb flow.*

Symplectization Given a contact form α on X, its *symplectization* is the symplectic manifold

$$(\mathbb{R} \times X, \omega = d(e^t\alpha)).$$

The Liouville vector field is $V = \partial_t$, which is positively transverse to all slices $\{t\} \times X$, where it induces the contact form $i_V\omega = e^t\alpha$. Note that the Reeb dynamics is the same in each slice (i.e., it is only rescaled by a constant positive multiple). In fact, the symplectization is the "universal neighbourhood" for every contact-type hypersurface:

Proposition 2.8. *Let $X \subset (M,\omega)$ be a contact-type hypersurface, with $\omega = d\lambda$ exact near X. Then, we can find sufficiently small $\epsilon > 0$, and an embedding*

$$\Phi : (-\epsilon, \epsilon) \times X \hookrightarrow M,$$

so that $\Phi^*\omega = d(e^t\alpha)$ where $\alpha = \lambda|_X$.

In other words, contact manifolds are always contact-type in some symplectic manifolds, and vice versa. We can summarize this discussion in the following motto: *contact geometry is $\mathbb{R}$-invariant symplectic geometry.*

Remark 2.9. One also calls the symplectic manifold $(\mathbb{R} \times X, \omega = d(r\alpha))$ the symplectization of α; this is related to the above by the obvious change of coordinates $r = e^t$. We shall use the two interchangeably. Note that $X = \{t = 0\} = \{r = 1\}$.

Digression: examples of symplectic manifolds from complex algebraic/ Kähler geometry

Example 2.10.

- (Projective varieties) The complex projective space $\mathbb{C}P^n$ admits a natural symplectic form, called the *Fubini–Study* form ω_{FS}, defined as follows. Let

$$K : \mathbb{C}^n \to \mathbb{R}$$
$$K(z) = \log\left(1 + \sum_{i=1}^n |z_i|^2\right).$$

In homogenous coordinates $(\zeta_0 : \cdots : \zeta_n)$ for $\mathbb{C}P^n$, let $U_\alpha = \{(\zeta_0 : \cdots : \zeta_n) : \zeta_\alpha \neq 0\}$ and

$$\varphi_\alpha : U_\alpha \to \mathbb{C}^n,$$

$$\varphi_\alpha(\zeta_0 : \cdots : \zeta_n) = \left(\frac{\zeta_0}{\zeta_i}, \ldots, \frac{\zeta_{i-1}}{\zeta_i}, \frac{\zeta_{i+1}}{\zeta_i}, \ldots, \frac{\zeta_n}{\zeta_i}\right) = (z_1^\alpha, \ldots, z_n^\alpha)$$

be the standard affine chart around $(0 : \cdots : 1 : \cdots : 0)$. Let $K_\alpha = K \circ \varphi_\alpha$, and define

$$\omega_\alpha = \sqrt{-1}\partial\overline{\partial}K_\alpha = \sum_{i,j=1}^n h_{ij}(z^\alpha)\mathrm{d}z_i^\alpha \wedge \mathrm{d}\overline{z}_j^\alpha.$$

Here, one computes

$$h_{ij}(z^\alpha) = \frac{\delta_{ij}\left(1+\sum_{i=1}^n |z_i^\alpha|^2\right) - z_i^\alpha \overline{z}_j^\alpha}{\left(1+\sum_{i=1}^n |z_i^\alpha|^2\right)^2}.$$

One checks that on overlaps $U_\alpha \cap U_\beta$, we have $\omega_\alpha = \omega_\beta$, and so, we get a well-defined global ω_{FS} so that $\omega_{FS}|_{U_\alpha} = \omega_\alpha$. The K_α are what is called a local Kähler potential (or plurisubharmonic function) for the Fubini–Study form. Every algebraic/analytic projective variety inherits a symplectic form via restriction of the ambient Fubini–study form.

- (Affine varieties: Stein manifolds) The standard complex affine space $\mathbb{C}^n$ carries the standard symplectic form via the identification $\mathbb{C}^n = \mathbb{R}^{2n}$, which in complex notation is

$$\omega_{\mathrm{std}} = \frac{\sqrt{-1}}{2}\sum_{i=1}^n \mathrm{d}z_i \wedge \mathrm{d}\overline{z}_j =: \frac{\sqrt{-1}}{2}\mathrm{d}z \wedge \mathrm{d}\overline{z} = -\mathrm{d}\lambda_{\mathrm{std}}$$

with $\lambda_{\mathrm{std}} = \frac{\sqrt{-1}}{4}(zd\overline{z} - \overline{z}\mathrm{d}z)$. This admits the standard plurisubharmonic function

$$f_{\mathrm{std}}(z) = |z|^2,$$

i.e., $\omega_{\mathrm{std}} = \sqrt{-1}\partial\overline{\partial}f_{\mathrm{std}}$. This function is exhausting (i.e., $\{z : f(z) \leq c\}$ is compact for every $c \in \mathbb{R}$), and is a Morse function (with a unique critical point at the origin).

By analogy as with the projective case, a Stein manifold X is a properly embedded complex submanifold of $\mathbb{C}^n$, endowed with the restriction of the standard symplectic form, the standard complex structure i, and the standard plurisubharmonic function. One may further assume (after a small perturbation) that f_{std} defines a Morse function on X.

The above examples (projective and affine) are all instances of Kähler manifolds, i.e., the symplectic form is suitably compatible with an integrable complex structure, and with a Riemannian metric. One way to obtain Stein manifolds from projective varieties is to remove a collection of generic hyperplane sections, i.e., the intersection of the variety with the zero sets of generic homogeneous polynomials of degree 1. A confusing point is that the Liouville form (i.e., the primitive of the resulting symplectic form), depends on the number of sections, as we illustrate as follows in the case of $\mathbb{C}P^n$ as the projective variety.

Continued digression: relationship with line bundles, connections, and Chern–Weil theory First, as a general fact, we recall that the Picard group of $\mathbb{C}P^n$ (i.e., the group of isomorphism classes of holomorphic line bundles, with tensor product as group operation) is isomorphic to $\mathbb{Z}$, each $k \in \mathbb{Z}$ corresponding to a line bundle $\mathcal{O}(k)$. For $k \geq 0$, the holomorphic sections of $\mathcal{O}(k)$ are precisely homogeneous polynomials of degree k on the homogeneous coordinates; $\mathcal{O}(k)$ has no holomorphic sections for $k < 0$, but admits meromorphic sections given by Laurent polynomials with poles of total order k. Moreover, the first Chern class of a line bundle is by definition the Poincaré dual of $Z(s)$, the zero set of a section s, generic in the sense that it is transverse to the zero section. The zero set of a generic polynomial of degree k is, by definition, a hypersurface of degree k. For very degenerate cases (i.e., when the polynomial factorizes into linear polynomials), this consists of a collection of hyperplanes, i.e., zero sets of linear polynomials as e.g., $H = \{\zeta_i = 0\}$, with total multiplicity k. One should think of $\mathbb{C}P^1$, where this zero set is simply a collection of points with total multiplicity k. This translates to the fact that first Chern class of $\mathcal{O}(k)$ is $c_1(\mathcal{O}(k)) = kh \in H^2(\mathbb{C}P^n, \mathbb{R})$, where h is the hyperplane class, the Poincaré dual to the homology class $[H] \in H_{2n-2}(\mathbb{C}P^n, \mathbb{R})$ of any hyperplane H, and a generator of the cohomology of $\mathbb{C}P^n$. On the other hand, Chern–Weil theory says that c_1 is represented by the curvature 2-form of a connection on $\mathcal{O}(k)$ (e.g., the Chern connection associated with the standard Hermitian metric). In practice, this means the following: for $k \geq 0$, take a holomorphic section $s_k \in \Gamma(\mathcal{O}(k))$, and consider $F_k = \sqrt{-1}\partial\overline{\partial}\log(|s_k|^2)$, which a $(1,1)$-form, defined on $X_k := \mathbb{C}P^n \backslash Z(s_k)$. We further have $F_k = -dd^{\mathbb{C}}\log|s_k|^2$, where $d^{\mathbb{C}}$ is defined via $d^{\mathbb{C}}\alpha(X) = \mathrm{d}\alpha(iX)$, and so, F_k is exact on X_k. Moreover, it is symplectic on X_k, which becomes a subset of $\mathbb{C}^n$ after choosing affine charts, and is in fact a Stein manifold, where the appropriate Liouville form for the symplectic form F_k is $\lambda_k = -d^{\mathbb{C}}\log|s_k|^2$. In other words, projective space is obtained from X_k by compactifying with a divisor $Z(s_k)$ "at infinity". Thinking of s_k as providing a local trivialization of $\mathcal{O}(k)$ over X_k, one checks that different choices of local trivializations give different F_k which glue together to a global $(1,1)$-form which is no longer exact, and actually its cohomology class is precisely $c_1(\mathcal{O}(k))$. Note that by construction, any standard chart U_α is of the form $\mathbb{C}P^n \backslash Z(s_1) \cong \mathbb{C}^n$, and $\omega_{FS}|_{U_\alpha} = F_1$, i.e., ω_{FS} is the curvature of the Chern connection on $\mathcal{O}(1)$ and hence Poincaré dual to h.

References Good references for Kähler and complex algebraic geometry are Griffiths–Harris [57], Huybrechts [79], and many others.

2.3. Open book decompositions

Definition 2.11. Let M be a closed manifold. A (concrete) open book decomposition on M is a fibration $\pi : M \backslash B \to S^1$, where $B \subset M$ is a closed, codimension-2 submanifold with trivial normal bundle. We further assume that $\pi(b, r, \theta) = \theta$ along some collar neighbourhood $B \times \mathbb{D}^2 \subset M$, where (r, θ) are polar coordinates on the disk factor.

Note that collar neighbourhoods of B exist, since they are trivializations of its normal bundle. B is called the *binding*, and the closure of the fibers $P_\theta = \overline{\pi^{-1}}(\theta)$ are called the *pages*, which satisfy $\partial P_\theta = B$ for every θ. We usually denote a concrete open book by the pair (π, B). See Fig. 2.

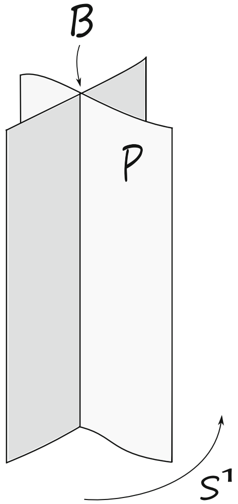


FIGURE 2. A neighbourhood of the binding look precisely like the pages of an open book, whose front cover has been glued to its back cover

The above concrete notion also admits an abstract version, as follows. Given the data of a typical page P (a manifold with boundary B), and a diffeomorphism $\varphi : P \to P$ with $\varphi = id$ in a neighbourhood of B, we can abstractly construct a manifold

$$M := \mathbf{OB}(P, \varphi) := B \times \mathbb{D}^2 \bigcup_{\partial} P_\varphi,$$

where $P_\varphi = P \times [0,1] \backslash (x, 0) \sim (\varphi(x), 1)$ is the associated mapping torus. By gluing the obvious fibration $P_\varphi \to S^1$ with the angular map $(b, r, \theta) \mapsto \theta$ defined on $B \times \mathbb{D}^2$, we see that this abstract notion recovers the concrete one. Reciprocally, every concrete open book can also be recast in abstract terms, where the choices are unique up to isotopy. However, while the two notions are equivalent from a topological perspective, it is important to make distinctions between the abstract and the concrete versions for instance when studying dynamical systems adapted to the open books (as we shall do below), since dynamics is in general very sensitive to isotopies.

Example 2.12.
- (trivial open book) Since the relative mapping class group of $\mathbb{D}^2$ is trivial, the only possible monodromy for an open book with disk-like pages is $S^3 = \mathbf{OB}(\mathbb{D}^2, \mathbb{1})$. Viewing $S^3 = \{(z_1, z_2) \in \mathbb{C}^2 : |z_1|^2 + |z_2|^2 = 1\}$, let $B = \{z_1 = 0\} \subset S^3$ be the binding (the unknot). The concrete version is e.g. $\pi : S^3 \backslash B \to S^1$, $\pi(z_1, z_2) = \frac{z_1}{|z_1|}$. See Fig. 3.
- (stabilized version) We also have $S^3 = \mathbf{OB}(\mathbb{D}^* S^1, \tau)$, where τ is the positive Dehn twist along the zero section S^1 of the annulus $\mathbb{D}^* S^1$. A concrete version is $\pi : S^3 \backslash L \to S^1$, $\pi(z_1, z_2) = \frac{z_1 z_2}{|z_1 z_2|}$, where $L = \{z_1 z_2 = 0\}$ is the Hopf link. This is the positive stabilization of the trivial open

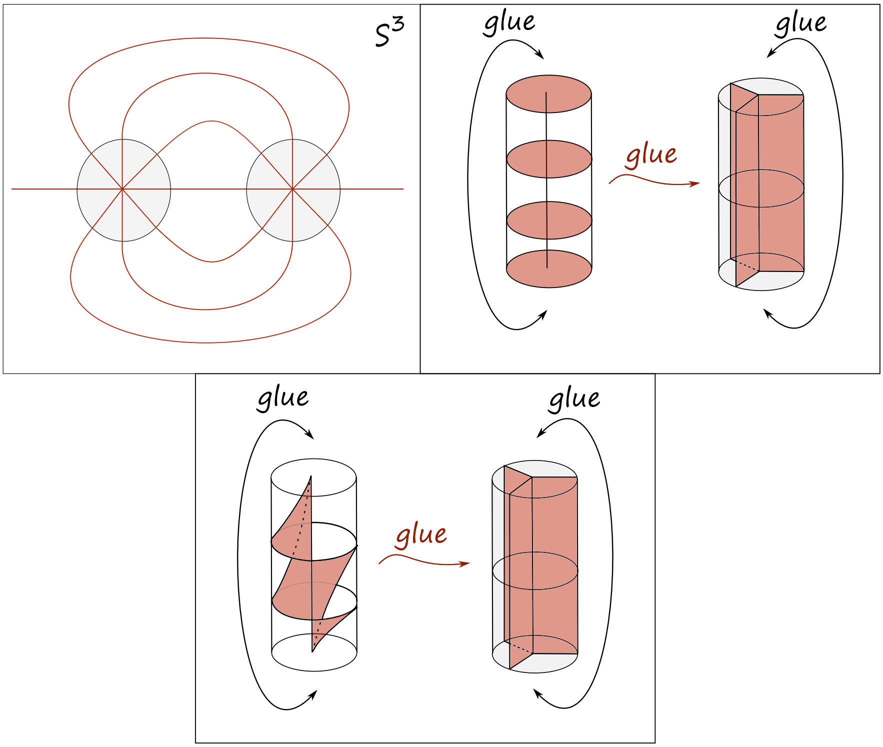


Figure 3. The disk-like pages of the trivial open book in S^3 (above) are obtained by gluing two foliations on two solid tori; similarly for its stabilized version (below), whose pages are annuli. Here, we use the genus 1 Heegaard splitting for S^3

book, an operation which does not change the manifold (see below). See Fig. 3.

- (Milnor fibrations) More generally, let $f : \mathbb{C}^2 \to \mathbb{C}$ be a polynomial which vanishes at the origin, and has no singularity in S^3 except perhaps the origin. Then, $\pi_f : S^3 \backslash B_f \to S^1$, $\pi_f(z_1, z_2) = \frac{f(z_1,z_2)}{|f(z_1,z_2)|}$, $B_f = \{f(z_1, z_2) = 0\} \cap S^3$, is an open book for S^3, called the Milnor fibration of the hypersurface singularity $(0, 0)$. The link B_f is the link of the singularity, and the binding of the open book, whereas the page is called the Milnor fiber. If f has no critical point at $(0, 0)$, then B_f is necessarily the unknot.
- We have $S^1 \times S^2 = \mathbf{OB}(\mathbb{D}^*S^1, \mathbb{1})$. This can be easily seen by removing the north and south poles of S^2 (whose S^1-fibers become the binding), and projecting the resulting manifold $\mathbb{D}^*S^1 \times S^1$ to the second factor.
- (Some lens spaces) We have $\mathbb{R}P^3 = \mathbf{OB}(\mathbb{D}^*S^1, \tau^2)$, as follows from taking the quotient of the stabilized open book in S^3 via the double cover $S^3 \to \mathbb{R}P^3$. More generally, for $p \geq 1$, we have $L(p, p-1) =$

$\mathbf{OB}(\mathbb{D}^*S^1, \tau^p)$, and for $p \leq 0$, $L(p,1) = \mathbf{OB}(\mathbb{D}^*S^1, \tau^p)$. Here, $L(p,q) = S^3/\mathbb{Z}_p$, is the lens space, where the generator $\zeta = e^{\frac{2\pi i}{p}} \in \mathbb{Z}_p$ acts via $\zeta \cdot (z_1, z_2) = (\zeta.z_1, \zeta^q.z_2)$. For $p = 0, 1, 2$, we recover the above examples.

In general, we have the following important result from smooth topology, which says that the open book construction achieves all closed, odd-dimensional manifolds:

Theorem E. (Alexander (dim $= 3$), Winkelnkemper (simply connected, dim ≥ 7), Lawson (dim ≥ 7), Quinn (dim ≥ 5)). *If M is closed and odd-dimensional, then M admits an open book decomposition.*

So far, we have discussed open books in terms of smooth topology. We now tie it with contact geometry, via the fundamental work of Emmanuel Giroux, which basically shows that contact manifolds can be studied from a purely topological perspective. One therefore usually speaks of the field contact *topology*, when the object of study is the contact manifold itself (as opposed, e.g., to a Reeb dynamical system on the contact manifold).

If M is oriented and endowed with an open book decomposition, then the natural orientation on the circle induces an orientation on the pages, which in turn induce the boundary orientation on the binding. The fundamental notion is the following:

Definition 2.13. (*Giroux*) Let (M, ξ) be an oriented contact manifold, and (π, B) an open book decomposition on M. Then, ξ is *supported* by the open book if one can find a positive contact form α for ξ (called a *Giroux form*), such that:

(1) $\alpha_B := \alpha|_B$ is a positive contact form for B;
(2) $d\alpha|_P$ is a positive symplectic form on the interior of every page P.

Here, the a priori orientations on binding and pages are the ones described above. Also, by a *positive* contact form, we mean a contact form α on M^{2n-1}, such that the orientation induced by the volume form $\alpha \wedge d\alpha^{n-1}$ coincides with the given orientation on M.

The above conditions are equivalent to:

(1)' $R_\alpha|_B$ is tangent to B;
(2)' R_α is positively transverse to the interior of every page.

In the above situation, $(B, \xi_B = \ker \alpha_B)$ is a codimension-2 contact submanifold, i.e., $\xi_B = \xi|_B$.

Theorem F. (Giroux [56]) *Every open book decomposition supports a unique isotopy class of contact structures. Any contact structure admits a supporting open book decomposition.*

Here, two contact structures are isotopic if they can be joined by a smooth path ξ_t of contact structures. An important result in contact geometry is *Gray's stability*, which says that isotopic contact structures are *contactomorphic*, i.e., there exists a diffeomorphism which carries one to the other. One may further assume that the pages in the above theorem are Stein manifolds, as discussed above. One may unequivocally use $\mathbf{OB}(P, \varphi)$ to denote the unique isotopy class of contact structures that this open book supports.

Giroux's result is actually much stronger in dimension 3, since it moreover states that the supporting open book is unique up to a suitable notion of *positive stabilization*, which can be thought of as two cancelling surgeries which therefore smoothly do not change the ambient manifold. This procedure consists of choosing a properly embedded path $l \subset P$ (a *stabilizing arc*) inside the surface P, attaching a 1-handle H along the attaching sphere $S^0 \cong \partial l \subset \partial P$, considering the loop γ obtained by gluing l with the core of H, and replacing the monodromy φ with $\varphi \circ \tau_\gamma$, where τ_γ is the right-handed Dehn twist along γ. In abstract notation

$$\mathbf{OB}(P, \varphi) \rightsquigarrow \mathbf{OB}(P \cup H, \varphi \circ \tau_\gamma).$$

The handle attachment on the page can be seen as an index 1 surgery on M, whereas composing with the monodromy adds a cancelling index 2 surgery, so that $\mathbf{OB}(P, \varphi) \cong \mathbf{OB}(P \cup H, \varphi \circ \tau_\gamma)$.

Theorem G. (Giroux's correspondence [56]) *If* $\dim(M) = 3$*, there is a 1:1 correspondence*

$$\{\textit{contact structures}\}/\textit{isotopy} \longleftrightarrow \{\textit{open books}\}/\textit{pos. stabilization.}$$

This bijection is why in dimension 3, one talks about Giroux's *correspondence*, which reduces the study of contact 3-manifolds to the topological study of open books. The analogous general uniqueness statement in higher dimensions is an open question to this day. Let us emphasize that in the above result, only the contact *structure* is fixed, and the contact form (and hence the dynamics) is auxiliary; Giroux's result is *not* dynamical, but rather topological/geometrical.

2.4. Global hypersurfaces of section

From a dynamical point of view, one wishes to adapt the underlying topology to the given dynamics, rather than vice versa. We therefore make the following:

Definition 2.14. Given a flow $\varphi_t : M \to M$ of an autonomous vector field on an odd-dimensional closed oriented manifold M carrying a concrete open book decomposition (π, B), we say that the open book is *adapted to the dynamics* if:

- B is φ_t-invariant;
- φ_t is positively transverse to the interior of each page;
- for each $x \in M \backslash B$ and P a page, then the orbit of x intersects the interior of P in the future, and in the past, i.e., there exists $\tau^+(x) > 0$ and $\tau^-(x) < 0$, such that $\varphi_{\tau^\pm(x)}(x) \in \mathrm{int}(P)$.

Note that the third condition actually follows from the second one, since we require it for every page and these foliate the complement of B. If φ_t is a Reeb flow, then the above is equivalent to asking that the (given) contact form is a Giroux form for the (auxiliary) open book. It follows from the definition that each page is a *global hypersurface of section*, defined as follows:

Definition 2.15. (*Global hypersurface of section*) A *global hypersurface of section* for an autonomous flow φ_t on a manifold M is a codimension-1 submanifold $P \subset M$, whose boundary (if non-empty) is flow-invariant, whose interior is transverse to the flow, such that the orbit of every point in $M \backslash \partial P$ intersects the interior of P in the future and past.

Poincaré return map Given a global hypersurface of section P for a flow φ_t, this induces a Poincaré return map, defined as

$$f : \mathrm{int}(P) \to \mathrm{int}(P), \; f(x) = \varphi_{\tau(x)}(x),$$

where $\tau(x) = \min\{t > 0 : \varphi_t(x) \in \mathrm{int}(P)\}$. This is clearly a diffeomorphism. And, by construction, periodic points of f (i.e., points p for which $f^k(p) = p$ for some $k \geq 1$) are in 1:1 correspondence with closed *spatial* orbits (those which are not fully contained in the binding).

Moreover, in the case of a Reeb dynamics, we have:

Proposition 2.16. *If φ_t is the Reeb flow of a contact form α, and P is a global hypersurface of section with induced return map f, then $\omega = d\alpha|_P = d\lambda$, with $\lambda = \alpha|_P$, is a symplectic form on int(P), and*

$$f : (int(P), \omega) \to (int(P), \omega)$$

is a symplectomorphism, i.e., $f^*\omega = \omega$.

In fact, f is an *exact* symplectomorphism, which means that $f^*\lambda = \lambda + d\tau$ for some smooth function τ (i.e., the return time). Differentiating this equation, we obtain $f^*\omega = \omega$. In dimension 2, a symplectic form is just an area form, and so the above proposition simply says that the return map is area-preserving.

The proof is quite simple: ω is symplectic precisely because the Reeb vector field, which spans the kernel of $\mathrm{d}\alpha$, is transverse to the interior of P (note, however, that it is degenerate at ∂P). For $x \in \mathrm{int}(P)$, $v \in T_xP$, we have

$$d_x f(v) = d_x\tau(v) R_\alpha(f(x)) + d_x\varphi_{\tau(x)}(v).$$

Using that φ_t satisfies $\varphi_t^*\alpha = \alpha$, we obtain

$$\begin{aligned}(f^*\lambda)_x(v) &= \alpha_{f(x)}(d_x f(v)) \\ &= d_x\tau(v) + (\varphi^*_{\tau(x)}\alpha)_x(v) \\ &= d_x\tau(v) + \lambda_x(v).\end{aligned} \tag{2.1}$$

Therefore

$$f^*\lambda = \mathrm{d}\tau + \lambda, \tag{2.2}$$

which proves the proposition.

Remark 2.17. In general, the return map might not necessarily extend to the boundary, and indeed, there are many examples on which this does not hold; this is a delicate issue which usually relies on analyzing the linearized flow equation along the normal direction to the boundary.

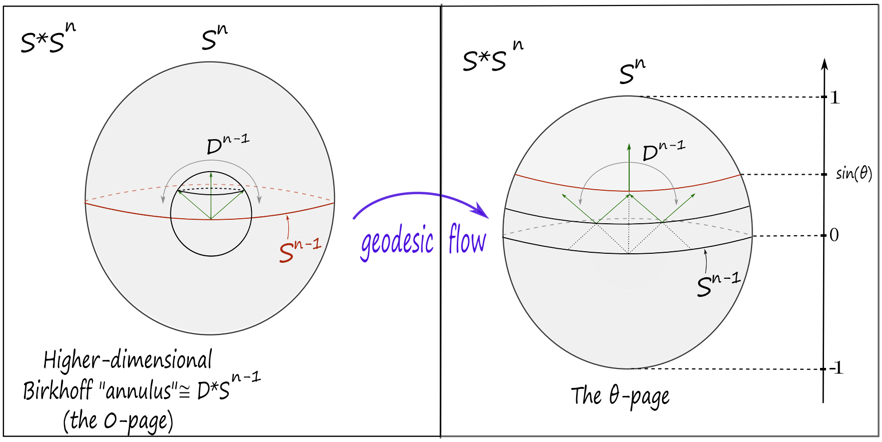


FIGURE 4. The geodesic open book for S^*S^n

2.5. Examples of adapted dynamics

Let us discuss two important but simple examples of open books supporting a Reeb dynamics.

Hopf flow The trivial open book on S^3, as well as its stabilized version, are both adapted to the Hopf flow.

Ellipsoids More generally, the trivial and stabilized open books on S^3 are adapted to the Reeb dynamics of every ellipsoid $E(a,b)$. In the trivial case, the return map on each page is the rotation by angle $2\pi\frac{a}{b}$; and in the stabilized case, we get a map of the annulus which rotates the two boundary components in the same direction (i.e., it is *not* a twist map).

2.6. Geodesic flow on S^n and the geodesic open book

We write

$$T^*S^n = \left\{(\xi,\eta) \in T^*\mathbb{R}^{n+1} = \mathbb{R}^{n+1} \oplus \mathbb{R}^{n+1} : \|\xi\| = 1,\ \langle \xi,\eta\rangle = 0\right\}.$$

The Hamiltonian for the geodesic flow is $Q = \frac{1}{2}\|\eta\|^2|_{T^*S^n}$ with Hamiltonian vector field

$$X_Q = \eta \cdot \partial_\xi - \xi \cdot \partial_\eta.$$

This is the Reeb vector field of the standard Liouville form λ_{std} on the energy hypersurface $\Sigma = Q^{-1}(\frac{1}{2}) = S^*S^n$. We have the invariant set

$$B := \{(\xi_0,\ldots,\xi_n;\eta_0,\ldots,\eta_n) \in \Sigma \mid \xi_n = \eta_n = 0\} = S^*S^{n-1}.$$

Define the circle-valued map

$$\pi_g : \Sigma\backslash B \longrightarrow S^1, \quad (\xi_0,\ldots,\xi_n;\eta_0,\ldots,\eta_n) \longmapsto \frac{\eta_n + i\xi_n}{\|\eta_n + i\xi_n\|}.$$

This is a concrete open book on S^*S^n, which we shall refer to as the *geodesic* open book. The page $\xi_n = 0$ and $\eta_n > 0$, i.e., the fiber over $1 \in S^1$, corresponds to a higher dimensional version of the famous *Birkhoff annulus* (when $n = 2$), and is a copy of $\mathbb{D}^*S^{n-1}$. Indeed, it consists of those (co)-vectors whose basepoint lies in the equator, and which point upwards to the upper hemisphere. See Fig. 4.

We then consider the angular form

$$\omega_g = d\pi_g = \frac{\eta_n d\xi_n - \xi_n d\eta_n}{\xi_n^2 + \eta_n^2}.$$

We see that $\omega_g(X_Q) = 1 > 0$, away from B. This means that (B, π_g) is a supporting open book for Σ and the pages of π_g are global hypersurfaces of section for X_Q. In fact, all of its pages are obtained from the Birkhoff annulus by flowing with the geodesic flow. In terms of the contact structure $\xi_{\text{std}} = \ker \lambda_{\text{std}}$, this open book corresponds to the abstract open book $(S^*S^n, \xi_{\text{std}}) = \mathbf{OB}(\mathbb{D}^*S^{n-1}, \tau^2)$ supporting ξ_{std}. Here, $\tau : \mathbb{D}^*S^{n-1} \to \mathbb{D}^*S^{n-1}$ is an exact symplectomorphism defined by Arnold in dimension 4 in [11] and extended by Seidel to higher dimensions (see, e.g., [115]), and is a generalization of the classical Dehn twist on the annulus. For $n = 2$, we reobtain the open book $\mathbb{R}P^3 = S^*S^2 = \mathbf{OB}(\mathbb{D}^*S^1, \tau^2)$.

2.7. Double cover of S^*S^2

We focus on $n = 2$, and consider

$$S^*S^2 = \{(\xi, \eta) \in T^*\mathbb{R}^3 : \|\xi\| = \|\eta\| = 1, \langle \xi, \eta \rangle = 0\},$$

the unit cotangent bundle of S^2, with canonical projection $\pi_0 : S^*S^2 \to S^2$, $\pi_0(\xi, \eta) = \xi$. It is easy to see that the map

$$\Phi : S^*S^2 \to SO(3),$$
$$\Phi(\xi, \eta) = (\xi, \eta, \xi \times \eta)$$

is a diffeomorphism, where we view ξ, η as column vectors, and so $S^*S^2 \cong SO(3) \cong \mathbb{R}P^3$. The projection π_0 on SO(3) becomes $\pi_0(A) = A(e_1)$, i.e., the first column of the matrix $A \in SO(3)$. We have $\pi_1(S^*S^2) = \mathbb{Z}_2$, generated by the S^1-fiber. By definition, the double cover of SO(3) is the Spin group Spin(3), which can be constructed as follows. Consider the quaternions

$$\mathbb{H} = \{a + bi + cj + dk : a, b, c, d \in \mathbb{R}\},$$

with $i^2 = j^2 = k^2 = -1$, $ij = k, jk = i, ki = j$. We identify $S^3 = Sp(1) := \{q \in \mathbb{H} : \|q\| = 1\}$, and $\mathbb{R}^3 = \mathrm{Im}(\mathbb{H}) = \langle i, j, k \rangle$ the set of purely imaginary quaternions. The conjugate of $q = a + bi + cj + dk$ is $\overline{q} = a - bi - cj - dk$. We then define

$$p : S^3 \to SO(3),$$
$$p(q)(v) = \overline{q} v q,$$

where $v \in \mathrm{Im}(\mathbb{H}) = \mathbb{R}^3$. We have $\|\overline{q} v q\| = \|q\|^2 \|v\| = \|v\|$, and $p(q)$ is seen to preserve orientation, so indeed $p(q) \in \mathrm{SO}(3)$. Clearly, $p(-q) = p(q)$, and the map p is in fact a double cover, so that $S^3 = \mathrm{Spin}(3)$.

Identifying i with e_1, we have $\pi_0(p(q)) = p(q)(i) = \overline{q} i q$. A short computation gives

$$\overline{q} i q = (a + bi + cj + dk)^* i (a + bi + cj + dk) = (a^2 + b^2 - c^2 - d^2) i$$

$$+2(bc - ad)j + 2(ac + bd)k.$$

On the other hand, the Hopf map may be defined as the map

$$\pi : S^3 \to S^2,\ \pi(z_1, z_2) = (|z_1|^2 - |z_2|^2, 2\mathrm{Re} z_1\overline{z_2}, 2\mathrm{Im} z_1\overline{z_2}),$$

where we view $S^3 = \{(z_1, z_2) \in \mathbb{C}^2; |z_1|^2 + |z_2|^2 = 1\}$ and $S^2 \subset \mathbb{R}^3$. Writing $q = a + bi + cj + dk = z_1 + z_2 j$, i.e. $z_1 = a + ib$, $z_2 = c + id$, one can easily check that

$$(|z_1|^2 - |z_2|^2, 2\mathrm{Re} z_1\overline{z_2}, 2\mathrm{Im} z_1\overline{z_2}) = (a^2 + b^2 - c^2 - d^2, 2(bc - ad), 2(ac + bd)).$$

We have proved the following:

Proposition 2.18. *The Hopf fibration is the fiber-wise double cover of the canonical projection π_0, i.e., we have a commutative diagram*

$$\begin{tikzcd}
S^1 \arrow[r, "z\mapsto z^2"] \arrow[d, hook] & S^1 \arrow[d, hook] \\
S^3 = Spin(3) \arrow[r, "p"] \arrow[d, "\pi"] & SO(3) = S^*S^2 \arrow[d, "\pi_0"] \\
S^2 \arrow[r, equal] & S^2
\end{tikzcd}$$

2.8. Magnetic flows and quaternionic symmetry

On this section, we expose the beautiful construction of [8] (to which we refer the reader for further details here omitted), relating the quaternions with Reeb flows on S^3, as double covers of magnetic flows on S^*S^2.

On S^2, consider an area form σ (the *magnetic field*), and the *twisted* symplectic form ω_σ, defined on T^*S^2 via

$$\omega_\sigma = \omega_{\mathrm{std}} - \pi_0^*\sigma,$$

where $\pi_0 : T^*S^2 \to S^2$ is the natural projection. Fixing a metric g on S^2, the Hamiltonian flow of the kinetic Hamiltonian $H(q,p) = \frac{\|p\|^2}{2}$, computed with respect to ω_σ, is called the *magnetic flow* of (g, σ). Note that $\sigma = 0$ corresponds to the geodesic flow of g. Physically, the magnetic flow models the motion of a particle on S^2 subject to a magnetic field (the terminology comes from Maxwell's equations, which can be recast in this language). From now on, we fix σ to be the standard area form on S^2, with total area 4π, and g the standard metric with constant Gaussian curvature 1.

On S^*S^2, we can choose a connection 1-form α satisfying $d\alpha = \pi^*\sigma$, which is a contact form (usually called a *prequantization form*). We identify $T^*S^2 \backslash S^2$ with $\mathbb{R}^+ \times S^*S^2$, and denoting by $r \in \mathbb{R}^+$ the radial coordinate, we have the associated symplectization form $d(r\alpha)$. Consider the S^1-family of symplectic forms

$$\omega_\theta = \cos\theta\ d(r\alpha) + \sin\theta\ d(r\alpha_{\mathrm{std}}),\ \theta \in \mathbb{R}/2\pi\mathbb{Z},$$

defined on $\mathbb{R}^+ \times S^*S^2 = T^*S^2\backslash S^2$, where $d(r\alpha_{\text{std}}) = \omega_{\text{std}}$. The Hamiltonian flow of the kinetic Hamiltonian H, with respect to ω_θ, and along $r = 1$, is easily seen to be the magnetic flow of $(g, -\cot\theta\cdot\sigma)$ up to constant reparametrization. In particular, for $\theta = \pi/2 \mod \pi$, we obtain the geodesic flow, whose orbits are great circles; for other values of θ, the strength of the magnetic field increases, and the orbits become circles of smaller radius with an increasing left drift. For $\theta = 0 \mod \pi$, the circles become points and the flow rotates the fibers of S^*S^2, i.e., this is the magnetic flow with "infinite" magnetic field.

We now construct the double covers of these magnetic flows on S^3, using the hyperkähler structure on $\mathbb{H} = \mathbb{R}^4 = \mathbb{C}^2$. We view S^3 as the unit sphere in $\mathbb{H}$. Every unit vector

$$c = c_1 i + c_2 j + c_2 k \in S^2 \subset \mathbb{R}^3$$

may be viewed as a complex structure on $\mathbb{H}$, i.e., $c^2 = -\mathbb{1}$. Denoting the radial coordinate on $\mathbb{R}^4$ by r, we obtain an S^2-family of contact forms on S^3 given by

$$\alpha_c = -2\mathrm{d}r \circ c|_{TS^2},\ c \in S^2.$$

The Reeb vector field of α_c is $R_c = \frac{1}{2}c\partial_r$. Note that α_i is the standard contact form on S^3, whose Reeb orbits are the Hopf fibers.

We then consider the quaternionic action of S^3 on itself, given by

$$\begin{aligned} l_a &: S^3 \to S^3 \\ u &\mapsto au, \end{aligned}$$

for $a \in S^3$. Recall that we also have the action of S^3 on S^2 via the $SO(3)$-action of the previous section, i.e., $a \cdot c = p(a)(c) = ac\overline{a} \in S^2$, for $a \in S^3$, $c \in S^2$, and $p : S^3 \to SO(3)$ the spin group double cover. One checks directly that $(l_a)_*\alpha_c = \alpha_{ac\overline{a}} = \alpha_{a\cdot c}$. In particular, $(l_a)_*\alpha_i = \alpha_{\pi(a)}$, where π is the Hopf fibration.

On the other hand, the stabilizer of $i \in S^2$ under the S^3-action is the circle

$$\mathrm{Stab}(i) = \{\cos(\varphi) + i\sin(\varphi) : \varphi \in S^1\} \cong S^1 \subset S^3.$$

The action of an element in this subgroup on S^3 then fixes α_i, but reparametrizes its Reeb orbits, i.e., rotates the Hopf fibers. We then consider an S^1-subgroup $\{a_\theta\} \subset S^3$ of unit quaternions which are transverse to this stabilizer, intersecting it only at the identity, given by

$$a_\theta = \cos(\theta/2) + k\sin(\theta/2),\ \theta \in [0, \pi]$$

for which

$$\pi(a_\theta) = a_\theta i \overline{a}_\theta = i\cos\theta + j\sin\theta.$$

Define

$$\alpha_\theta := \alpha_{\pi(a_\theta)} = \cos\theta\ \alpha_i + \sin\theta\ \alpha_j,$$

with Reeb vector field $R_\theta := R_{\pi(a_\theta)}$. One further checks that

$$\alpha_\theta = p^*(\cos\theta\ \alpha + \sin\theta\ \alpha_{\text{std}}),$$

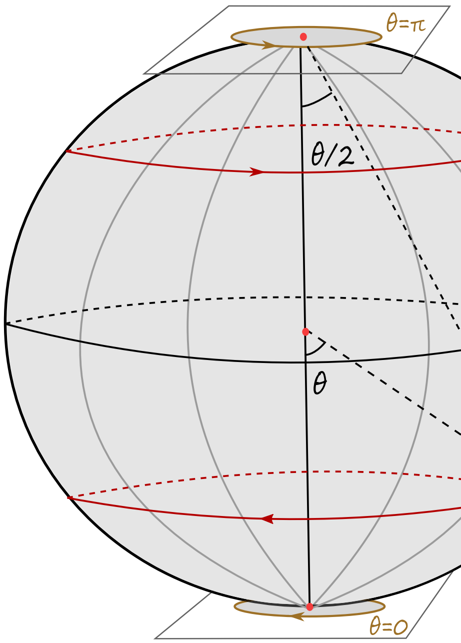


Figure 5. The binding of the magnetic open book $\overline{p}_\theta$ (in red), consisting of two circles of latitude θ and $\pi - \theta$, doubly covered by two Reeb orbits of α_θ. At $\theta = \pi$, the action of a_π maps the Hopf fiber over a point to the Hopf fiber over its antipodal (cf. [8, Fig. 1])

and so

$$\widetilde{\omega}_\theta := \mathrm{d}\alpha_\theta = p^*\omega_\theta|_{S^*S^2}$$

is the double cover of the twisted symplectic form ω_θ along the unit cotangent bundle (alternatively, we can also think of $\widetilde{\omega}_\theta$ as being defined on $\mathbb{R}^4 \backslash \{0\} = \mathbb{R}^+ \times S^3$ as the symplectization of α_θ). We have obtained:

Theorem H. [8] *There are contact forms α_i, α_j and an S^1-action on S^3, sending α_i to contact forms $\alpha_\theta = \cos\theta\ \alpha_i + \sin\theta\ \alpha_j$, $\theta \in S^1$, such that the Reeb flow of α_θ doubly covers the magnetic flow of ω_θ.*

Remark 2.19. Note that for $\theta = 0$, corresponding to the infinite magnetic flow, this reduces to the statement of Proposition 2.18. For $\theta = \pi/2$, this says that we can lift the geodesic flow on S^2 to (a rotated version of) the Hopf flow. Of course, this statement depends on choices; we could have arranged that the lift is precisely the Hopf flow by changing our choice of coordinates.

2.9. The magnetic open book decompositions

We now tie the previous discussion with open book decompositions. We have seen that the geodesic open book on S^*S^2 is constructed in such a way that it is adapted to the geodesic flow of the round metric. On the other hand, by considering the action on S^3 of the subgroup $\{a_\theta\} \subset S^3$ of the previous section, we obtain an S^1-family $\{p_\theta : S^3 \backslash a_\theta(L) \to S^1\}$ of open book decompositions on S^3 (here, L is the Hopf link). These are, respectively, adapted to the Reeb dynamics of α_θ, and start from the stabilized open book p_0 on S^3 (adapted to α_i by the example discussed above); they are all just rotations of each other.

Note that Proposition 2.18, the push-forward of p_0 under the Hopf map, i.e. $\overline{p}_0 := \pi_*(p_0) = p_0 \circ \pi^{-1} : S^*S^2 \backslash B_0 \to S^1$ where B_0 is the disjoint union of the unit cotangent fibers over the north and south poles N, S in S^2 (i.e., the image of the Hopf link under π), is adapted to the infinite magnetic flow. The pages are cylinders obtained as follows: $S^*S^2 \backslash B_0 \cong ((-1,1) \times S^1) \times S^1$ is a trivial bundle over $S^2 \backslash \{N, S\} \cong (-1,1) \times S^1$ (the Euler class of S^*S^2 is -2), and $\overline{p}_0$ is the trivial fibration.

The push-forward $\overline{p}_\theta = \pi_*(p_\theta) : S^*S^2 \backslash B_\theta \to S^1$ is then an open book decomposition on S^*S^2, which coincides with the geodesic open book at $\theta = \pi$. The binding B_θ consists of two magnetic geodesics for ω_θ; see Fig. 5. We call any element of the family $\{\overline{p}_\theta\}$, a *magnetic open book decomposition.*

Digression: open books and Heegaard splittings A 3-dimensional genus g (orientable) handlebody H_g is the 3-manifold with boundary resulting by taking the boundary connected sum of g copies of the solid 2-torus $S^1 \times \mathbb{D}^2$ (here, we set $H_0 = B^3$ the 3-ball). H_g can also be obtained by attaching a sequence of g 1-handles to B^3. Its boundary is Σ_g, the orientable surface of genus g. A *Heegaard splitting* of genus g of a closed 3-manifold X is a decomposition

$$X = H_g \bigcup_f H_g',$$

where $f : \Sigma_g = \partial H_g \to \Sigma_g = \partial H_g'$ is a homeomorphism of the boundary of two copies of H_g. The surface Σ_g is called the splitting surface. Different choices of f in the mapping class group of Σ_g give, in general, different 3-manifolds. In fact, it is a fundamental theorem of 3-dimensional topology that every closed 3-manifold admits a Heegaard splitting. We have also touched upon another structural result for 3-manifolds: namely, that every closed 3-manifold admits an open book decomposition. Let us then discuss how to induce a Heegaard splitting from an open book.

Starting from a concrete open book decomposition $M \backslash B \to S^1 = \mathbb{R}/\mathbb{Z}$ of abstract type $M = \mathbf{OB}(P, \varphi)$, we obtain a Heegaard splitting via

$$H_g = \pi^{-1}([0, 1/2]) \cup B, \; H_g' = \pi^{-1}([1/2, 1]) \cup B,$$

where the splitting surface $\Sigma_g = P_0 \cup_B P_{1/2}$ is the double of the page $P_0 = \pi^{-1}(0)$, obtained by gluing P_0 to its "opposite" $P_{1/2} = \pi^{-1}(1/2)$. The gluing map f is simply given by φ on P_0, and the identity on $P_{1/2}$. Stabilizing the open book translates into a stabilization of the Heegaard splitting.

This shows that the Heegaard diagram thus induced is rather special, since the gluing map is trivial on "half" of the splitting surface. In fact, not every Heegaard splitting arises this way, as is easy to see (e.g., the lens spaces are precisely the 3-manifolds with Heegaard splittings of genus 1, but only the

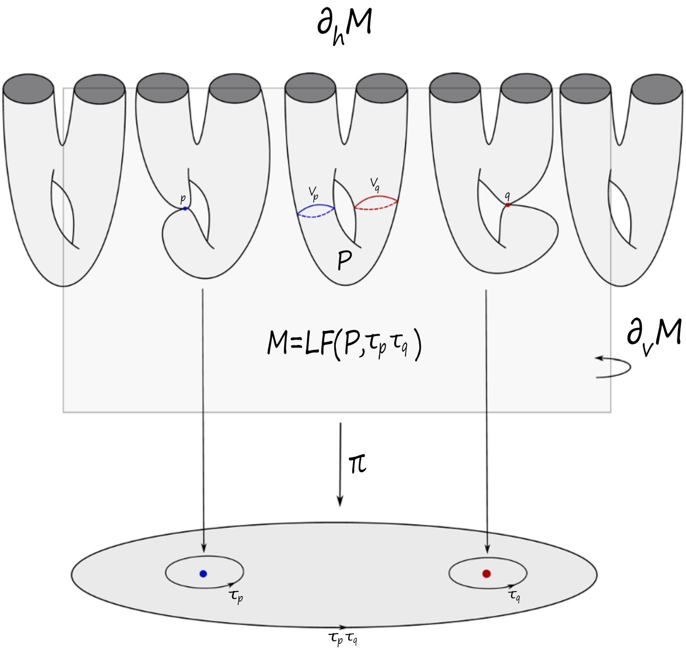


FIGURE 6. The Lefschetz fibration $\mathbf{LF}(P, \tau_p\tau_q)$ over $\mathbb{D}^2$

lens spaces discussed in Example 2.12 arise from an open book with annulus page, since its relative mapping class group is generated by the Dehn twist).

Digression: open books and Lefschetz fibrations/pencils We now explore some further interplay between symplectic and algebraic geometry.

Definition 2.20. (*Lefschetz fibration*) Let M be a compact, connected, oriented, smooth 4-manifold with boundary. A *Lefschetz fibration* on M is a smooth map $\pi : M \to S$, where S is a compact, connected, oriented surface with boundary, such that each critical point p of π lies in the interior of M and has a local complex coordinate chart $(z_1, z_2) \in \mathbb{C}^2$ centered at p (and compatible with the orientation of M), together with a local complex coordinate z near $\pi(p)$, such that $\pi(z_1, z_2) = z_1^2 + z_2^2$ in this chart.

In other words, each critical point has a local (complex) Morse chart, and is therefore non-degenerate. We then have finitely many critical points due to compactness of M. One may also (up to perturbation of π) assume that there is a single critical point on each fiber of π. The regular fibers are connected oriented surfaces with boundary, whereas the singular fibers are immersed oriented surfaces with a transverse self-intersection (or node). This singularity is obtained from nearby fibers by pinching a closed curve (the *vanishing cycle*) to a point. See Fig. 6.

The boundary of a Lefschetz fibration splits into two pieces

$$\partial M = \partial_h M \cup \partial_v M,$$

where

$$\partial_h M = \bigcup_{b \in S} \partial \pi^{-1}(b), \quad \partial_v M = \pi^{-1}(\partial S).$$

By construction, $\partial_h M$ is a circle fibration over S, and $\partial_v M$ is a surface fibration over ∂S. If we focus on the case $S = \mathbb{D}^2$, the two-disk, denoting the regular fiber P and $B = \partial P$, we necessarily have that $\partial_h M$ is trivial as a fibration, and $\partial_v M$ is the mapping torus P_ϕ of some monodromy $\phi : P \to P$. Therefore

$$\partial M = \partial_h M \cup \partial_v M = B \times \mathbb{D}^2 \bigcup P_\phi = \mathbf{OB}(P, \phi).$$

Now, the monodromy ϕ is not arbitrary, since orientations here play a crucial role (Fig. 7). While every element in the symplectic mapping class group of a surface is a product of powers of Dehn twists along some simple closed loops, it turns out that ϕ is necessarily a product of *positive* powers of Dehn twists (once orientations are all fixed). In fact, $\phi = \prod_{p \in \mathrm{crit}(\pi)} \tau_p$, where $\tau_p = \tau_{V_p}$ is the positive (or right-handed) Dehn twist along the corresponding vanishing cycle $V_p \cong S^1 \subset P$. This can be algebraically encoded via the monodromy representation

$$\rho : \pi_1(\mathbb{D}^2 \backslash \mathrm{critv}(\pi)) \to \mathrm{MCG}(P, \partial P),$$

where $\mathrm{critv}(\pi) = \{x_1, \ldots, x_n\}$, $x_i = \pi(p_i)$, is the finite set of critical values of π. We have

$$\pi_1(\mathbb{D}^2 \backslash \{x_1, \ldots, x_n\}) = \left\langle g_\partial, g_1, \ldots, g_n : g_\partial = \prod_{i=1}^n g_i \right\rangle,$$

where g_i is a small loop around x_i and $g_\partial = \partial \mathbb{D}^2$, and ρ is defined via $\rho(g_i) = \tau_{V_{p_i}}$.

Reciprocally, a 4-dimensional Lefschetz fibration on M over $\mathbb{D}^2$ is abstractly determined by the data of the regular fiber P (a surface with non-empty boundary) and a collection of simple closed loops $V_1, \ldots, V_n \subset P$. This determines a monodromy $\phi = \prod_{i=1}^n \tau_{V_i}$, a product of positive Dehn twists along the vanishing cycles V_i. The recipe to construct M works as follows: decompose $P = \mathbb{D}^2 \bigcup H_1 \cup \cdots \cup H_k$ into a handle decomposition with a single 0-handle $\mathbb{D}^2$ and a collection of 2-dimensional 1-handles $H_1, \ldots, H_k \cong \mathbb{D}^1 \times \mathbb{D}^1$. One starts with the trivial Lefschetz fibration $M_0 = \mathbb{D}^2 \times \mathbb{D}^2 \to \mathbb{D}^2$ with disk fiber, and then, one attaches (thickened) 4-dimensional 1-handles $H_i \times \mathbb{D}^2$ to M_0 to obtain the trivial Lefschetz fibration $M_1 = P \times \mathbb{D}^2 \to \mathbb{D}^2$ with fiber P. To add the singularities, one attaches one 4-dimensional 2-handle $H = \mathbb{D}^2 \times \mathbb{D}^2$ along $V_i \subset P \times \{1\} \subset \partial M_1$, viewed as the attaching sphere $V_i = S^1 \times \{0\} \subset S^1 \times \mathbb{D}^2 \subset \partial H$. At each step of the 2-handle attachments, we obtain a fibration with monodromy representation ρ_i extending ρ_{i-1} and satisfying $\rho_i(g_i) = \tau_{V_i}$, starting from the trivial representation $\rho_0 = \mathbb{1} : \pi_1(\mathbb{D}^2) = \{1\} \to \mathrm{MCG}(P, \partial P)$. We denote the resulting manifold as $M = \mathbf{LF}(P, \phi)$, for which we have a handle description with handles of index $0, 1, 2$.

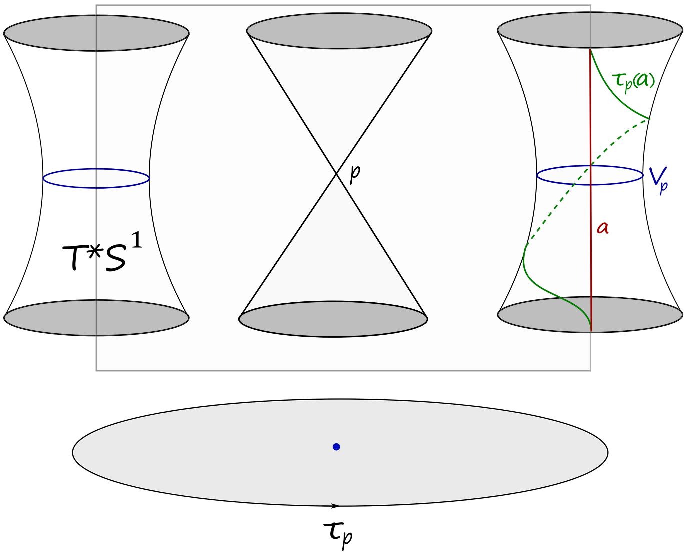


Figure 7. The local model for a Lefschetz singularity

Remark 2.21. The notation $\mathbf{LF}(P, \phi)$, although simple, is a bit misleading: we need to remember the factorization of ϕ, since different factorizations lead in general to different smooth 4-manifolds. One should perhaps use $\mathbf{LF}(P; V_1, \ldots, V_n)$ instead, although we hope that this will not lead to confusion.

Having said that, we summarize this discussion in the following:

Lemma 2.22. (Relationship between Lefschetz fibrations and open books) *We have*

$$\partial \mathbf{LF}(P, \phi) = \mathbf{OB}(P, \phi),$$

for $\phi = \prod_{i=1}^n \tau_{V_i}$ a product of positive Dehn twists along a collection of vanishing cycles $V_1, \ldots, V_n$ in P.

While so far this has been a discussion in the smooth category, one may upgrade this to the symplectic/contact category. While we have seen that open books support contact structures in the sense of Giroux, Lefschetz fibrations also support symplectic structures. This is encoded in the following:

Definition 2.23. (*Symplectic Lefschetz fibrations*) An (exact) *symplectic* Lefschetz fibration on an exact symplectic 4-manifold $(M, \omega = d\lambda)$ is a Lefschetz fibration π for which the vertical and horizontal boundary are convex, and the fibers $\pi^{-1}(b)$ are symplectic with respect to ω, also with convex boundary.

Here, convexity means that the Liouville vector field is outwards pointing. Note that, by Stokes's theorem and exactness of ω, a symplectic Lefschetz fibration cannot have contractible vanishing cycles, since otherwise there would be a non-constant symplectic sphere in a fiber. The description of Lefschetz fibrations in terms of handle attachments can also be upgraded to the sympectic category via the notion of a *Weinstein handle*. After smoothing out the corner $\partial_h M \cap \partial_v M$, the boundary ∂M becomes contact-type via $\alpha = \lambda|_{\partial M}$, and the contact structure $\xi = \ker \alpha$ is supported by the open book at the boundary. The contact manifold $(\partial M, \xi)$ is said to be *symplectically filled* by (M, ω) (see the discussion below on symplectic fillings of contact manifolds).

Since the space of symplectic forms on a two-manifold is convex and hence contractible, one can show that, given the Lefschetz fibration $\mathbf{LF}(P, \phi)$, an *adapted* symplectic form (i.e., as in the definition above) exists and is unique up to symplectic deformation. Therefore, similarly as in Giroux's correspondence, one can talk about $\mathbf{LF}(P, \phi)$ as a symplectomorphism class of symplectic manifolds.

Example 2.24. An example which is relevant for the spatial CR3BP is that of T^*S^2. We consider the *Brieskorn variety*

$$V_\epsilon = \left\{ (z_0, \dots, z_n) \in \mathbb{C}^{n+1} : \sum_{j=0}^{n} z_j^2 = \epsilon \right\},$$

and the associated *Brieskorn manifold* $\Sigma_\epsilon = V_\epsilon \cap S^{2n+1}$. If $\epsilon = 0$, V_0 has an isolated singularity at the origin, and Σ_0 is called the *link of the singularity*. For $\epsilon \neq 0$, the domain $V_\epsilon^{\mathrm{cpt}} = V_\epsilon \cap B^{2n+2}$ is a smooth manifold, with boundary $\Sigma_\epsilon \cong \Sigma_0$; the manifold V_ϵ also inherits a symplectic form by restriction of ω_{std} on $\mathbb{C}^{n+1}$. Similarly, Σ_ϵ inherits a contact form by restriction of the standard contact form $\alpha_{\mathrm{std}} = i \sum_j z_j \mathrm{d}\overline{z}_j - \overline{z}_j \mathrm{d}z_j$. In fact, V_ϵ is a Stein manifold, and $V_\epsilon^{\mathrm{cpt}}$ is a Stein filling of Σ_ϵ; see the discussion on Stein manifolds above and fillings below.

A standard fact is the following: the map

$$(V_1, \omega_{\mathrm{std}}) \to (T^*S^n \subset T^*\mathbb{R}^{n+1}, \omega_{\mathrm{can}}),\ z = q + ip \mapsto (\|q\|^{-1} q, \|q\| p)$$

is a symplectomorphism, which restricts to a contactomorphism

$$(\Sigma_0, \alpha_{\mathrm{std}}) \to (S^*S^n \subset T^*\mathbb{R}^{n+1}, \lambda_{\mathrm{can}}).$$

The standard Lefschetz fibration on T^*S^n can be obtained from the Brieskorn variety model as

$$V_1 \to \mathbb{C},\ (z_0, \dots, z_n) \mapsto z_0.$$

This induces the geodesic open book on S^*S^n at the boundary, given by the same formula.

The above map induces the Lefschetz fibration $T^*S^2 = \mathbf{LF}(T^*S^1, \tau^2)$, where τ is the Dehn twist along the vanishing cycle $S^1 \subset T^*S^1$, the zero section. We conclude again that $S^*S^2 = \mathbb{R}P^3 = \mathbf{OB}(\mathbb{D}^*S^1, \tau^2)$. See Fig. 8.

To tie the above discussion with classical algebraic geometry, we introduce the following notion (in the closed case):

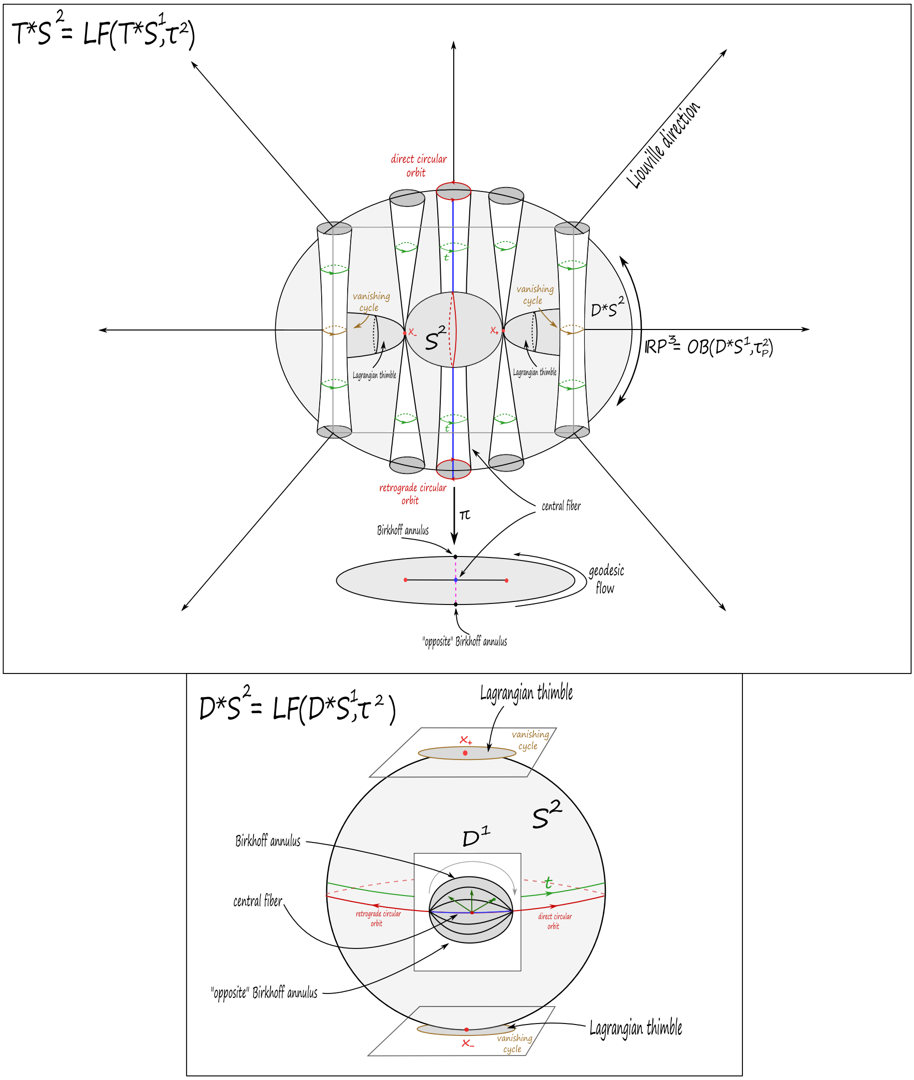


Figure 8. The standard Lefschetz fibration on $\mathbb{D}^*S^2 = \mathbf{LF}(\mathbb{D}^*S^1, \tau^2)$, where τ is the Dehn twist along the zero section $S^1 \subset \mathbb{D}^*S^1$. In the picture above, we draw T^*S^2, and the fibers on $\mathbb{D}^*S^2$ are obtained by projecting along the Liouville direction. These are drawn in the picture below. The two critical points induce the monodromy τ^2. We call the equators transversed in both directions the direct/retrograde (circular) orbits, for reasons that will become apparent

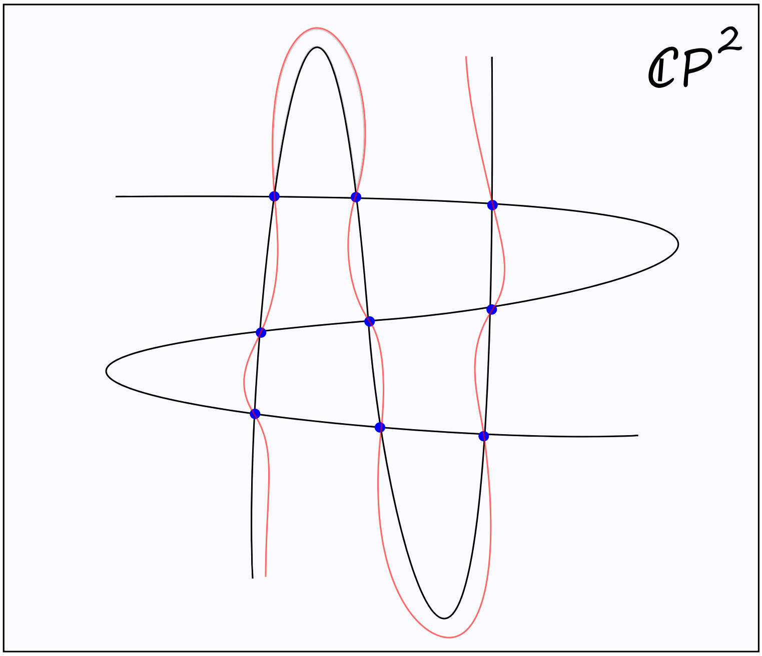


Figure 9. A cartoon of a pencil of cubics, where L consists of 9 points, and each fiber has genus 1

Definition 2.25. (*Lefschetz pencil*) Let M be a closed, connected, oriented, smooth 4-manifold. A *Lefschetz pencil* on M is a Lefschetz fibration $\pi : M\backslash L \to \mathbb{C}P^1$, where $L \subset M$ is a finite collection of points, such that near each base point $p \in L$ there exists a complex coordinate chart (z_1, z_2) in which π looks like the Hopf map $\pi(z_1, z_2) = [z_1 : z_2]$.

Lefschetz pencils arise naturally in the study of projective varieties, and linear systems of line bundles over them (Fig. 9). The basic construction is the following: Consider two distinct homogeneous polynomials $F(x, y, z), G(x, y, z)$ of degree d in projective coordinates $[x : y : z] \in \mathbb{C}P^2$ (i.e., sections of the holomorphic line bundle $\mathcal{O}(d)$), generic in the sense that $V(F) = \{F = 0\}$ and $V(G) = \{G = 0\}$ are smooth degree d curves, of genus $g = \frac{(d-1)(d-2)}{2}$ by the genus-degree formula, and so that the base locus $V(F) \cap V(G) = L$ consists of a collection of d^2 distinct points (by Bézout's theorem). Consider the *degree d pencil* $\{C_{[\lambda:\mu]}\}_{[\lambda:\mu]\in\mathbb{C}P^1}$, where

$$C_{[\lambda:\mu]} = V(\lambda F + \mu G) \subset \mathbb{C}P^2.$$

Through any point in $\mathbb{C}P^2\backslash L$, there is a unique $C_{[\lambda:\mu]}$ which contains it. We then have a Lefschetz pencil

$$\pi : \mathbb{C}P^2\backslash L \to \mathbb{C}P^1,$$

where $\pi([x : y : z]) = [\lambda : \mu]$ if $C_{[\lambda:\mu]}$ is the unique degree d curve in the family passing through $[x : y : z]$.

By construction, every curve in the pencil meets at the d^2 points in L. One can further perform a complex blow-up along each of these points, by adding an exceptional divisor (a copy of $\mathbb{C}P^1$) of all possible incoming directions at a given point, and the result is a Lefschetz fibration

$$Bl_L\pi : Bl_L\mathbb{C}P^2 \to \mathbb{C}P^1.$$

By construction, this Lefschetz fibration has plenty of spheres, i.e., the exceptional divisors, which are sections of the fibration.

The above construction also extends to the case of closed 4-dimensional projective varieties in some ambient projective space. Moreover, as we have already mentioned, projective varieties are Kähler, and in particular symplectic. It is a very deep fact that the above construction extends beyond the algebraic case to the general case of *all* closed symplectic 4-manifolds:

Theorem I. (Donaldson [34]) *Any closed symplectic* 4*-manifold* (M, ω) *admits Lefschetz pencils with symplectic fibers. In fact, if* $[\omega] \in H^2(M;\mathbb{Z})$ *is integral, the fibers are Poincaré dual to* $k[\omega]$ *for some sufficiently large* $k \gg 0$.

The above implies that techniques from algebraic geometry can also be applied in the symplectic category, and the interplay is very rich. From the above discussion, after blowing up a finite number of points on the given closed symplectic 4-manifold (M, ω), we obtain a Lefschetz fibration.

Digression: symplectic cobordisms and fillings We have already seen the fundamental relationship between contact and symplectic geometry. We now touch upon this a bit further.

Definition 2.26. (*Symplectic cobordism*) A (strong) symplectic cobordism from a closed contact manifold (X_-, ξ_-) to a closed contact manifold (X_+, ξ_-) is a compact symplectic manifold (M, ω) satisfying:

- $\partial M = X_+ \bigsqcup X_-$;
- $\omega = d\lambda_\pm$ is exact near $X_\pm$, and the (local) Liouville vector field $V_\pm$ (defined via $i_{V_\pm}\omega = \lambda_\pm$) is inwards pointing along X_- and outwards pointing along X_+;
- $\ker \lambda_\pm|_{X_\pm} = \xi_\pm$.

If $\omega = d\lambda$ is globally exact and the Liouville vector field is outwards/ inwards pointing along $X_\pm$, we say that (M, ω) is a *Liouville* cobordism. The boundary component X_+ is called *convex* or *positive*, and X_-, *concave* or *negative*. Note that a symplectic cobordism is *directed*; in general, there might be such a cobordism from X_- to X_+ but not vice versa. In fact, the relation $(X_-, \xi_-) \preceq (X_+, \xi_+)$ whenever there exists a symplectic cobordism as above, is reflexive, transitive, but *not* symmetric. We remark that the opposite convention on the choice of *to* and *from* is also used in the literature.

Definition 2.27. (*Symplectic filling/Liouville domain*) A (strong, Liouville) symplectic filling of a contact manifold (X, ξ) is a (strong, Liouville) compact symplectic cobordism from the empty set to (X, ξ). A Liouville filling is also called a *Liouville domain.*

The *Liouville manifold* associated with a Liouville domain (M, ω) is its *Liouville completion*, obtained by attaching a cylindrical end

$$(\widehat{M}, \widehat{\omega} = d\widehat{\lambda}) = (M, \omega = d\lambda) \cup_{\partial M} ([1, +\infty) \times \partial M, d(r\alpha)),$$

where $\alpha = \lambda|_{\partial M}$ is the contact form at the boundary. Liouville manifolds are therefore "convex at infinity".

It is a fundamental question of contact topology whether a contact manifold is fillable or not, and, if so, how many fillings it admits (say, up to symplectomorphism, diffeomorphism, homeomorphism, homotopy equivalence, s-cobordism, h-cobordism,…). Note that, given a filling, one may choose to perform a symplectic blow-up in the interior, which does not change the boundary but changes the symplectic manifold; to remove this trivial ambiguity, one usually considers *symplectically aspherical* fillings, i.e., symplectic manifolds (M, ω) for which $[\omega]|_{\pi_2(M)} = 0$ (this holds if, e.g., ω is exact, as the case of a Liouville filling).

For example, the standard sphere $(S^{2n-1}, \xi_{\text{std}})$ admits the unit ball $(B^{2n}, \omega_{\text{std}})$ as a Liouville filling. A fundamental theorem of Gromov [59, p. 311] says that this is unique (strong, symplectically aspherical=:ssa) filling up to symplectomorphism in dimension 4; this is known up to diffeomorphism in higher dimensions by a result of Eliashberg–Floer–McDuff [94], but unknown up to symplectomorphism. This was generalized to the case of *subcritically* Stein fillable contact manifolds in [14]. Another example is a unit cotangent bundle (S^*Q, ξ_{std}), which admits the standard Liouville filling $(\mathbb{D}^*Q, \omega_{\text{std}})$. There are known examples of manifolds Q with (S^*Q, ξ_{std}) admitting only one ssa filling up to symplectomorphism (e.g., $Q = \mathbb{T}^2$, [126]; if $n \geq 3$ and $Q = \mathbb{T}^n$, this also holds up to diffeomorphism [21,51]), but there are other examples with non-unique ssa fillings which are not blowups of each other (e.g., $Q = S^n$, $n \geq 3$ [107]). See also [87,88,117]. The literature on fillings is vast (especially in dimension 3) and this list is by all means non-exhaustive.

Remark 2.28. There are also other notions of symplectic fillability: weak, Stein, Weinstein …which we will not touch upon. The set of contact manifolds admitting a filling of every such type is related via the following inclusions:

$$\{\text{Stein}\} \subset \{\text{Weinstein}\} \subset \{\text{Liouville}\} \subset \{\text{strong}\} \subset \{\text{weak}\}.$$

The first inclusion is an equality by a deep result of Eliashberg [27]. All others are strict inclusions, something that has been in known in dimension 3 for some time [19,35,52], but has been fully settled in higher dimensions only very recently [20,21,93,128].

A very broad class for which very strong uniqueness results hold is the following. We say that a contact 3-manifold (X, ξ) is *planar* if ξ is supported (in the sense of Giroux) by an open book whose page has genus zero.

Theorem J. (Wendl [126]) *Assume that (M, ω) is a strong symplectic filling of a planar contact 3-manifold (X, ξ), and fix a supporting open book of* *genus zero pages, i.e., $M = \mathbf{OB}(P, \phi)$ with $g(P) = 0$. Then, (M, ω) is symplectomorphic to a (symplectic) blow-up of the symplectic Lefschetz fibration $\mathbf{LF}(P, \phi)$.*

If we assume that the strong filling is *minimal*, in the sense that it does not have symplectic spheres of self-intersection -1 (i.e. exceptional divisors), such a filling is then uniquely determined. It follows as a corollary that a planar contact manifold is strongly fillable if and only if every supporting planar open book has monodromy isotopic to a product of positive Dehn twists. This reduces the study of strong fillings of a planar contact 3-manifolds to the study of factorizations of a given monodromy into product of positive Dehn twists, a problem of geometric group theory in the mapping class group of a genus zero surface.

References A good introductory textbook to contact topology is Geiges' book [49]; see also [50] by the same author for a very nice survey on the history of contact geometry and topology, including connections to the work of Sophus Lie on differential equations (which gave rise to the contact condition), Huygens' principle on optics, and the formulation of classical thermodynamics in terms of contact geometry. For an introduction to symplectic topology, McDuff–Salamon [95] is a must-read. Anna Cannas da Silva [23] is also a very good source, touching on Kähler geometry as well as toric geometry, relevant for the classical theory of integrable systems. For open books and Giroux's correspondence in dimension 3, Etnyre's notes [36] is a good place to learn. For open books in complex singularity theory (i.e., Milnor fibrations), the classical book by Milnor [97] is a gem. For related reading on Brieskorn manifolds in contact topology, Lefschetz fibrations, and further material, Kwon–van Koert [86] is a great survey. Another good source for symplectic geometry in dimension 4, Lefschetz pencils, and its relationship to holomorphic curves and rational/ruled surfaces is Wendl's recent book [127].

3. The three-body problem

After paving the way, we now discuss a very old conundrum. The setup of the classical 3-body problem consists of three bodies in $\mathbb{R}^3$, subject to the gravitational interactions between them, which are governed by Newton's laws of motion. Given initial positions and velocities, the problem consists in predicting the future positions and velocities of the bodies. The understanding of the resulting dynamical system is quite a challenge, and an outstanding open problem.

We consider three bodies: earth (E), moon (M), and satellite (S), with masses m_E, m_M, m_S. We have the following special cases:

- (restricted) $m_S = 0$ (the satellite is negligible wrt the *primaries* E and M);
- (circular) Each primary moves in a circle, centered around the common center of mass of the two (as opposed to general ellipses);
- (planar) S moves in the plane containing the primaries;

- (spatial) The planar assumption is dropped, and S is allowed to move in three-space.

The restricted problem then consists in understanding the dynamics of the trajectories of the Satellite, whose motion is affected by the primaries, but not vice versa. For simplicity, we will use the acronym CR3BP=circular restricted three-body problem. We denote the *mass ratio* by $\mu = \frac{m_M}{m_E+m_M} \in [0,1]$, and we normalize, so that $m_E + m_M = 1$, and so, $\mu = m_M$.

In a suitable inertial plane spanned by the E and M, the position of the Earth becomes $E(t) = (\mu\cos(t), \mu\sin(t))$, and the position of the Moon is $M(t) = (-(1-\mu)\cos(t), -(1+\mu)\sin(t))$. The time-dependent Hamiltonian whose Hamiltonian dynamics we wish to study is then

$$\begin{aligned} &H_t : \mathbb{R}^3\backslash\{E(t), M(t)\} \to \mathbb{R} \\ &H_t(q,p) = \frac{1}{2}\|p\|^2 - \frac{\mu}{\|q-M(t)\|} - \frac{1-\mu}{\|q-E(t)\|}, \end{aligned}$$

i.e., the sum of the kinetic energy plus the two Coulomb potentials associated to each primary. Note that this Hamiltonian is time-dependent. To remedy this, we choose rotating coordinates, in which both primaries are at rest; the price to pay is the appearance of angular momentum term in the Hamiltonian which represents the centrifugal and Coriolis forces in the rotating frame. Namely, we undo the rotation of the frame, and assume that the positions of Earth and Moon are $E = (\mu, 0, 0)$, $M = (-1+\mu, 0, 0)$. After this (time-dependent) change of coordinates, which is just the Hamiltonian flow of $L = p_1q_2 - p_2q_1$, the Hamiltonian becomes

$$\begin{aligned} &H : \mathbb{R}^3\backslash\{E, M\} \times \mathbb{R}^3 \to \mathbb{R} \\ &H(q,p) = \frac{1}{2}\|p\|^2 - \frac{\mu}{\|q-M\|} - \frac{1-\mu}{\|q-E\|} + p_1q_2 - p_2q_1, \end{aligned}$$

and in particular is *autonomous*. By preservation of energy, this means that it is a preserved quantity of the Hamiltonian motion. The planar problem is the subset $\{p_3 = q_3 = 0\}$, which is clearly invariant under the Hamiltonian dynamics.

There are precisely five critical points of H, called the *Lagrangian points* $L_i, i = 1, \ldots, 5$, ordered, so that $H(L_1) < H(L_2) < H(L_3) < H(L_4) = H(L_5)$ (in the case $\mu < 1/2$; if $\mu = 1/2$, we further have $H(L_2) = H(L_3)$). L_1, L_2, L_3, all saddle points, lie in the axis between Earth and Moon (they are the *collinear* Lagrangian points). L_1 lies between the latter, while L_2 on the opposite side of the Moon, and L_3 on the opposite side of the Earth. The others, L_4, L_5, are maxima, and are called the *triangular* Lagrangian points. For $c \in \mathbb{R}$, consider the energy hypersurface $\Sigma_c = H^{-1}(c)$. If

$$\pi : \mathbb{R}^3\backslash\{E, M\} \times \mathbb{R}^3 \to \mathbb{R}^3\backslash\{E, M\},\ \pi(q,p) = q,$$

is the projection onto the position coordinate, we define the *Hill's region* of energy c as

$$\mathcal{K}_c = \pi(\Sigma_c) \in \mathbb{R}^3\backslash\{E, M\}.$$

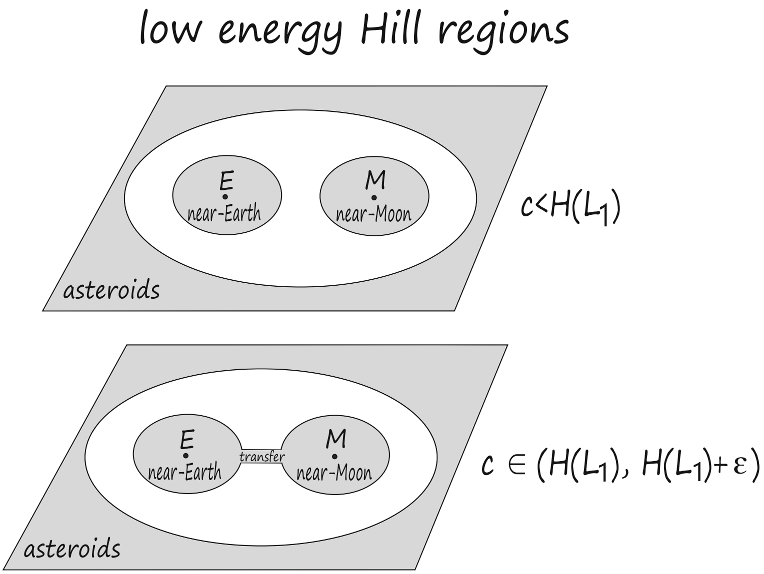


FIGURE 10. The low-energy Hill regions

This is the region in space where the satellite of energy c is allowed to move. If $c < H(L_1)$ lies below the first critical energy value, then $\mathcal{K}_c$ has three connected components: a bounded one around the Earth, another bounded one around the Moon, and an unbounded one. Namely, if the Satellite starts near one of the primaries, and has low energy, then it stays near the primary also in the future. The unbounded region corresponds to asteroids which stay away from the primaries. Denote the first two components by $\mathcal{K}_c^E$ and $\mathcal{K}_c^M$, as well as $\Sigma_c^E = \pi^{-1}(\mathcal{K}_c^E) \cap \Sigma_c$, $\Sigma_c^M = \pi^{-1}(\mathcal{K}_c^M) \cap \Sigma_c$, the components of the corresponding energy hypersurface over the bounded components of the Hill region. As c crosses the first critical energy value, the two connected components $\mathcal{K}_c^E$ and $\mathcal{K}_c^M$ get glued to each other into a new connected component $\mathcal{K}_c^{E,M}$, which topologically is their connected sum. Then, the Satellite in principle has enough energy to transfer between Earth and Moon. In terms of Morse theory, crossing critical values corresponds precisely to attaching handles, so similar handle attachments occur as we sweep through the energy values until the Hill region becomes all of position space. See Fig. 10.

4. Moser regularization

The 5-dimensional energy hypersurfaces are non-compact, due to collisions of the massless body S with one of the primaries, i.e., when if $q = M$ or $q = E$. Note that the Hamiltonian becomes singular at collisions because of the Coulomb potentials, and conservation of energy implies that the momenta necessarily explodes whenever S collides (i.e., $p = \infty$). Fortunately, there are ways to regularize the dynamics even after collision. Intuitively, the effect is:

whenever S collides with a primary, it bounces back to where it came from, and hence, we continue the dynamics beyond the catastrophe. More formally, one is looking for a compactification of the energy hypersurface, which may be viewed as the level set of a new Hamiltonian on another symplectic manifold, in such a way that the Hamiltonian dynamics of the compact, regularized level set is a *reparametrization* of the original one (time is forgotten under regularization).

Two body collisions can be regularized via Moser's recipe. This consists in interchanging position and momenta, and compactifying by adding a point at infinity corresponding to collisions (where the velocity explodes). The bounded components Σ_c^E and Σ_c^M [for $c < H(L_1)$), as well as $\Sigma_c^{E,M}$ (for $c \in (H(L_1), H(L_1)+\epsilon)$], are thus compactified to compact manifolds $\overline{\Sigma}_c^E$, $\overline{\Sigma}_c^M$, and $\overline{\Sigma}_c^{E,M}$. The first two are diffeomorphic to $S^*S^3 = S^3 \times S^2$, and should be thought of as level sets in (two different copies of) $(T^*S^3, \omega_{\text{std}})$ of a suitable regularized Hamiltonian $Q : T^*S^3 \to \mathbb{R}$. The fiber of the level sets $\overline{\Sigma}_c^E$, $\overline{\Sigma}_c^M$ over (a momenta) $p \in S^3$ is a 2-sphere allowed positions q to have fixed energy. If $p = \infty$ is the North pole, the fiber, called the *collision locus*, is the result of a real blow-up at a primary, i.e., we add all possible "infinitesimal" positions nearby (which one may think of as all unit directions in the tangent space of the primary) (Fig. 11). On the other hand, $\overline{\Sigma}_c^E$ is a copy of $S^*S^3 \# S^*S^3$, which can be understood in terms of handle attachments along a critical point of index 1. In the planar problem, the situation is similar: we obtain copies of $S^*S^2 = \mathbb{R}P^3$ and $\mathbb{R}P^3 \# \mathbb{R}P^3$.

In terms of formulas, this can be done as follows.

4.1. Stark–Zeeman systems

We will only do the subcritical case $c < H(L_1)$. By restricting the Hamiltonian to the Earth or Moon component, we can view the three-body problem as a *Stark–Zeeman* system, which is a more general class of mechanical systems.

To define such systems in general, consider a twisted symplectic form

$$\omega = \mathrm{d}\vec{p} \wedge \mathrm{d}\vec{q} + \pi^* \sigma_B,$$

with $\sigma_B = \frac{1}{2} \sum B_{ij} \mathrm{d}q_i \wedge \mathrm{d}q_j$ a 2-form on the position variables (a *magnetic* term, which physically represents the presence of an electromagnetic field, as in Maxwell's equations), and $\pi(q,p) = q$ the projection to the base. A *Stark–Zeeman system* for such a symplectic form is a Hamiltonian of the form

$$H(\vec{q}, \vec{p}) = \frac{1}{2}\|\vec{p}\|^2 + V_0(\vec{q}) + V_1(\vec{q}),$$

where $V_0(\vec{q}) = -\frac{g}{\|\vec{q}\|}$ for some positive coupling constant g, and V_1 is an extra potential.[1]

We will make two further assumptions.

[1] In this section, we will use the symbol $\vec{\ }$ for vectors in $\mathbb{R}^3$ to make our formulas for Moser regularization simpler. We will use the convention that $\xi \in \mathbb{R}^4$ has the form $(\xi_0, \vec{\xi})$.

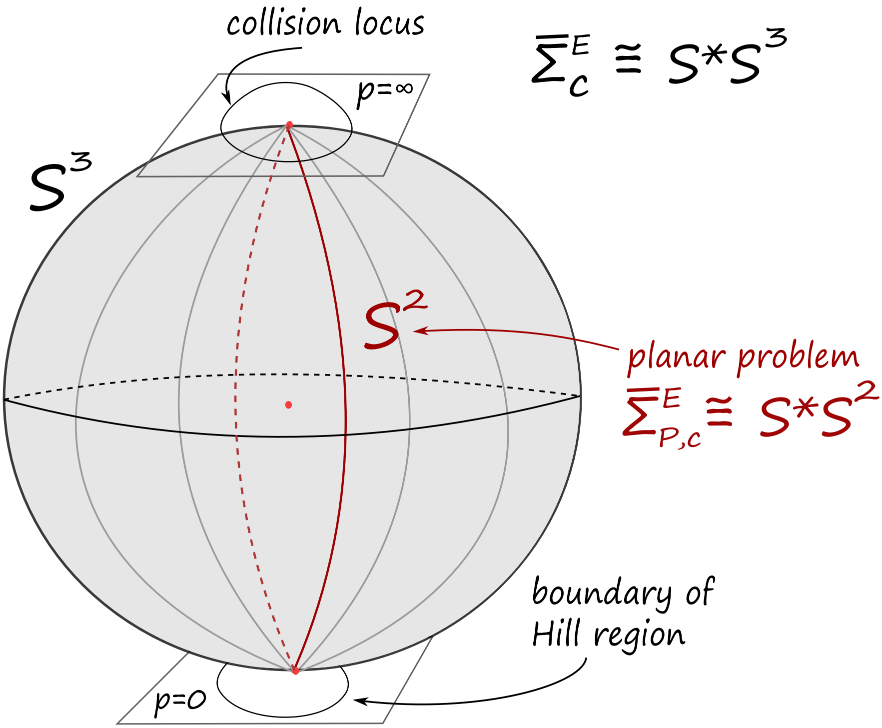


FIGURE 11. In Moser regularization near the Earth, we add a Legendrian sphere of collisions at the North pole (for fixed energy). The planar problem, which also contains collisions, is an invariant subset

Assumption. (A1) We assume that the magnetic field is exact with primitive 1-form $\vec{A}$. Then, with respect to $d\vec{p} \wedge d\vec{q}$, we can write

$$H(\vec{q}, \vec{p}) = \frac{1}{2}\|\vec{p} + \vec{A}(\vec{q})\|^2 + V_0(\vec{q}) + V_1(\vec{q}).$$

(A2) We assume that $\vec{A}(\vec{q}) = (A_1(q_1, q_2), A_2(q_1, q_2), 0)$, and that the potential satisfies that symmetry $V_1(q_1, q_2, -q_3) = V_1(q_1, q_2, q_3)$.

Observe that these assumptions imply that the planar problem, defined as the subset $\{(\vec{q}, \vec{p}) : q_3 = p_3 = 0\}$, is an invariant set of the Hamiltonian flow. Indeed, we have

$$\dot{q}_3 = \frac{\partial H}{\partial p_3} = p_3, \text{ and } \dot{p}_3 = -\frac{\partial H}{\partial q_3} = -\frac{g q_3}{\|\vec{q}\|^3} - \frac{\partial V_1}{\partial q_3}. \tag{4.1}$$

Both these terms vanish on the subset $q_3 = p_3 = 0$ by noting that the symmetry implies that $\frac{\partial V_1}{\partial q_3}|_{q_3=0} = 0$.

For non-vanishing g, Stark–Zeeman systems have a singularity corresponding to two-body collisions, which we will regularize by Moser regularization. To do so, we will define a new Hamiltonian Q on T^*S^3 whose dynamics correspond to a reparametrization of the dynamics of H. We will describe the scheme for energy levels $H = c$, which we need to *fix* a priori (i.e., the regularization is not in principle for all level sets at once). Define the intermediate Hamiltonian

$$K(\vec{q}, \vec{p}) := (H(\vec{q}, \vec{p}) - c)\|\vec{q}\|.$$

For $\vec{q} \neq 0$, this function is smooth, and its Hamiltonian vector field equals

$$X_K = \|\vec{q}\| \cdot X_H + (H - c)X_{\|\vec{q}\|}.$$

We observe that X_K is a multiple of X_H on the level set $K = 0$. Writing out K gives

$$K = \left(\frac{1}{2}(\|\vec{p}\|^2 + 1) - (c + 1/2) + \langle \vec{p}, \vec{A}\rangle + \frac{1}{2}\|\vec{A}\|^2 + V_1(\vec{q})\right) \|\vec{q}\| - g.$$

Stereographic projection We now substitute with the stereographic coordinates. The basic idea is to switch the role of momentum and position in the $\vec{q}, \vec{p}$-coordinates, and use the $\vec{p}$-coordinates as position coordinates in $T^*\mathbb{R}^n$ (for any n), where we think of $\mathbb{R}^n$ as a chart for S^n. We set

$$\vec{x} = -\vec{p}, \quad \vec{y} = \vec{q}.$$

We view T^*S^n as a symplectic submanifold of $T^*\mathbb{R}^{n+1}$, via

$$T^*S^n = \{(\xi, \eta) \in T^*\mathbb{R}^{n+1} \mid \|\xi\|^2 = 1,\ \langle \xi, \eta\rangle = 0\}.$$

Let $N = (1, 0, \dots, 0) \in S^n$ be the north pole. To go from $T^*S^n \backslash T_N^*S^n$ to $T^*\mathbb{R}^n$, we use the stereographic projection, given by

$$\begin{aligned} \vec{x} &= \frac{\vec{\xi}}{1 - \xi_0} \\ \vec{y} &= \eta_0 \vec{\xi} + (1 - \xi_0)\vec{\eta}. \end{aligned} \tag{4.2}$$

To go from $T^*\mathbb{R}^n$ to $T^*S^n \backslash T_N^*S^n$, we use the inverse given by

$$\begin{aligned} \xi_0 &= \frac{\|\vec{x}\|^2 - 1}{\|\vec{x}\|^2 + 1} \\ \vec{\xi} &= \frac{2\vec{x}}{\|\vec{x}\|^2 + 1} \\ \eta_0 &= \langle \vec{x}, \vec{y}\rangle \\ \vec{\eta} &= \frac{\|\vec{x}\|^2 + 1}{2}\vec{y} - \langle \vec{x}, \vec{y}\rangle \vec{x}. \end{aligned} \tag{4.3}$$

These formulas imply the following identities:

$$\frac{2}{\|\vec{x}\|^2 + 1} = 1 - \xi_0, \quad \|\vec{y}\| = \frac{2\|\eta\|}{\|\vec{x}\|^2 + 1} = (1 - \xi_0)\|\eta\|,$$

which allows us to simplify the expression for K. Setting $n = 3$, we obtain a Hamiltonian $\tilde{K}$ defined on T^*S^3, given by

$$\begin{aligned}\tilde{K} &= \left(\frac{1}{1-\xi_0} - (c+1/2) - \frac{1}{1-\xi_0}\langle \vec{\xi}, \vec{A}(\xi,\eta)\rangle + \frac{1}{2}\|\vec{A}(\xi,\eta)\|^2 + V_1(\xi,\eta)\right)\\ &\quad (1-\xi_0)\|\eta\| - g\\ &= \|\eta\| \Big(1 - (1-\xi_0)(c+1/2) - \langle \vec{\xi}, \vec{A}(\xi,\eta)\rangle + (1-\xi_0)\\ &\quad \left(\frac{1}{2}\|\vec{A}(\xi,\eta)\|^2 + V_1(\xi,\eta)\right)\Big) - g.\end{aligned}$$

Put

$$\begin{aligned}f(\xi,\eta) &= 1 + (1-\xi_0)\left(-(c+1/2) + \frac{1}{2}\|\vec{A}(\xi,\eta)\|^2 + V_1(\xi,\eta)\right) - \langle \vec{\xi}, \vec{A}(\xi,\eta)\rangle\\ &= 1 + (1-\xi_0)b(\xi,\eta) + M(\xi,\eta), \end{aligned} \tag{4.4}$$

where

$$\begin{aligned}b(\xi,\eta) &= -(c+1/2) + \frac{1}{2}\|\vec{A}(\xi,\eta)\|^2 + V_1(\xi,\eta)\\ M(\xi,\eta) &= -\langle \vec{\xi}, \vec{A}(\xi,\eta)\rangle.\end{aligned}$$

Note that the collision locus corresponds to $\xi_0 = 1$, i.e., the cotangent fiber over N. The notation is supposed to suggest that $(1-\xi_0)b(\xi,\eta)$ vanishes on the collision locus and M is associated with the magnetic term; it is not the full magnetic term, though. We then have that

$$\tilde{K} = \|\eta\| f(\xi,\eta) - g.$$

To obtain a smooth Hamiltonian, we define the Hamiltonian

$$Q(\xi,\eta) := \frac{1}{2} f(\xi,\eta)^2 \|\eta\|^2.$$

The dynamics on the level set $Q = \frac{1}{2}g^2$ are a reparametrization of the dynamics of $\tilde{K} = 0$, which in turn correspond to the dynamics of $H = c$.

Remark 4.1. We have chosen this form to stress that Q is a deformation of the Hamiltonian describing the geodesic flow on the round sphere, which is given by level sets of the Hamiltonian

$$Q_{\text{round}} = \frac{1}{2}\|\eta\|^2.$$

This is the dynamics that one obtains in the regularized Kepler problem (the two-body problem; see below), corresponding to the Reeb dynamics of the contact form given by the standard Liouville form. As we have seen, this is a Giroux form for the open book $S^*S^3 = \mathbf{OB}(\mathbb{D}^*S^2, \tau^2)$, supporting the standard contact structure on S^*S^3.

Formula for the restricted three-body problem Since the restricted three-body problem is our main interest, we conclude this section by giving the explicit formula for this problem. By completing the squares, we obtain

$$H(\vec{q},\vec{p}) = \frac{1}{2}\left((p_1+q_2)^2 + (p_2-q_1)^2 + p_3^2\right) - \frac{\mu}{\|\vec{q}-\vec{m}\|} - \frac{1-\mu}{\|\vec{q}-\vec{e}\|} - \frac{1}{2}(q_1^2+q_2^2).$$

This is then a Stark–Zeeman system with primitive

$$\vec{A} = (q_2, -q_1, 0),$$

coupling constant $g = \mu$, and potential

$$V_1(\vec{q}) = -\frac{1-\mu}{\|\vec{q}-\vec{e}\|} - \frac{1}{2}(q_1^2 + q_2^2), \tag{4.5}$$

both of which satisfy Assumptions (A1) and (A2).

After a computation, we obtain

$$\begin{aligned} f(\xi,\eta) = 1 &+ (1-\xi_0)\left(-(c+1/2) + \xi_2\eta_1 - \xi_1\eta_2\right) - \xi_2\,(1-\mu) \\ &- \frac{(1-\mu)(1-\xi_0)}{\|\vec{\eta}(1-\xi_0) + \vec{\xi}\eta_0 + \vec{m} - \vec{e}\|}, \end{aligned} \tag{4.6}$$

and we have

$$b(\xi,\eta) = -(c+1/2) - \frac{(1-\mu)}{\|\vec{\eta}(1-\xi_0) + \vec{\xi}\eta_0 + \vec{m} - \vec{e}\|} \tag{4.7}$$

$$M(\xi,\eta) = (1-\xi_0)(\xi_2\eta_1 - \xi_1\eta_2) - \xi_2(1-\mu). \tag{4.8}$$

4.2. Levi–Civita regularization

We follow the exposition in [47]. Consider the map

$$\begin{aligned} &\mathcal{L} : \mathbb{C}^2\backslash(\mathbb{C}\times\{0\}) \to T^*\mathbb{C}\backslash\mathbb{C}, \\ &(u,v) \mapsto \left(\frac{u}{\bar{v}}, 2v^2\right), \end{aligned}$$

where we view $\mathbb{C} \subset T^*\mathbb{C}$ as the zero section. Using $\mathbb{C}$ as a chart for S^2 via the stereographic projection along the north pole, this map extends to a map

$$\mathcal{L} : \mathbb{C}^2\backslash\{0\} \to T^*S^2\backslash S^2,$$

which is a degree 2 cover. Writing (p,q) for coordinates on $T^*\mathbb{C} = \mathbb{C}\times\mathbb{C}$ (this is the *opposite* to the standard convention, and comes from the Moser regularization), the Liouville form on $T^*\mathbb{C}$ is $\lambda = q_1 \mathrm{d}p_1 + q_2 \mathrm{d}p_2$, with associated Liouville vector field $X = q_1\partial_{q_1} + q_2\partial_{q_2}$. One checks that

$$\mathcal{L}^*\lambda = 2(v_1\mathrm{d}u_1 - u_1\mathrm{d}v_1 + v_2\mathrm{d}u_2 - u_2\mathrm{d}v_2),$$

whose derivative is the symplectic form

$$\omega = \mathrm{d}\lambda = 4(\mathrm{d}v_1 \wedge \mathrm{d}u_1 + \mathrm{d}v_2 \wedge \mathrm{d}u_2).$$

Note that λ and ω are *different* from the standard Liouville and symplectic forms (resp.) on $\mathbb{C}^2$. However, the associated Liouville vector field defined via $i_V\omega = \lambda$ coincides with the standard Liouville vector field

$$V = \frac{1}{2}(u_1\partial_{u_1} + u_2\partial_{u_2} + v_1\partial_{v_1} + v_2\partial_{v_2}),$$

and we have $\mathcal{L}^*X = V$. We conclude the following:

Lemma 4.2. *A closed hypersurface $\Sigma \subset T^*S^2$ is fiber-wise star-shaped if and only if $\mathcal{L}^{-1}(\Sigma) \subset \mathbb{C}^2\backslash\{0\}$ is star-shaped.*

Note that $\Sigma \cong S^*S^2 \cong \mathbb{R}P^3$, and $\mathcal{L}^{-1}(\Sigma) \cong S^3$, and so, $\mathcal{L}$ induces a two-fold cover between these two hypersurfaces.

4.3. Kepler problem

We now work out the Moser and Levi–Civita regularizations of the *Kepler problem* at energy $-\frac{1}{2}$. This is the well-known two-body problem, whose Hamiltonian is given by

$$E : T^*(\mathbb{R}^2\backslash\{0\}) \to \mathbb{R},$$
$$E(q,p) = \frac{1}{2}\|p\|^2 - \frac{1}{\|q\|}.$$

The result of Moser regularization is the Hamiltonian

$$K(p,q) = \frac{1}{2}\left(\|q\|\left(E(-q,p)+\frac{1}{2}\right)+1\right)^2 = \frac{1}{2}\left(\frac{1}{2}\left(\|p\|^2+1\right)\|q\|\right)^2.$$

This is the kinetic energy of the “momentum” q, with respect to the round metric, viewed in the stereographic projection chart. It follows that its Hamiltonian flow is the round geodesic flow. Moreover, we have

$$X_K|_{E^{-1}(-1/2)}(p,q) = \|q\|X_E|_{E^{-1}(-1/2)}(-q,p),$$

so that the Kepler flow is a reparametrization of the round geodesic flow.

To understand the Levi–Civita regularization, we consider the shifted Hamiltonian $H = E + \frac{1}{2}$ (which has the same Hamiltonian dynamics). After substituting variables via the Levi–Civita map $\mathcal{L}$, we obtain

$$H(u,v) = \frac{\|u\|^2}{2\|v\|^2} - \frac{1}{2\|v\|^2} + \frac{1}{2}.$$

We then consider the Hamiltonian

$$Q(u,v) = \|v\|^2 H(u,v) = \frac{1}{2}(\|u\|^2 + \|v\|^2 - 1).$$

The level set $Q^{-1}(0) = H^{-1}(0)$ is the 3-sphere, and the Hamiltonian flow of Q, a reparametrization of that of H, is the flow of two uncoupled harmonic oscillators. This is precisely the Hopf flow. We summarize this discussion in the following:

Proposition 4.3. *The Moser regularization of the Kepler problem is the geodesic flow on S^2. Its Levi–Civita regularization is the Hopf flow on S^3, i.e., the double cover of the geodesic flow on S^2 (cf. Remark 2.19).*

5. Historical remarks

This section contains a historical account, from the Poincaré approach to finding closed orbits in the three-body problem, to some current developments in symplectic geometry. This is by all means non-exhaustive, and tilted towards the author’s interests and biased understanding of the developments.

The perturbative philosophy One of the most basic approaches that underlies mathematics and physics is the perturbative approach. Basically, it means understanding a simplified situation first, where everything can be explicitly understood, and attempt to understand ”nearby” situations by perturbing the parameters relevant to the problem in question.

In the context of celestial/classical mechanics, this was precisely the approach of Poincaré. The idea is to start with a limit case, which is *completely integrable* (i.e., an integrable system), perturb it, and study what remained. Integrable systems, roughly speaking, are those which allow enough symmetries, so that the solutions to the equations of motion can be "explicitly" solved for (however, quantitative questions need to allow sufficiently many functions, e.g., special functions such as elliptic integrals). The solutions tend to admit descriptions in terms of algebraic geometry. In the classical setting of celestial mechanics, if phase space is $2n$-dimensional and the Hamiltonian H Poisson-commutes with other $n-1$ Hamiltonians (which are therefore preserved under the Hamiltonian flow of H), the well-known Arnold–Liouville theorem provides action-angle coordinates in which the symplectic manifold is foliated by flow-invariant tori, along which the Hamiltonian flow is linear, with varying slopes (the *frequencies*). In good situations, the generic tori are half-dimensional (and *Lagrangian*, i.e., the symplectic form vanishes along them), whereas there might also be degenerate lower dimensional tori. This is the natural realm of toric symplectic geometry, dealing with symplectic manifolds which admit a Hamiltonian action of the torus, and the study of the corresponding moment maps and their associated Delzant polytopes. There is also a related theory in algebraic geometry, where the polytope is replaced with a fan. However, in general (e.g., the Euler problem), we get only an $\mathbb{R}^n$-action, which is unfortunately beyond the scope of toric geometry. See [67] for more connections between the theory of integrable systems, and differential and algebraic geometry.

The study of what remains after a small perturbation of an integrable system is the realm of KAM theory, as well as complementary weaker versions such as Aubry–Mather theory. Roughly speaking, the original version of the KAM theorem (due to Kolmogorov–Arnold–Moser) says that if one perturbs a "sufficiently irrational" Liouville torus, i.e., the vector of frequencies of the action is very badly approximated by rational numbers (it is *diophantine*), and moreover, the Hessian with respect to action variables is non-degenerate, then the Liouville tori survives to an invariant tori whose frequencies are close to the original one, and hence is foliated by orbits which are *quasi-periodic*, in the sense that they are dense in the tori and never close up. Aubry–Mather theory is meant to deal with the rest of the tori, including resonant ones which are foliated by closed orbits and non-diophantine non-resonant ones, as well as large deformations (as opposed to sufficiently small perturbations). This theory provides invariant subsets which are usually Cantor-like, and obtained via measure-theoretical means (they are the supports of invariant measures minimizing certain action functionals).

The Poincaré–Birkhoff theorem, and the planar three-body problem The problem of finding closed orbits in the planar case of the restricted three-body problem goes back to ground-breaking work in celestial mechanics of Poincaré [109, 110], building on work of G.W. Hill on the lunar problem [62, 63]. The basic scheme for his approach may be reduced to:

(1) Finding a global surface of section for the dynamics;

(2) Proving a fixed-point theorem for the resulting first return map.

This is the setting for the celebrated Poincaré–Birkhoff theorem, proposed and confirmed in special cases by Poincaré and later proved in full generality by Birkhoff in [16]. The statement can be summarized as: if $f : A \to A$ is an area-preserving homeomorphism of the annulus $A = [-1, 1] \times S^1$ that satisfies a *twist* condition at the boundary (i.e., it rotates the two boundary components in opposite directions), then it admits infinitely many periodic points of arbitrary large period. The fact that the area is preserved is a consequence of Liouville's theorem for Hamiltonian systems; we have basically used this in our proof of Proposition 2.16.

The whole point of a global surface of section is to reduce a *continuous* flow on a 3-manifold to the *discrete* dynamics of a map on a 2-manifold, thus reducing by one the degrees of freedom. It is perhaps fair to say, that this key (and beautiful) idea is responsible for motivating the well-studied area of dynamics on surfaces, a huge industry in its own right.

The direct and retrograde orbits The actual physical Moon is in *direct* motion around the Earth (i.e., it rotates in the same direction around the Earth as the Earth around the Sun). The opposite situation is a *retrograde* motion. In [62,63], while attempting to model the motion of the Moon, Hill indeed finds both direct and retrograde orbits. While still an idealized situation, such direct orbit is a reasonable approximation to the actual orbit of the Moon, and Hill even goes further to find better approximations via perturbation theory, something which deeply impressed Poincaré himself. Let us remark that direct orbits are usually the more interesting to astronomers, since most moons are in direct motion around their planet. Topologically, one may think of the retrograde/direct Hill orbits as obtained from a Hopf link in S^3, via the double cover to $\mathbb{R}P^3$. This is the binding of the open book $\mathbb{R}P^3 = \mathbf{OB}(\mathbb{D}^*S^1, \tau^2)$, where τ is the positive Dehn twist along $S^1 \subset \mathbb{D}^*S^1$.

Brouwer's and Frank's theorem To find the direct orbit away from the lunar problem, Birkhoff had in mind finding a disk-like surface of section whose boundary is precisely the retrograde orbit. The direct orbit would then be found via Brouwer's translation theorem: every area-preserving homeomorphism of the open disk admits a fixed point. Removing the fixed point, we obtain an area-preserving homeomorphism of the open annulus, which, via a theorem of Franks, admits either none or infinitely many periodic points. All this combined, one has: an area-preserving homeomorphism of an open disk admits either one or infinitely many periodic points. Note that if the boundary is also an orbit, we obtain 2 or infinitely many. If furthermore we have twist, the Poincaré–Birkohff theorem provides infinitely many orbits. This is a classical heuristic for finding orbits that has survived to this day in several guises, as we will see below. See Fig. 12.

Perturbative results As we have seen, we have $\mathbb{R}P^3 = \mathbf{OB}(\mathbb{D}^*S^1, \tau^2)$ as smooth manifolds, and one would hope that a concrete version of this open book is adapted to the (Moser-regularized) planar dynamics, and that the return map is a Birkhoff twist map. For $c < H(L_1)$ and $\mu \sim 0$ small, one can interpret from this perspective that Poincaré [110] proved this by perturbing

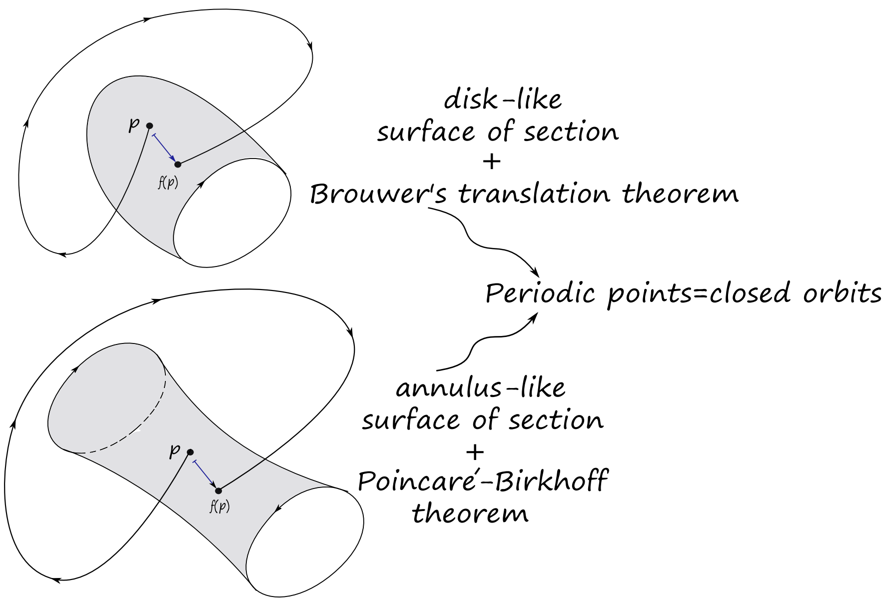


FIGURE 12. Obtaining closed orbits in the planar problem

the rotating Kepler problem (when $\mu = 0$), which is an integrable system for which the return map is a twist map. Of course, he never stated it in these words. In the case where $c \ll H(L_1)$ is very negative and $\mu \in (0, 1)$ is arbitrary, this was done by Conley [29] (also perturbatively), who checked the twist condition and used Poincaré–Birkhoff. In [96], McGehee provides a disk-like global surface of section for the rotating Kepler problem problem for $c < H(L_1)$, and computes the return map.

Non-perturbative results More generally and *non-perturbatively*, the existence of this adapted open book was obtained in [77, Theorem 1.18] for the case where (μ, c) lies in the *convexity range* via holomorphic curve methods due to Hofer–Wysocki–Zehnder [73] (see also [5,6]). This non-perturbative approach, which implies the use of modern techniques of symplectic and contact geometry, will be discussed below.

The search of closed geodesics: a very brief survey After suitable regularization, the round geodesic flow on S^2 appears as an integrable limit case in the planar restricted three-body problem, when the Jacobi constant c converges to $-\infty$. Poincaré was aware of this fact, which brought him, near the end of his life, to study the geodesic flow of "near-integrable" metrics on S^2, i.e., perturbations of the round one. One may well argue that this was one of the starting points of the very long and fruitful search of closed geodesics that ensued later throughout the 20th century.

A basic argument for finding closed geodesics, sometimes attributed to Birkhoff, was already present in work of Hadamard in 1898, who studied the case of surfaces with negative curvature. This is a variational argument on the loop space, in the sense that closed geodesics are viewed as loops which happen to be geodesics (as opposed to the dynamical point of view, where a closed geodesic is a geodesic path which happens to close up). It works as follows: on a compact manifold, one chooses a sequence of loops in a fixed homotopy class whose length converges to the infinimum in such class, and appeals to the Arzelà–Ascoli theorem. If the infimum is non-zero, this gives a non-trivial closed geodesic. This argument works if the fundamental group is non-trivial; it gives a geodesic in each non-trivial free homotopy class, and hence infinitely many if the genus is at least 1. This leaves out the case of S^2, for which it gives nothing. The program of finding geodesics for general manifolds was picked up by Birkhoff in a more systematic way, who proved existence of at least one geodesic for the case of all surfaces and certain higher dimensional manifolds including spheres. For the case where the infimum in the above variational argument is zero, Birkhoff introduced the famous minmax argument. For S^2, this works as follows: take the foliation of S^2 minus the north and south poles, whose leaves are the circles given by the parallels (think of the standard embedding, but where the metric is not the standard one). Choose a curve-shortening procedure for each non-trivial leaf (there are several, the simplest one being replacing two nearby points on a loop by a geodesic arc; this is a tricky business, however, since the resulting loop might have self-intersections). This gives a sequence of foliations, and we may choose the loop with maximal length for each. These lengths are bounded from below for topological reasons. Again by Arzelà–Ascoli, the limit of such curves, being invariant under the shortening procedure, is a geodesic.

Before Birkhoff, Poincaré himself [111] had the idea of obtaining a geodesic for the case of S^2 embedded in $\mathbb{R}^3$ as a convex surface S (with the induced metric), by considering the shortest simple closed curve γ dividing S into two pieces of equal total Gaussian curvature. A simple argument using Gauss–Bonnet shows that γ should be a geodesic. The full details of this beautiful argument were carried out by Croke in 1982 [32], who considered the more general case of a convex hypersurface in $\mathbb{R}^n$.

Poincaré further proposed that, also in the case of a convex S^2 in $\mathbb{R}^3$, there should be at least 3 closed geodesics with no self-intersections (i.e., simple). A short proof of this was published by Lusternik–Schnirelmann in 1929 [91,92]. Their proof relied on two steps: first, to consider the space of all simple circles (great and small) and a continuous curve-shrinking procedure which keeps all such circles simple; and second, the fact that the space of non-oriented round geodesics is a copy of $\mathbb{R}P^2$ (it can be identified with the space of planes in $\mathbb{R}^3$ through the origin), together with the fact that every Morse function on $\mathbb{R}P^2$ has at least 3 critical points. Unfortunately, there were gaps in both steps. These were filled in by Ballmann in 1978 [12], who also considered the case of arbitrary genus; Gage–Hamilton and Grayson also developed the curvature flow (or curve-shortening flow), which may be viewed as the gradient flow of the length functional. It has the property that, if a smooth simple closed curve undergoes the curvature flow, it remains smoothly embedded without self-intersections.

Existence of at least one geodesic for arbitrary closed Riemannian manifolds was finally proved by Lusternik–Fet in 1951–1952 [37,90]. Their approach was based on Morse theory; and indeed, the problem of finding geodesics was the initial motivation for Morse himself. Geodesics are the critical points of the energy functional on the loop space. Moreover, the space $\mathcal{L}M$ of parametrized closed curves on M cannot be retracted into the subspace $\mathcal{L}^0M$ of homotopically trivial closed curves, and Lusternik–Schnirelmann theory applies to give a critical point outside of $\mathcal{L}^0M$.

Even though the loop space of a manifold is infinite dimensional, if the manifold is compact, then the energy functional satisfies the compactness condition of Palais–Smale, which in practice means that it behaves as a Morse function on a finite-dimensional manifold. However, the main difficulty in this approach is that each geodesic can be iterated, and this corresponds to distinct points in the loop space. Distinguishing two geometrically distinct geodesics is a subtle, hard problem.

So far, all the above methods provide only finitely many geodesics, so how about infinitely many? In this direction, another beautiful idea due to Birkhoff, for a Riemannian S^2, is that of an annulus global surface of section; we have of course seen this in the previous sections. One considers a closed geodesic γ (which Birkhoff proved to exist via the minmax argument explained above), dividing the sphere in a upper and a lower hemishpere. One then considers vectors along γ which point towards the upper hemisphere (this is an annulus) as initial values of geodesics, starts shooting orbits along these vectors, and considers the first return map. However, for this annulus to be a global surface of section, one needs that no geodesic gets "trapped" in the upper hemisphere (this will be satisfied for example when the Gaussian curvature is strictly positive). Moreover, one needs to further check the twist condition at the boundary to apply the Poincaré–Birkhoff theorem. Here, note that Birkhoff only stated the existence of at least two fixed points, but a simple argument which Birkhoff seems to have overlooked was provided by Neumann [103], thus obtaining infinitely many periodic points (not related by iterations); this is the version of the Poincaré–Birkhoff theorem we stated above. In the case where we do have a well-defined Birkhoff map, what if the return map does not twist? This is where the theorem due to Franks from 1992 [46] that we mentioned above (which is a statement about the open annulus) comes into play; he obtained infinitely many geodesics on S^2 for this case. In the case where the Birkhoff annulus is not a global section and so there is no return map, an argument of Bangert from 1993 [13] shows that, if geodesics get trapped, they need to do so around a small "waist" (a "short" geodesic), or more formally, geodesics with no conjugate points. Moreover, he shows that the existence of a waist forces the existence of infinitely many geodesics. One key observation is that the Birkhoff return map sends a point on the boundary (lying on a geodesic) to its second conjugate point along this geodesic, and so, some of the ideas were already present in Birkhoff's work. This filled in the general case, finally (after almost 90 years) obtaining the existence of infinitely many geodesics for an arbitrary metric on S^2. We further mention that in 1993, Nancy Hingston, building on work of other people (see [65] and references therein), also provided a full proof of a quantitative estimate on the growth of the number of geodesics with respect to length; if $N(l)$ is the number of geodesics with length at most l then $N(l) \gtrsim l/log(l)$, i.e., the same growth rate as prime numbers.

One should further mention that Katok [82] (see also Ziller's account [130]) has famously constructed examples of non-reversible Finsler metrics on $S^n, \mathbb{C}P^n$ with only finitely many closed geodesics. For instance, the case of S^2 can be described as the round geodesic flow, but on a frame rotating along the z-axis with irrational angle of rotation (and the metric is arbitrarily close to the round one); so that the only closed geodesics are the equator in both directions. This example shows that the general Finsler case is very different from the Riemannian case, and hence, the $\mathbb{Z}_2$-action which allows to reverse geodesics should be used in a significant way to obtain infinitely many geodesics.

Another celebrated result in this story is that of Gromoll–Meyer 1969 [58]: if the sequence of Betti numbers of the free loop space $\mathcal{L}M$ of M is unbounded, then M admits infinitely many geodesics (for *any* metric). Morse had previously, in his 1932 book "Calculus of variations in the large" (although unfortunately with mistakes), computed the homology of $\mathcal{L}M$ in the non-degenerate case. For this, one may use a spectral sequence whose terms in the E^1-page consist of the homology of the base (constant loops) and the homology of each geodesic, endowed with a local coefficient system, and degree shifted by the Morse index. Note that non-degeneracy is in the Morse–Bott sense, since we can always reparametrize loops (which we consider unoriented) via the action of $O(2)$ on S^1, and so we see one circle for each orientation in this homology group. Another ingredient is Bott's famous iteration formula for the index [17], which implies that $\mu(\gamma^m)$ grows linearly with m. When combined with the homology computation via the above Morse–Bott spectral sequence, one sees that if the set of primitive geodesics is finite, then the Betti numbers of $\mathcal{L}M$ are bounded, and hence the result by Gromoll–Meyer follows in the non-degenerate case. The degenerate case, roughly speaking, is obtained by the fact that every degenerate orbit is the limit of a *finite* number of non-degenerate ones, and contributes to the homology in a bounded index window.

This leaves the question of when the sequence of Betti numbers of $\mathcal{L}M$ is unbounded. In [121], Vigué–Poirrier–Sullivan show, via the above result and algebraic calculations, that if M has finite fundamental group, then the Betti numbers of $\mathcal{L}M$ are unbounded if and only if $H_*(M;\mathbb{Q})$ requires at least 2 generators as a ring. Ziller proves this holds for symmetric spaces of rank > 1 [129]. This covers many cases, but it leaves out many important ones e.g. $S^n, \mathbb{R}P^n, \mathbb{C}P^n, \mathbb{H}P^n, CaP^2$.

On the other hand, one can consider the case of a *generic* metric (or "bumpy", i.e., for which all geodesics are non-degenerate). For such a case, on any manifold with finite fundamental group, Gromov has also shown the following quantitative estimate: there exist constants a, b, such that $N(l) \geq \frac{a}{l}\sum_{i=1}^{bl} b_i(\mathcal{L}M)$. Rademacher [112] has shown the existence of infinitely many geodesics for bumpy metrics on manifolds with finite fundamental group. This result builds on work of Klingenberg–Takens [85], Klingenberg [84], who reduced to the case where all orbits are hyperbolic; and Hingston [64], who covered the bumpy case for $S^n, \mathbb{R}P^n, \mathbb{C}P^n, \mathbb{H}P^n, CaP^2$, under the hyperbolic-orbits-only assumption.

One therefore clearly sees that, while a "simpler" problem than finding closed orbits in the three-body problem, finding infinitely many closed geodesics is significantly complicated. This is a problem that has inspired enormous amounts of work, has spanned most of the 20th century, and still is not known in the general case. Indeed, it is still an open question whether any Riemannian metric on a given closed simply connected manifold admits infinitely many closed geodesics. In particular, it is unknown for $S^n, n \geq 3$, for a general metric.

Remarks on Floer theory, and modern symplectic geometry As we have seen, symplectic geometry is the geometry of classical mechanics, dealing with Hamiltonians and their associated evolution equations, and in particular closed Hamiltonian orbits of period 1. In this context, Arnold [10] proposed his famous conjecture on the minimal number of such orbits for a non-degenerate Hamiltonian on a closed symplectic manifold M: there should be at least as many as the sum of the Betti numbers of M. This is naturally related to the classical Morse inequalities. It is notable that Arnold proposed this conjecture *as a version of the Poincaré–Birkhoff theorem* (here, note that the sum of Betti numbers of the annulus is 2).

It was from this conjecture that one of the cornerstones of the modern methods of symplectic geometry was introduced; namely, Floer theory. Together with the introduction of holomorphic curves due to Gromov in 1985 [59], these two developments form the building bricks of the symplectician's toolkit and daily musings.

The approach of Floer to the Arnold conjecture [38–43] is again based on the ideas of Morse theory. Indeed, one can view Hamiltonian orbits as the critical points of a suitable action functional on the loop space, in such a way that flow-lines correspond to cylinders satisfying an elliptic PDE (the Floer equation). One defines a differential which counts these solutions, and the resulting homology theory is actually isomorphic to the Morse homology of the underlying manifold, so that the Arnold conjecture follows. Floer proved it under some technical assumptions, i.e., symplectic asphericity, and the symplectic Calabi–Yau condition; these have been lifted after work of several authors (Ono [108], Hofer–Salamon [70], Liu–Tian [89], Fukaya–Ono [48],...), at least for the case of rational coefficients. The technical details are very difficult (needing the introduction of virtual techniques) and have been subject of heated debate. Lifting the result to integer coefficients is subject of ongoing efforts, most notably due to Abouzaid–Blumberg [2], who, amongst other results, prove it for every finite field.

As we have seen, a special case of closed Hamiltonian orbits is that of Reeb orbits in a contact-type level set. Since every contact manifold is contact-type in some symplectic manifold (i.e., its symplectization), one can view the problem of finding closed Reeb orbits as an odd-dimensional version of the Hamiltonian problem. In this setting, an important statement related to the Arnold conjecture is the *Weinstein conjecture*, which claims the existence of at least one closed Reeb orbit for any contact form on a given contact manifold. Recalling that geodesic flows are particular cases of Reeb flows, this includes the statement that every Riemannian metric admits a closed geodesic (proved by Lusternik–Fet, as mentioned above). In dimension three, it was established by Taubes [119] (based on Seiberg–Witten theory), thus culminating a large body of work by several people extending over more than 2 decades. There are also further striking results in dimension 3, e.g., Irie's results on equidistribution of closed orbits in the generic case [80, 81], or the "2 or infinitely many" dichotomy for torsion contact structures [30]. This dichotomy uses the combination of Brouwer and Frank's theorem as discussed above as the fixed-point theorem, and Hutching's embedded contact homology (ECH) to find the disk-like global surface of section; and so fits in well with the basic two-step approach by Poincaré. Irie's results rely on the relationship between volume and ECH capacities as proved by Cristofaro-Gardiner–Hutchings–Ramos [31]. In higher dimensions, though there are several partial results (e.g., [9, 45, 71, 72, 124]), the Weinstein conjecture is still open.

While the Arnold conjecture is stated for closed symplectic manifolds, a natural class of symplectic manifolds with non-empty boundary is that of Liouville domains. There is an associated Floer theory for such manifolds, which goes under the name of *symplectic homology*. The first version of such theory was due to Floer–Hofer [44], and can be traced to the Ekeland–Hofer capacities and their relation to early versions of S^1-equivariant symplectic homology[2]; see also section 5 in [68] for an even previous and non-equivariant version, called symplectology. There is also a version due to Viterbo [122, 123] (see also [18, 28] for more recent versions), who showed that symplectic homology of a cotangent bundle is the homology of the free loop space of the base, a bridge between the classical story of finding geodesics, and the modern Floer-theoretic approach (see also [1, 114]).

In the Liouville setting, as opposed to the closed setting, the difference is that the associated Floer theory recovers not only the homology of the manifold, but also dynamical data at the contact-type boundary (i.e., closed Reeb orbits). Of course, one of the motivations for such a theory is the Weinstein conjecture, at least for those contact manifolds which bound a Liouville domain (i.e., Liouville fillable ones). Heuristically, if the symplectic homology is infinite dimensional or zero, then there is at least one orbit at the boundary (since the homology of the manifold is finite dimensional and non-zero, although, strictly speaking, here we need consider the case of "finite-type" Liouville domains; see, e.g., [105] for a nice survey, containing these and related ideas).

[2] This was discussed at the opening lectures by Hofer and Floer in Fall 1988 at the symplectic program at the MSRI Berkeley, although unfortunately is written nowhere. Hofer gave a lecture on capacities and the S^1-equivariant symplectic homology at a conference in Durham in 1989, whose proceedings are published in [33], and contains the non-equivariant part of the story. I thank Hofer for these clarifications.

The Arnold conjecture is a statement about *fixed* points (or 1-periodic orbits) of Hamiltonian maps, and predicts a finite number of such. On the other hand, one could want to estimate the number of *periodic* points (recall the same situation for the Poincaré–Birkhoff theorem, whose original version predicted 2 fixed points, although one can also obtain infinitely many periodic points, as was observed after Birkhoff). The analogous statement for Hamiltonian or Reeb flows is the *Conley conjecture*. Roughly speaking, for a "vast" collection of closed symplectic manifolds, every Hamiltonian map has infinitely many simple periodic orbits and, moreover, simple periodic orbits of unbounded minimal period whenever the fixed points are isolated. This was proved by Ginzburg for closed symplectically aspherical symplectic manifolds in [53] (see [54] and references therein, for a survey and history of the problem; and [55] for what the author understands is the current state of the art). One of the key inputs is a special class of critical points introduced by Hingston, and later called symplectic degenerate maxima/minima (SDM) by Ginzburg. The presence of an SDM forces the existence of infinitely many closed orbits (cf. [65,66] for the case of geodesics on S^2).

We conclude this section with the following (clearly debatable but rather convincing from the above story) meta-mathematical claim: *the three-body problem inspired large portions of modern symplectic geometry*. In all probability, it would also be fair to make the same claim for most of the modern theory of dynamical systems.

Final remark on different approaches Amongst the approaches that we have discussed (by all means non-exhaustive), we point out that the advantage of KAM theory (in the pertubative case), when compared to more abstract approaches via general fixed point theorems, is that in favourable situations, one can localize periodic (or quasi-periodic) orbits in bounded regions of phase space, and obtain better qualitative information on these. This is, of course, much more complicated in non-perturbative situations, where rigorous numerics is usually the preferred approach. See [47] for examples of return maps on a disk-like global surface of section, obtained numerically, for the planar problem.

More references A nice basic introduction to the classical KAM theorem is, e.g., [125]. Another very nice exposition on the basics behind Mather theory is, e.g., [118]. A beautiful and very detailed account on the three-body problem and Poincaré's work are the notes by Chenciner [25]. A very recent and detailed survey on open questions on geodesics, illustrating the vastness and richness of their search, is that of Burns and Matveev [22]. I also based parts of the above brief survey on very nice lectures by Nancy Hingston given at the summer school "Current Trends in Symplectic Topology", July 2019, at the Centre de recherches mathématiques, Université de Montréal, Canada; where I happened to be in the audience. Of course, this is a classical story and there are plenty of other sources; see, e.g., Oancea's much more detailed account [106] and references therein (as well as the appendix due to Hrynewicz on the story for S^2), with a view towards symplectic geometry.

6. Contact geometry in the restricted three-body problem

The next result opens up the possibility of using modern techniques from contact and symplectic geometry on the CR3BP (holomorphic curves, Floer theory,...). Denote by $\overline{\Sigma}_c^E$ and $\overline{\Sigma}_c^M$ the bounded components of the Moser-regularized energy hypersurfaces for the spatial problem and $c < H(L_1)$, and let $\overline{\Sigma}_c^{E,M}$ be the connected sum bounded component, for $c \in (H(L_1), H(L_2))$. Similarly, use $\overline{\Sigma}_{P,c}^E$, $\overline{\Sigma}_{P,c}^M$ and $\overline{\Sigma}_{P,c}^{E,M}$ for the case of the planar problem.

Theorem K. ([7] (planar problem), [26] (spatial problem)) *If $c < H(L_1)$, the Moser-regularized energy hypersurfaces $\overline{\Sigma}_c^E, \overline{\Sigma}_c^M, \overline{\Sigma}_{P,c}^E, \overline{\Sigma}_{P,c}^M$ are all contact-type. The same holds for $\overline{\Sigma}_c^{E,M}, \overline{\Sigma}_{P,c}^{E,M}$, if $c \in (H(L_1), H(L_1)+\epsilon)$ for sufficiently small $\epsilon > 0$. As contact manifolds, we have*

$$\overline{\Sigma}_c^E \cong \overline{\Sigma}_c^M \cong (S^*S^3, \xi_{\mathrm{std}}), \text{ if } c < H(L_1),$$
$$\overline{\Sigma}_{P,c}^E \cong \overline{\Sigma}_{P,c}^M \cong (S^*S^2, \xi_{\mathrm{std}}), \text{ if } c < H(L_1),$$

and

$$\overline{\Sigma}_c^{E,M} \cong (S^*S^3, \xi_{\mathrm{std}}) \# (S^*S^3, \xi_{\mathrm{std}}), \text{ if } c \in (H(L_1), H(L_1)+\epsilon).$$
$$\overline{\Sigma}_{P,c}^{E,M} \cong (S^*S^2, \xi_{\mathrm{std}}) \# (S^*S^2, \xi_{\mathrm{std}}), \text{ if } c \in (H(L_1), H(L_1)+\epsilon).$$

In all above cases, the planar problem is a codimension-2 contact submanifold of the spatial problem. □

Recall that the above just means that there exists a Liouville vector field which is transverse to the regularized level sets; in fact, this is just the fiber-wise Liouville vector field $q\partial_q$. The regularized level sets, as contact manifolds, are standard and well known, so not very interesting from a geometrical perspective. However, their interest lies in the given non-standard *dynamics* for the underlying standard geometry. The Hamiltonian dynamics for the problem now becomes the Reeb dynamics, and the planar problem (from a dynamical perspective rather than a geometric one) is actually invariant under the Reeb flow. We will refer as the *low-energy range* to the interval $(-\infty, H(L_1)+\epsilon)$ of energies c for which the above result holds.

Remark 6.1. The contact condition is in fact lost for sufficiently high Jacobi constant c; see [104].

Remark 6.2. (Weinstein handles) In the above statement, the connected sum is to be interpreted in the contact category; this amounts to attaching a *Weinstein* 1-handle to the disjoint union of two copies of $(S^*S^3, \xi_{\mathrm{std}})$. Roughly speaking, this means removing two Darboux balls and identifying their boundaries via attaching a 1-handle, which is endowed with the extra structure of a symplectic form which glues well to the symplectization form of the standard contact form at the boundary of each ball. The result is a *Liouville/Weinstein cobordism* having $(S^*S^3, \xi_{\mathrm{std}}) \bigsqcup (S^*S^3, \xi_{\mathrm{std}})$ at the negative end, and $(S^*S^3, \xi_{\mathrm{std}}) \# (S^*S^3, \xi_{\mathrm{std}})$ at the positive one. Note that here the terms positive/negative are relevant: the Liouville vector field is outwards/inwards pointing at the corresponding boundary components, respectively, and so these cobordisms are oriented. This is always the local Morse-theoretical picture for a non-degenerate index 1 critical point of a Hamiltonian (as is the case of L_1). To learn about Weinstein manifolds, see, e.g., [27]; this source also provides deep connections between this notion and that of Stein manifolds.

References For a very detailed and well-exposed overview of contact geometry and holomorphic curves in the planar case of the CR3BP, we refer to Frauenfelder–van Koert [47]. Indeed, the subject of this book is precisely the direction outlined in this document, but focused on the planar problem, and so the reader is specially encouraged to delve in it.

6.1. Non-perturbative methods: holomorphic curves

We now discuss the non-perturbative approach coming from the theory of holomorphic curves.

Hofer–Wysocki–Zehnder We begin with a definition. A connected compact hypersurface $\Sigma \subset \mathbb{R}^4$ is said to *be strictly convex* if there exists a domain $W \subset \mathbb{R}^4$ and a smooth function $\phi : \mathbb{R}^4 \to \mathbb{R}$ satisfying:

(i) (Regularity) $\Sigma = \{\phi = 0\}$ is a regular level set;
(ii) (Bounded domain) $W = \{z \in \mathbb{R}^4 : \phi(z) \leq 0\}$ is bounded and contains the origin; and
(iii) (Positive-definite Hessian) $\nabla^2\phi_z(h, h) > 0$ for $z \in W$ and for each non-zero tangent vector $h \in T\Sigma$.

In this case, the radial vector field is transverse to Σ, and so Σ is a contact-type 3-sphere, inheriting a contact form α induced by the standard Liouville form in $\mathbb{R}^4$.

Remark 6.3. In the planar restricted three-body problem, the values of energy/mass ratio (c, μ) for which the Levi–Civita regularization is dynamically convex is called the *convexity range*. This is implied by strict convexity. See the following page for a precise definition of dynamical convexity.

In [73], Hofer–Wysocki–Zehnder prove the following:

Theorem L. [73] *A strictly convex hypersurface* $(\Sigma, \alpha) \subset \mathbb{R}^4$ *has either* 2 *or infinitely many periodic orbits.*

The strategy of the proof is finding a disk-like global surface of section, and use the combination Brouwer–Franks mentioned as a heuristic above. The difficulty is precisely finding the section. These are to be thought of as the (holomorphic) pages of a trivial open book on $\Sigma \cong S^3 = \mathbf{OB}(\mathbb{D}^2, \mathbb{1})$, which is adapted to the given Reeb dynamics. The rough idea is as follows.

Consider the symplectization $(M, \omega) = (\mathbb{R} \times \Sigma, d(e^t\alpha))$ of (Σ, α). Its tangent space splits as $TM = \xi \oplus \langle \partial_t, R_\alpha \rangle$. A (cylindrical, α-compatible) almost complex structure is an endomorphism $J \in \mathrm{End}(TM)$ satisfying:

- $J^2 = -\mathbb{1}$ (i.e. J is a "90-degree rotation" at each tangent space);
- $J(\xi) = \xi$, $J(\partial_t) = R_\alpha$;

- J is $\mathbb{R}$-invariant;
- $g = d\alpha(\cdot, J\cdot)$ defines a J-invariant Riemannian metric on ξ.

A *J-holomorphic plane* is then a map

$$u : (\mathbb{C}, i) \to (M, J),$$

intertwining the complex structures, i.e., satisfying the non-linear *Cauchy–Riemann* equation

$$J \circ du = du \circ i.$$

The *Hofer-energy* of such a plane is the quantity

$$\mathbf{E}(u) = \sup_{\varphi \in \mathcal{P}} \int_{\mathbb{C}} u^* \omega_\varphi,$$

where $\mathcal{P} = \{\varphi : \mathbb{R} \to (0,1) : \varphi' \geq 0\}$ is the set of orientation preserving diffeomorphisms between $\mathbb{R}$ and $(0,1)$, and $\omega_\varphi = d(e^{\varphi(t)}\alpha)$ is a symplectic form. The choice of J implies that the integrand is point-wise non-negative and so $\mathbf{E}(u) \geq 0$. A fundamental property is that non-constant finite energy J-holomorphic planes are asymptotic to closed Reeb orbits (originally noted by Hofer in his proof of the Weinstein conjecture for overtwisted contact 3-manifolds):

Proposition 6.4. [69] *If $\mathbf{E}(u) < +\infty$ and $u = (a, v) \in \mathbb{R} \times \Sigma$ is non-constant, then $0 < \int v^* d\alpha := T < +\infty$, and there exists a sequence $R_k \to +\infty$, such that $\lim_{k \to +\infty} u(R_k e^{2\pi i t}) = \gamma(tT)$, for a closed Reeb orbit γ.*

Moreover, under a non-degeneracy condition for γ, the above convergence is exponential and $\lim_{R \to +\infty} u(Re^{2\pi i t}) = \gamma(tT)$, $\lim_{R \to +\infty} a(Re^{2\pi i t}) = +\infty$. A further fundamental property is *positivity of intersections*; since M is 4-dimensional, generically two planes intersect at a finite number of points, and if they are holomorphic, the intersection numbers are positive. However, there is an an obvious drawback: planes are non-compact, and so, the classical intersection pairing is not homotopy invariant, since intersections can disappear to infinity. The solution to this issue was provided by Siefring [116], who, using the very explicit asymptotic behaviour of finite energy planes, defined an intersection pairing with all the desired properties. In particular, it is homotopy invariant, takes into consideration interior intersections as well as those "coming from infinity", and two holomorphic planes have vanishing Siefring intersection if and only if their images do not intersect at all. Moreover, in such a case, their projections to Σ do not intersect unless their images coincide. (As the attentive reader might have already noticed, Siefring's work is posterior to the above result; but we will ignore this for the purposes of this rough discussion.)

With these preambles, the main idea for the proof of Theorem L is as follows. One assumes the existence of a special Reeb orbit γ, in the sense that is unknotted and linked to every other Reeb orbit (necessary conditions to be the binding of a trivial open book for S^3), non-degenerate, has minimal period, and satisfies $\mu_{CZ}(\gamma) = 3$. Here, we use the *Conley–Zehnder* index μ_{CZ}, which is roughly speaking a winding number associated with the paths of symplectic matrices which are suitably non-degenerate, and is used to assign to every Reeb orbit γ an integer $\mu_{CZ}(\gamma)$ (which depends on a trivialization of the tangent bundle along a choice of disk bounded by γ; in the case of S^3, where $\pi_2(S^3) = 0$, this is independent on choices). One then considers the moduli space $\mathcal{M}$ of finite energy J-holomorphic planes asymptotic to this Reeb orbit γ, and having vanishing Siefring self-intersection, modulo the action of $\mathbb{R}$-translation in the image (recall J is $\mathbb{R}$-invariant) and conformal reparametrizations of the domain $\mathbb{C}$. Its expected dimension is $\dim \mathcal{M} = \mu_{CZ}(\gamma) - 2 = 1$, by the Riemann–Roch formula for the Fredholm index. Moreover, the miraculous 4-dimensional phenomenon of automatic transversality shows that $\mathcal{M}$ is a manifold for any cylindrical J. The properties of the Siefring pairing imply that the projections of planes in $\mathcal{M}$ are immersed, do not intersect, and provide a local foliation of Σ. A further step needed in order to get a global foliation is a way to compactify $\mathcal{M}$. This is provided by Gromov's compactification (or the SFT compactification), obtained by adding strata of nodal curves and "holomorphic buildings" with potentially several "floors"; strictly speaking, these a priori are no longer planes. However, the fact that γ is linked to every other orbit can be used to show that no extra strata needs to be added to $\mathcal{M}$, and is in fact *a priori* compact. The result is that $\mathcal{M} \cong S^1$, and projecting the planes in $\mathcal{M}$ to Σ provides a global foliation of Σ. The leaves of this foliation are the S^1-family of pages of an open book with binding γ, and are in fact global surfaces of section for the Reeb dynamics.

While the assumption on the existence of γ above might seem far-fetched, it is implied by *dynamical convexity* [73, Theorem 1.3]. One says that (Σ, α) is dynamically convex if $\mu_{CZ}(\gamma) \geq 3$ for Reeb every orbit γ. This condition is implied by strict convexity [73, Theorem 3.4]; intuitively, this implies that there is "enough winding" of the linearized Reeb flow along each orbit (and so, at the end of the day when the open book is obtained, this condition applied to the binding γ implies that the arising return map extends to the boundary). The special Reeb orbit is found by first considering the case of an ellipsoid, in which it is explicitly found, then interpolating to the dynamically convex case by considering a symplectic cobordism, and finally using properties of finite energy planes in cobordisms; see Section 4 in [73].

Conclusion The main message to take away from this discussion is that the global surfaces of section are the (holomorphic) pages of a trivial open book on $\Sigma \cong S^3 = \mathbf{OB}(\mathbb{D}^2, \mathbb{1})$, which is *a posteriori* adapted to the given Reeb dynamics. The way that this result ties up with the planar CR3BP is via the Levi–Civita regularization; one says that (μ, c) lies in the convexity range whenever the Levi–Civita regularization is dynamically convex (cf. Proposition 4.3). The holomorphic open book provided by Hofer–Wysocki–Zehnder, given suitable symmetries, descends to a *rational* open book on the Moser-regularized hypersurface $\mathbb{R}P^3$ (i.e. the pages are disks, but their boundary is doubly covered). Alternatively, [77, Theorem 1.18] provides an honest open book with annuli fibers for $\mathbb{R}P^3 = \mathbf{OB}(\mathbb{D}^*S^1, \tau^2)$, adapted to the planar dynamics. This circle of ideas has also been fruitfully exploited in e.g. [74–76]; see [78] for a very nice survey and references therein, especially for the applications on the planar CR3BP.

7. Holomorphic curve techniques on the spatial CR3BP

In this section, we present some (yet unpublished) results of the author, in co-authorship with Otto van Koert. The main direction is to generalize the approach of Poincaré in the planar problem [i.e., Steps (1) and (2) outlined above] to the *spatial* problem.

7.1. Step (1): Global hypersurfaces of section

We first state a structural result, which provides the basic architecture and scaffolding for the problem:

Theorem M. (Moreno–van Koert [98]) *Fix a mass ratio $\mu \in (0,1]$. Denote a connected, bounded component of the regularized, spatial, circular restricted three-body problem for energy level c by Σ_c. Then, Σ_c is of contact-type and admits a supporting open book decomposition for energies $c < H(L_1)$ that is adapted to the Hamiltonian dynamics. Furthermore, if $\mu < 1$, then there is $\epsilon > 0$, such that the same holds for $c \in (H(L_1), H(L_1)+\epsilon)$. The open books have the following abstract form:*

$$\Sigma_c \cong \begin{cases} \boldsymbol{OB}(\mathbb{D}^*S^2, \tau^2), & \text{if } c < H(L_1) \\ \boldsymbol{OB}(\mathbb{D}^*S^2 \natural \mathbb{D}^*S^2, \tau_1^2 \circ \tau_2^2), & \text{if } c \in (H(L_1), H(L_1)+\epsilon) \text{ and } \mu < 1. \end{cases}$$

Here, $\mathbb{D}^*S^2$ is the unit cotangent bundle of the 2-sphere, τ is the positive Dehn–Seidel twist along the Lagrangian zero section $S^2 \subset \mathbb{D}^*S^2$, and $\mathbb{D}^*S^2 \natural \mathbb{D}^*S^2$ denotes the boundary connected sum of two copies of $\mathbb{D}^*S^2$. The monodromy of the second open book is the composition of the square of the positive Dehn–Seidel twists along both zero sections (they commute). The binding is the planar problem $\Sigma_c^P \cong \mathbb{R}P^3$.

See Fig. 13 for an abstract representation (see also Fig. 14). We wish to emphasize that Theorem M holds for c in the whole low-energy range. A heuristical reason is the following: while in the planar case finding the invariant subset is non-trivial (the search for the direct and retrograde orbits indeed has a long history), the invariant subset in the spatial case is immediately obvious; it is the planar problem. The technique of proof does not rely on holomorphic curves, since one can directly write down the open book explicitly; it is rather elementary, but the calculations are very involved.

The above result is motivated by the following observation. We consider a Stark–Zeemaan system satisfying Assumptions (A1) and (A2). In unregularized (or physical) coordinates, we put

$$B_u := \{(\vec{q}, \vec{p}) \in H^{-1}(c) \mid q_3 = p_3 = 0\},$$

the planar problem. Its normal bundle is trivial, and we have the following map to S^1:

$$\pi_u : H^{-1}(c) \backslash B_u \longrightarrow S^1, (\vec{q}, \vec{p}) \longmapsto \frac{q_3 + ip_3}{\|q_3 + ip_3\|}. \tag{7.1}$$

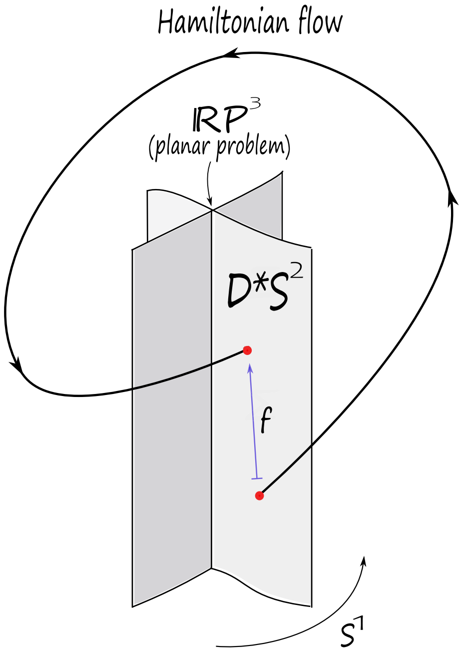


FIGURE 13. The open book for Σ_c, with $c < H(L_1)$, and the first return map f

We will refer to this map as the *physical* open book. We consider the angular 1-form

$$\mathrm{d}\pi_u := \frac{\Omega_p^u}{p_3^2 + q_3^2},$$

where

$$\Omega_p^u = p_3 \mathrm{d}q_3 - q_3 \mathrm{d}p_3, \tag{7.2}$$

is the unregularized numerator. We need to see whether $d\pi_u(X_H)$ is non-negative, and vanishes only along the planar problem.

From Eq. (4.1), we have

$$\mathrm{d}\pi_u(X_H) = \frac{p_3^2 + q_3^2\left(\frac{g}{\|\vec{q}\|^3} + \frac{1}{q_3}\frac{\partial V_1}{\partial q_3}(\vec{q})\right)}{p_3^2 + q_3^2}. \tag{7.3}$$

Note that Assumption (A2) implies that $\frac{\partial V_1}{\partial q_3}(\vec{q}) = aq_3 + O(q_3^2)$ near $q_3 = 0$, and so, $\frac{1}{q_3}\frac{\partial V_1}{\partial q_3}(\vec{q})$ is well defined at $q_3 = 0$. In order for the above expression to satisfy the required non-negativity condition, we impose the following:

Assumption. (A3) We assume that the function

$$F(\vec{q}) = \frac{g}{\|\vec{q}\|^3} + \frac{1}{q_3}\frac{\partial V_1}{\partial q_3}(\vec{q})$$

is everywhere positive.

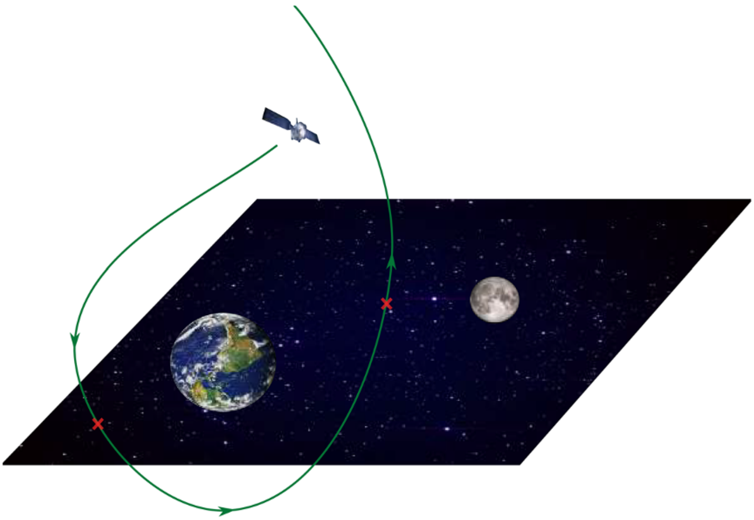


FIGURE 14. Theorem M admits a physical interpretation: away from collisions, the orbits of the negligible mass point intersect the plane containing the primaries transversely. This is intuitively clear from a physical perspective, and translates (after regularization) to the fact that the "pages" $\{q_3 = 0, p_3 > 0\}$, $\{q_3 = 0, p_3 < 0\}$ of the "physical" open book are global hypersurfaces of section outside of the collision locus. Unfortunately, this does not extend continuously to the latter, as explained in Fig. 15. The binding is the planar problem

Note that it suffices that the second summand be non-negative.

Remark 7.1. In the restricted three-body problem, from Eq. (4.5), we obtain

$$\frac{\partial V_1}{\partial q_3}(\vec{q}) = q_3 \frac{1-\mu}{\|\vec{q}-\vec{e}\|^3},$$

and therefore, the corresponding expression in Eq. (7.3) is non-negative, vanishing if and only if $p_3 = q_3 = 0$.

The obvious problem of the above computation is that it a priori does not extend to the collision locus, and indeed, it cannot (see Fig. 15). In fact, one needs to interpolate with the *geodesic* open book described in Sect. 2.6, which is well behaved near the collision locus. This creates an interpolation region where fine estimates are needed, and this is the main difficulty in the proof; we refer to [98] for the details.

Symmetries Consider the symplectic involution of $(\mathbb{R}^6, dp \wedge dq)$ given by

$$r : (q_1, q_2, q_3, p_1, p_2, p_3) \mapsto (q_1, q_2, -q_3, p_1, p_2, -p_3).$$

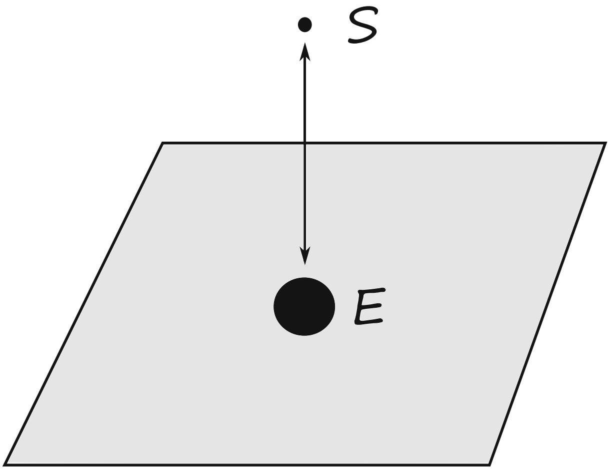


FIGURE 15. For the rotating Kepler problem, there exist (regularized) collision orbits which are periodic and "bounce" vertically over a primary, always staying on the region $q_3 > 0$ (or $q_3 < 0$). We call them the *polar* orbits. This means that the "pages" $\{q_3 = 0, p_3 > 0\}$, $\{q_3 = 0, p_3 < 0\}$ are *not* transverse to the regularized dynamics

We also have the anti-symplectic involutions

$$\rho_1 : (q_1, q_2, q_3, p_1, p_2, p_3) \mapsto (q_1, -q_2, -q_3, -p_1, p_2, p_3)$$
$$\rho_2 : (q_1, q_2, q_3, p_1, p_2, p_3) \mapsto (q_1, -q_2, q_3, -p_1, p_2, -p_3),$$

satisfying the relations $\rho_1 \circ \rho_2 = \rho_2 \circ \rho_1 = r$, and so generating the abelian group $\{1, r, \rho_1, \rho_2\} \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$, which is the natural symmetry group of the spatial circular restricted three-body problem.

After regularization, the symplectic involution admits the following intrinsic description. Consider the smooth reflection $R : S^3 \to S^3$ along the equatorial sphere $S^2 \subset S^3$. Then, r is the physical transformation it induces on T^*S^3, given by

$$r : T^*S^3 \to T^*S^3$$
$$r(q, p) = (R(q), [(d_qR)^*]^{-1}(p)).$$

This map preserves the unit cotangent bundle S^*S^3. The maps ρ_1, ρ_2 also have regularized versions. The following emphasizes the symmetries present in our setup:

Proposition 7.2. [98] *Let $c < H(L_1)$, and consider the symplectic involution $r : S^*S^3 \to S^*S^3$. The open book decomposition $\Sigma_c = \mathbf{OB}(\mathbb{D}^*S^2, \tau^2)$ is symmetric with respect to r, in the sense that*

$$r(P_\theta) = P_{\theta+\pi}, \quad Fix(r) = B = \Sigma_c^P.$$

Moreover, the anti-symplectic involutions preserve B and satisfy

$$\rho_1(P_\theta) = P_{-\theta}, \ \rho_2(P_\theta) = P_{-\theta+\pi}.$$

In particular, ρ_1 *preserves* P_0 *and* P_π*, whereas* ρ_2 *preserves* $P_{\pi/2}$ *and* $P_{-\pi/2}$*.*

In other words, the open book is compatible with all the symmetry group $\mathbb{Z}_2 \oplus \mathbb{Z}_2$.

The return map First, we recall a standard definition. We say that a symplectomorphism $f : (M, \omega) \to (M, \omega)$ is *Hamiltonian* if $f = \phi_K^1$, where $K : \mathbb{R} \times M \to \mathbb{R}$ is a smooth (time-dependent) Hamiltonian, and ϕ_K^t is the Hamiltonian isotopy it generates. This is defined by $\phi_K^0 = id$, $\frac{d}{dt}\phi_K^t = X_{K_t} \circ \phi_K^t$, and X_{H_t} is the Hamiltonian vector field of H_t defined via $i_{X_{H_t}}\omega = -dH_t$. Here, we write $K_t = K(t, \cdot)$.

In the SCR3BP, for $c < H(L_1)$, and after fixing a page $P = \pi^{-1}(1)$ of the corresponding open book, Theorem M implies the existence of a Poincaré return map $f : \mathrm{int}(P) \to \mathrm{int}(P)$. Moreover, as in Proposition 2.16, we can consider the 2-form ω obtained by restriction to P of $d\alpha$, where α is the contact form on Σ_c for the spatial problem, whose restriction to the binding α_P is the contact form for the planar problem (cf. Fig. 16). Recall that ω is symplectic only along the *interior* of P (which may be thought of as an *ideal* Liouville domain). Moreover, we have a smooth identification $\mathrm{int}(P) \cong \mathrm{int}(\mathbb{D}^* S^2)$, giving a symplectomorphism $G : \mathrm{int}(P) \to \mathrm{int}(\mathbb{D}^* S^2)$ on the interior which extends smoothly to the boundary B, but its inverse G^{-1}, although continuous at B, is *not* differentiable along B since ω becomes degenerate there. After conjugating f with G and considering $\widetilde{\omega} = G_*\omega$, we obtain a symplectomorphism $\widetilde{f} := G \circ f \circ G^{-1} : (\mathrm{int}(\mathbb{D}^* S^2), \widetilde{\omega}) \to (\mathrm{int}(\mathbb{D}^* S^2), \widetilde{\omega})$, where $\widetilde{\omega}$ is a Liouville filling of (B, α_P). In particular, $\widetilde{\omega}$ is non-degenerate at B.

Theorem N. (Moreno–van Koert [98]) *For every* $\mu \in (0, 1]$, $c < H(L_1)$*, the associated Poincaré return map* f *extends smoothly to the boundary* ∂P*, and in the interior, it is an exact symplectomorphism*

$$f = f_{c,\mu} : (int(P), \omega) \to (int(P), \omega),$$

where $\omega = d\alpha$ *(depending on* c, μ*). Moreover,* f *is Hamiltonian in the interior.*

After conjugating with G, $\widetilde{f}$ *extends continuously to the boundary, is Hamiltonian in the interior, and the Liouville completion of* $\widetilde{\omega}$ *is symplectomorphic to the standard symplectic form* ω_{std} *on* T^*S^2*.*

The fact that f is an exact symplectomorphism follows from Proposition 2.16. The fact that f extends to the boundary is non-trivial, and relies on second-order estimates near the binding: it suffices to show that the Hamiltonian giving the spatial problem is positive definite on the symplectic normal bundle to the binding. This non-degeneracy condition can be interpreted as a convexity condition that plays the role, in this setup, of the notion of dynamical convexity due to Hofer–Wysocki–Zehnder [73]. Note that if a continuous extension exists, then by continuity, it is unique.

The fact that f is Hamiltonian in the interior follows from:

(1) The monodromy of the open book is Hamiltonian (here, the Hamiltonian is allowed to move the boundary).
(2) The general fact that the return map f is always symplectically isotopic to a representative of the monodromy, via a boundary-preserving isotopy.

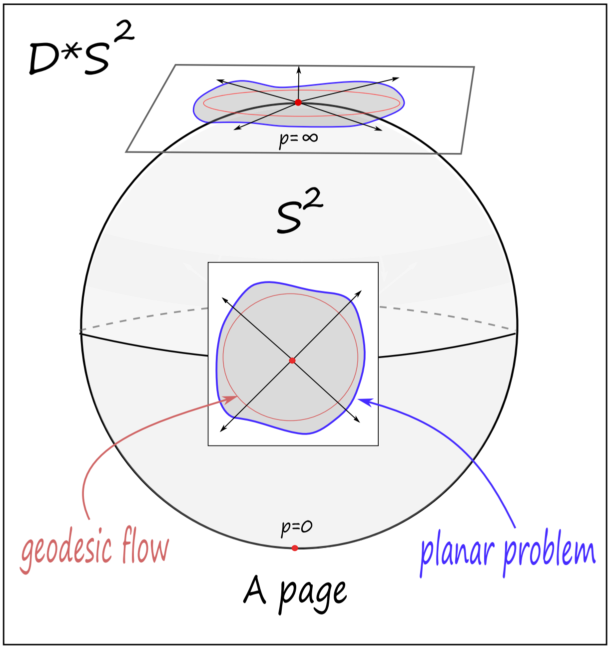


FIGURE 16. A page of the open book as a symplectic filling of the planar problem, viewed as a fiber-wise star-shaped domain in T^*S^2. The geodesic flow corresponds to the unit cotangent bundle

(3) $H^1(P;\mathbb{R}) = 0$, so that every symplectic isotopy is Hamiltonian.

7.2. Step (2): Fixed-point theory of Hamiltonian twist maps

The periodic points of τ are either boundary periodic points, which give planar orbits, or interior periodic points which are in 1:1 correspondence with spatial orbits. We are interested in finding *interior* periodic points, and we follow Poincaré's philosophy to try to find them.

The Hamiltonian twist condition We propose a generalization of the twist condition introduced by Poincaré, for the Hamiltonian case and for arbitrary Liouville domains. Let $(W, \omega = d\lambda)$ be a $2n$-dimensional Liouville domain, and consider a Hamiltonian symplectomorphism τ. Let $(B, \xi) = (\partial W, \ker \alpha)$ be the contact manifold at the boundary where $\alpha = \lambda|_B$, and R_α the Reeb vector field of α. The Liouville vector field V_λ is defined via $i_{V_\lambda}\omega = \lambda$.

Definition 7.3. (*Hamiltonian twist map*) We say that τ is a *Hamiltonian twist map* (with respect to α), if τ is generated by a *smooth* Hamiltonian $H : \mathbb{R} \times W \to \mathbb{R}$ which satisfies $X_{H_t}|_B = h_t R_\alpha$ for some *positive* and smooth function $h : \mathbb{R} \times B \to \mathbb{R}^+$.

In particular, $H_t|_B \equiv$ *const* on B, and $\tau(B) \subset B$. We have $h_t = dH_t(V_\lambda)|_B$ is the derivative of H_t in the Liouville direction V_λ along B, which we assume strictly positive. Also, $\tau|_B$ is the time-1 map of a positive reparametrization of the Reeb flow on B. But, note that, while the latter condition is only localized at B, the twist condition is of a *global* nature, as it requires global smoothness of the generating Hamiltonian.

Here is a simple example illustrating why the smoothness of the Hamiltonian is relevant for the purposes of fixed points:

Example 7.4. (Integrable twist maps) Let $M = S^n$ for $n \geq 1$ with the round metric, and $H : T^*M \to \mathbb{R}$, $H(q,p) = 2\pi|p|$ (*not* smooth at the zero section); ϕ^1_H extends to all of $\mathbb{D}^*M$ as the identity. It is a positive reparametrization of the Reeb flow at S^*M, a full turn of the geodesic flow, and all orbits are fixed points with fixed period. If we smoothen H near $|p| = 0$ to $K(q,p) = 2\pi g(|p|)$, with $g(0) = g'(0) = 0$, then $\tau = \phi^1_K : \mathbb{D}^*M \to \mathbb{D}^*M$, $\tau(q,p) = \phi_H^{2\pi g'(|p|)}(q,p)$, is now a Hamiltonian twist map. If $g'(|p|) = l/k \in \mathbb{Q}$ with l, k coprime, then τ has a simple k-periodic orbit; therefore, τ has simple interior orbits of arbitrary large period (cf. [83, p. 350], [102], for the case $M = S^1$).

Remark 7.5. In what follows, we shall appeal to the symplectic homology (or the Floer homology) of a Liouville domain (W, λ), denoted $SH_\bullet(W, \lambda)$. This is a homology theory, which keeps track of both dynamical and topological data; it is, roughly speaking, the homology of a chain complex generated by critical points of a Morse function on the interior of W, as well as by Reeb orbits at the boundary ∂W. These are the 1-periodic orbits of an *admissible* Hamiltonian, i.e., linear at infinity and C^2-small and Morse in the interior. Formally, one needs to take a direct limit over admissible Hamiltonians whose slope increases to infinity, so that we capture orbits at the boundary with all possible periods. The grading in symplectic homology comes from the Conley–Zehnder index (whenever orbits are non-degenerate); for the degenerate case, one can also use the Robbin–Salamon index. The details behind its definition are beyond the scope of this survey; we refer, e.g., to [18, 28].

The Hamiltonian twist condition will be used to extend the Hamiltonian to a Hamiltonian that is admissible for computing symplectic homology. The extended Hamiltonian can have additional 1-periodic orbits and these, as well as 1-periodic orbits on the boundary, need be distinguished from the interior periodic points of τ. We impose the following conditions to do so.

Index growth We consider a suitable index growth condition on the dynamics on the boundary, which is satisfied in the three-body problem whenever the *planar* dynamics is strictly convex. This assumption will allow us to separate boundary and extension orbits from interior ones via the index.

We call a strict contact manifold $(Y, \xi = \ker \alpha)$ *strongly index-definite* if the contact structure $(\xi, d\alpha)$ admits a symplectic trivialization ϵ with the property that

- There are constants $c > 0$ and $d \in \mathbb{R}$, such that for every Reeb chord $\gamma : [0,T] \to Y$ of Reeb action $T = \int_0^T \gamma^*\alpha$, we have
$$|\mu_{RS}(\gamma; \epsilon)| \geq cT + d,$$
where μ_{RS} is the Robbin–Salamon index [113].

Index-positivity is defined similarly, where we drop the absolute value. A variation of this notion was explored in Ustilovsky's thesis [120]. He imposed the additional condition $\pi_1(Y) = 0$, so that index-positivity becomes independent of the choice of trivialization, although the exact constants c and d still depend on the trivialization ϵ. The global trivialization is important when considering extensions of our Hamiltonians, as it allows us to measure the index growth of potential new orbits. The point in the above definition is that the index of boundary orbits grows to infinity under iterations of our return map, and so, these do not contribute to symplectic homology.

A general condition for index-positivity to hold, which is also relevant for the restricted three-body problem, is the following:

Lemma 7.6. *Suppose that (Σ, α) is a strictly convex hypersurface in $\mathbb{R}^4$. Then, (Σ, α) is strongly index-positive.*

Fixed-point theorems We propose the following generalization of the Poincaré–Birkhoff theorem:

Theorem O. (Moreno–van Koert [99]. Generalized Poincaré–Birkhoff theorem) *Suppose that τ is an exact symplectomorphism of a connected Liouville domain (W, λ), and let $\alpha = \lambda|_B$. Assume the following:*

- *(Hamiltonian twist map) τ is a Hamiltonian twist map, where the generating Hamiltonian is at least C^2. In addition, assume all fixed points of τ are isolated;*
- *(Index-definiteness) If $\dim W \geq 4$, then assume $c_1(W)|_{\pi_2(W)} = 0$, and $(\partial W, \alpha)$ is strongly index-definite;*
- *(Symplectic homology) $SH_\bullet(W)$ is infinite dimensional.*

Then, τ has simple interior periodic points of arbitrarily large (integer) period.

Remark 7.7. Let us discuss some aspects of the theorem:

(1) (Grading) We need impose the assumptions $c_1(W)|_{\pi_2(W)} = 0$ (i.e., W is symplectic Calabi–Yau) to have a well-defined integer grading on symplectic homology.

(2) (Surfaces) If W is a surface, then the condition that $SH_\bullet(W)$ is infinite dimensional just means that $W \neq D^2$; for D^2, we have $SH_\bullet(D^2) = 0$, and a rotation on D^2 gives an obvious counterexample to the conclusion. In the surface case, the argument simplifies, and one can simply work with homotopy classes of loops rather than the grading on symplectic homology. The Hamiltonian twist condition recovers the classical twist condition for $W = \mathbb{D}^* S^1$, due to orientations, and hence, the above is clearly a version of the classical Poincaré–Birkhoff theorem.

(3) (Cotangent bundles) The symplectic homology of the cotangent bundle of a closed manifold is infinite dimensional, due to a result of Viterbo [122,123] (see also [1,114]), combined, e.g., with a theorem of Gromov [60, Sec. 1.4]. We have $c_1(T^*M) = 0$ whenever M is orientable. As for the existence of a global trivialization of the contact structure $(\xi, d\lambda_{can})$, we note:

- if Σ is an oriented surface, then $S^*\Sigma$ admits such a global symplectic trivialization;
- if M^3 is an orientable 3-manifold, then S^*M^3 also admits such a global symplectic trivialization;
- symplectic trivializations of the contact structure on (S^*S^2, λ_{can}) are unique up to homotopy.

(4) (Fixed points) If fixed points are non-isolated, then we vacuously obtain infinitely many of them, although we cannot conclude that their periods are arbitrarily large; "generically", one expects finitely many fixed points.

(5) (Long orbits) If W is a global hypersurface of section for some Reeb dynamics, with return map τ, interior periodic points with long (integer) period for τ translates into spatial Reeb orbits with long (real) period. See Appendix C in [99].

(6) (Katok examples) There are well-known examples due to Katok [82] of Finsler metrics on spheres with only finitely many simple geodesics, which are arbitrarily close to the round metric. Moreover, they admit global hypersurfaces of section with Hamiltonian return maps, for which the index-definiteness and the condition on symplectic homology both hold. It follows that the return map does not satisfy the twist condition for any choice of Hamiltonians.

(7) (Spatial restricted three-body problem) From the above discussion and [98], we gather: the only standing obstruction for applying the above result to the spatial restricted three-body problem, in case where the planar problem is strictly convex, is the Hamiltonian twist condition. Here, note that symplectic homology is invariant under deformations of Liouville domains; see, e.g., [15] for a paper with detailed proofs. This would give a proof of existence of *spatial* long orbits in the spirit of Conley [29], which could in principle be collision orbits (these may be excluded, at least perturbatively, by different methods). Since the geodesic flow on S^2 arises as a limit case (i.e., the Kepler problem), it should be clear from the discussion on Katok examples that this is a subtle condition. In [98], we have computed a generating Hamiltonian for the integrable case of the rotating Kepler problem; it does *not* satisfy the twist condition in the spatial case (in the planar case, a Hamiltonian twist map was essentially found by Poincaré). This does not mean a priori that there is not *another* generating Hamiltonian which does, but this seems rather unlikely and difficult to check.

As a particular case of Theorem O, we state the above result for star-shaped domains in cotangent bundles, as a case of independent interest (cf. [61]):

Theorem P. (Moreno–van Koert [99]) *Suppose that W is a fiber-wise star-shaped domain in the Liouville manifold (T^*M, λ_{can}), where M is simply connected, orientable and closed, and assume that $\tau : W \to W$ is a Hamiltonian twist map. If the Reeb flow on ∂W is strongly index-positive, and if*

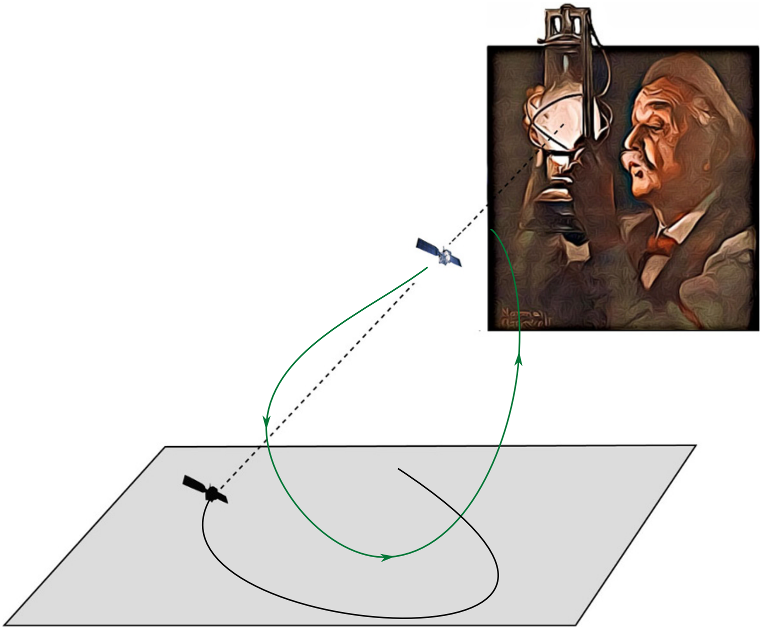


FIGURE 17. Philosophy: To shed some light on a complicated higher dimensional problem, try first to look at the shadow that your lantern is producing!

all fixed points of τ are isolated, then τ has simple interior periodic points of arbitrarily large period.

The above also holds for $M = S^2$, as explained in Remark 7.7 (2). A difference with [61] is that in this setup, we conclude that periodic points are interior, to the expense of imposing index-positivity.

7.3. Alternative approach: dynamics on moduli spaces

An alternative approach to that of a fixed-point theorem is the following construction (see Fig. 17 for the philosophy). We start by recalling that the page $\mathbb{D}^*S^2 = \mathbf{LF}(\mathbb{D}^*S^1, \tau_P^2)$ of the open book of Theorem M has a Lefschetz fibration with genus zero fibers over the 2-disk, with monodromy the Dehn twist τ_P (P here is for "planar", to differentiate from the monodromy τ used for the spatial case; recall Fig. 8). The main geometric observation for what follows is: the leaf space $\mathcal{M}$ of such fibers (i.e., the moduli space parametrizing them) is a copy of S^3. Indeed, each page $\mathbb{D}^*S^2$ of the open book $S^2 \times S^3 = \mathbf{OB}(\mathbb{D}^*S^2, \tau^2)$ is a 2-disk worth of fibers; we moreover have an S^1-family of such pages, all of them sharing the boundary $\mathbb{R}P^3$ (the binding), and such that their Lefschetz fibration all induce the S^1-family of pages of the open book $\mathbb{R}P^3 = \mathbf{OB}(\mathbb{D}^*S^1, \tau_P^2)$. It follows that the leaf space carries the trivial open book $\mathcal{M} = \mathbf{OB}(\mathbb{D}^2, \mathbb{1}) \cong S^3$, whose disk-like page corresponds to the base of the page in $S^2 \times S^3$, and whose binding $\mathcal{M}_B$ is the S^1-family of pages for $\mathbb{R}P^3$. See Figs. 18 and 19.

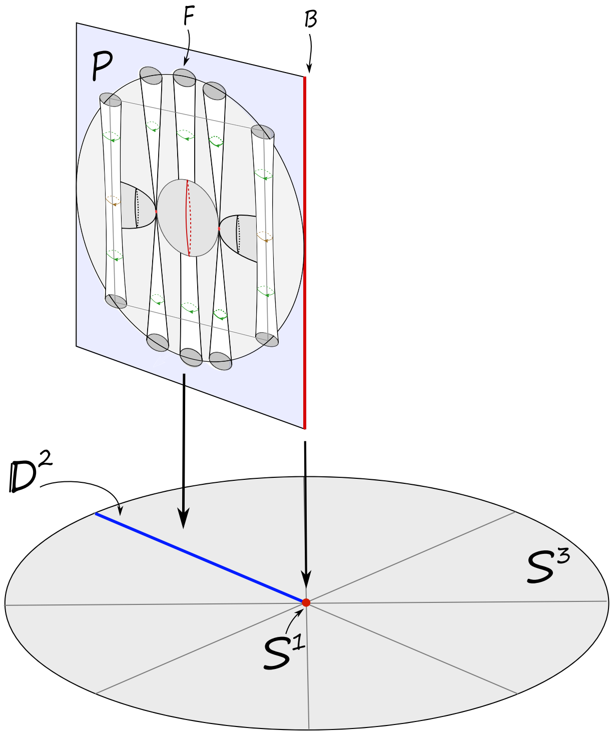


FIGURE 18. The moduli space of curves is a copy of $S^3 = \mathbf{OB}(\mathbb{D}^2, \mathbb{1})$

Rotating Kepler problem In [98, App. A], we discuss the completely integrable limit case of the rotating Kepler problem, where $\mu = 0$, and so, there is only one primary. The return map can be studied explicitly. Geometrically, this map may be understood via the following proposition (recall Fig. 8):

Proposition 7.8. ([98], Integrable case) *In the rotating Kepler problem, the return map f preserves the annuli fibers of the standard Lefschetz fibration $\mathbb{D}^*S^2 = \mathbf{LF}(\mathbb{D}^*S^1, \tau_P^2)$, where it acts as a classical integrable twist map on regular fibers, and fixes the two (unique) nodal singularities on the singular fibers.*

The two fixed points are the north and south poles of the zero section S^2, and correspond to the two periodic collision orbits bouncing on the primary (one for each of the half-planes $q_3 > 0$, $q_3 < 0$).

The abstract case We now consider an abstract situation where the previous argument also holds. Consider a *concrete* open book decomposition $\pi : M \backslash B \to S^1$ on a contact 5-manifold $(M, \xi_M) = \mathbf{OB}(P, \phi)$. We assume that P (abstractly) admits the structure of a 4-dimensional Lefschetz fibration over $\mathbb{D}^2$ whose fibers are surfaces of genus zero and perhaps several boundary components. We abstractly write $P = \mathbf{LF}(F, \phi_F)$, where ϕ_F is the

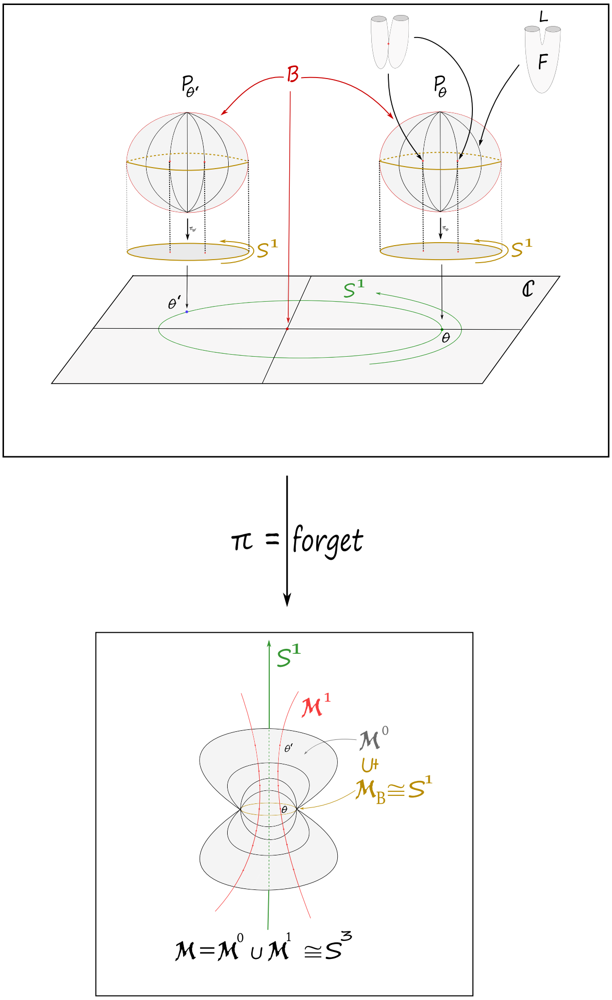


Figure 19. The moduli space $\mathcal{M} \cong S^3$ has two strata: the open strata $\mathcal{M}^0$ consisting of regular fibers, and the nodal strata $\mathcal{M}^1$ consisting of singular fibers

monodromy of the Lefschetz fibration on P (as we have discussed, necessarily a product of positive Dehn twists on the genus zero surface F).

Following [3], we will refer to the open book on M as an *iterated planar* (IP) open book decomposition, and the contact manifold M as iterated planar. As observed in [4, Lemma 4.1], a contact 5–manifold is iterated planar if and only if it admits an open book decomposition supporting the contact structure, whose binding is planar (i.e., admits a 3-dimensional supporting open book whose pages have genus zero). In fact, we have $B = \mathbf{OB}(F, \phi_F)$.

We wish to adapt the underlying planar structure to a *given* Reeb dynamics on M (and hence the need to work with concrete open books, rather than the abstract version). We then assume that the concrete open book on M is adapted to the Reeb dynamics of a *fixed* contact form α_M, i.e., α_M is a Giroux form for the open book (whose dynamics we wish to study). In particular, $\omega_\theta := d\alpha_M|_{P_\theta}$ is a symplectic form on P_θ for each $\theta \in S^1$. Therefore, $(P_\theta, \omega_\theta)$ is a Liouville filling of the binding $(B, \xi_B = \ker \alpha_B)$, where $\alpha_B = \alpha_M|_B$, for each θ. We will further assume that we have a *concrete* planar open book on the 3-manifold $B = \mathbf{OB}(F, \phi_F)$, which is adapted to the Reeb dynamics of α_B and where ϕ_F is a product of positive Dehn twists in the genus zero surface F. We will denote $L = \partial F$, which is a link in B (the binding of the open book for B, and Reeb orbits for α_B). Given the above situation, we will say that the Giroux form α_M is an *IP Giroux form.*

This is precisely the situation in the SCR3BP whenever the planar dynamics is strictly convex/dynamically convex, as follows from [77, Theorem 1.18], combined with Theorem M above. We now state the general construction:

Theorem Q. ([100], IP foliation) *There is a foliation $\mathcal{M}$ of $M\backslash L$, consisting of immersed $d\alpha_M$-holomorphic curves whose boundary is L. Away from B, its elements are arranged as fibers of Lefschetz fibrations $\pi_\theta : P_\theta \to \mathbb{D}^2_\theta$, $\theta \in S^1$, all of which induce the same fixed concrete open book at B. The π_θ are all generic, i.e., each fiber contains at most a single critical point. We have $\mathcal{M} \cong S^3$, and it is endowed with the trivial open book whose θ-page is identified with $\mathbb{D}^2_\theta$, and its binding is $\mathcal{M}_B \cong S^1$, the family of pages of the open book at B.*

The point here is that the above result is in principle non-perturbative; it applies whenever there is an adapted open book at B. It should be thought of as an S^1-parametric version of Wendl's result (Theorem J above), and as the "correct" higher dimensional analogue of the finite energy foliations introduced by Hofer–Wysocki–Zehnder for the study of 3-dimensional Reeb flows. We can further endow the moduli space with extra structure:

Theorem R. ([100], contact and symplectic structures on moduli) *The moduli space $\mathcal{M}$ carries a natural contact structure $\xi_{\mathcal{M}}$ which is supported by the trivial open book on S^3 (and hence it is isotopic to the standard contact structure ξ_{std} by Giroux correspondence). Moreover, the symplectization form on $\mathbb{R} \times M$ associated with any Giroux form α_M on M induces a tautological symplectic form on $\mathbb{R} \times \mathcal{M}$ by leaf-wise integration, which is naturally the symplectization of a contact form $\alpha_{\mathcal{M}}$ for $\xi_{\mathcal{M}}$, whose Reeb flow is adapted to the trivial open book on $\mathcal{M}$.*

The contact form can be written down via the following tautological formula:

$$(\alpha_{\mathcal{M}})_u(v) = \int_{z \in u} \alpha_z(v(z))\mathrm{d}z,$$

where $u \in \mathcal{M}$, $v \in T_u\mathcal{M} = \ker \mathbf{D}_u$ for $\mathbf{D}_u$ the linearized CR-operator of u, and $\mathrm{d}z = \mathrm{d}\alpha|_u$ is an area form along u. The contact structure $\xi_{\mathcal{M}} = \ker \alpha_{\mathcal{M}}$ and the 1-dimensional distribution $\ker \mathrm{d}\alpha_{\mathcal{M}}$ can then be thought of as the average of the contact planes ξ_z, and respectively of $\ker d\alpha_z$, for $z \in u$, that is

$$\xi_{\mathcal{M}} = \int_{z\in u} \pi_*(\xi_z)\mathrm{d}z,$$
$$\ker d\alpha_{\mathcal{M}} = \int_{z\in u} \pi_*(\ker \mathrm{d}\alpha_z)\mathrm{d}z,$$

where $\pi : M\backslash L \to S^3$ is the quotient map to the leaf space. This means that the Reeb vector field $R_{\mathcal{M}}$ of $\alpha_{\mathcal{M}}$ spans the average direction in the "shadowing cone" $C_\alpha = \pi_*(\ker \mathrm{d}\alpha) \subset TS^3$.

The holomorphic shadow We define the *holomorphic shadow* of the Reeb dynamics of α_M on M to be the Reeb dynamics of the associated contact form $\alpha_{\mathcal{M}}$ on S^3, provided by Theorem R. The flow of $\alpha_{\mathcal{M}}$ can be viewed as a flow $\phi_t^{M;\mathcal{M}}$ on $M\backslash L$ which leaves the holomorphic foliation $\mathcal{M}$ invariant (i.e., it maps holomorphic curves to holomorphic curves). It is the "best approximation" of the Reeb flow of α_M with this property, as its generating vector field is obtained by reparametrizing the projection of the original Reeb vector field to the tangent space of $\mathcal{M}$, via a suitable L^2-orthogonal projection. Concretely, we have

$$R_{\mathcal{M}}(u) = \frac{P_u(R_\alpha|_u)}{(\alpha_{\mathcal{M}})_u(P_u(R_\alpha|_u))} \in T_u\mathcal{M},$$

where $P_u : W^{1,2}(N_u) \to \ker \mathbf{D}_u$ denotes the L^2-orthogonal projection with respect to the metric

$$g_u(v,w) = \int_{z\in u} g_z(v(z), w(z))\mathrm{d}z,$$

with $g_z = \mathrm{d}\alpha_z(\cdot, J\cdot) + \alpha_z \otimes \alpha_z + \mathrm{d}t \otimes \mathrm{d}t$, and $v, w \in W^{1,2}(N_u)$ sections of the normal bundle N_u to u. It may also be viewed as a Reeb flow $\phi_t^{S^3;\mathcal{M}}$ on S^3, related to the one on M via a semi-conjugation

$$\begin{array}{ccc} M\backslash L & \xrightarrow{\phi_t^{M;\mathcal{M}}} & M\backslash L \\ \downarrow{\scriptstyle\pi} & & \downarrow{\scriptstyle\pi} \\ S^3 & \xrightarrow{\phi_t^{S^3;\mathcal{M}}} & S^3 \end{array}$$

where π is the projection to the leaf space $\mathcal{M} \cong S^3$. We will now focus on the global properties of the correspondence $\alpha_M \mapsto \alpha_{\mathcal{M}}$.

For F a genus zero surface, let $\mathbf{Reeb}(F, \phi_F)$ denote the collection of contact forms whose flow is adapted to some concrete planar open book $\pi_B : B\backslash L \to S^1$ on a given 3-manifold B, of abstract form $B = \mathbf{OB}(F, \phi_F)$. Iteratively, we define $\mathbf{Reeb}(\mathbf{LF}(F, \phi_F), \phi)$ to be the collection of contact forms with flow adapted to some concrete IP open book $\pi_M : M\backslash B \to S^1$ on a 5-manifold M, of abstract form $M = \mathbf{OB}(\mathbf{LF}(F, \phi_F), \phi)$, whose restriction to the binding $B = \mathbf{OB}(F, \phi_F)$ belongs to $\mathbf{Reeb}(F, \phi_F)$. We call elements in $\mathbf{Reeb}(\mathbf{LF}(F, \phi_F), \phi)$ *IP contact forms*, or *IP Giroux forms.*

We then have a map

$$\mathbf{HS}:\mathbf{Reeb}(\mathbf{LF}(F,\phi_F),\phi)\rightarrow\mathbf{Reeb}(\mathbb{D}^2,\mathbb{1}),$$

given by taking the holomorphic shadow with respect to an auxiliary almost complex structure J associated with α_M. We refer to $\mathbf{HS}^{-1}(\alpha_{\mathrm{std}})$ as the *integrable fiber*, where α_{std} denotes the standard contact form in S^3.

Theorem S. ([100] Reeb flow lifting theorem) **HS** *is surjective.*

In other words, for some J, we may lift any Reeb flow on S^3 adapted to the trivial open book, as the holomorphic shadow of the Reeb flow of an IP Giroux form adapted to *any* choice of concrete IP contact 5-fold. The map **HS** is clearly not in general injective, as it forgets dynamical information in the fibers. While the above lifting procedure is not precisely an extension of the flow, the above theorem says that Reeb dynamics on an IP contact 5-fold is at least as complex as Reeb dynamics on the standard contact 3-sphere. Recalling that the Levi–Civita regularization of the planar restricted three-body problem (for subcritical energy) gives a Reeb flow on S^3; this gives a concrete "measure" of the complexity of the spatial three-body problem. Namely, *the* **spatial** *three-body problem is dynamically at least as complex as the* **planar** *three-body problem.*

Somewhat related, we point out that higher dimensional Reeb flows encode the complexity of all flows on arbitrary compact manifolds (i.e., they are *universal*) [24].

Dynamical applications We wish to apply the above results to the SCR3BP (cf. Fig. 20). We first introduce the following general notion. Consider an IP 5-fold M with an IP Reeb dynamics, endowed with an IP holomorphic foliation $\mathcal{M}$ as in Theorem Q. Fix a page P in the IP open book of M, and consider the associated Poincaré return map $f:\mathrm{int}(P)\rightarrow\mathrm{int}(P)$. A (spatial) point $x\in\mathrm{int}(P)$ is said to be *leaf-wise* (or *fiber-wise*) *k-recurrent* with respect to $\mathcal{M}$ if $f^k(x)\in\mathcal{M}_x$, where $\mathcal{M}_x$ is the leaf of $\mathcal{M}$ containing x, and $k\geq 1$. This means that $f^k(\mathrm{int}(\mathcal{M}_x))\cap\mathrm{int}(\mathcal{M}_x)\neq\emptyset$. This is, roughly speaking, a symplectic version of the notion of *leaf-wise intersection* introduced by Moser [101] for the case of the isotropic foliation of a coisotropic submanifold.

In the integrable case of the rotating Kepler problem, where the mass ratio $\mu=0$, the holomorphic foliation provided by Theorem Q can be obtained directly; cf. Proposition 7.8. Denote this "integrable" holomorphic foliation on S^*S^3 by $\mathcal{M}_{int}$. Since the return map for $\mu=0$ preserves fibers, every point is leaf-wise 1-recurrent with respect to $\mathcal{M}_{int}$. If the mass ratio is sufficiently small, then the leaves of $\mathcal{M}_{int}$ will still be symplectic with respect to $d\alpha$, where α is the corresponding perturbed contact form on the unit cotangent bundle S^*S^3.

We have the following perturbative result:

Theorem T. ([100]) *In the SCR3BP, for any choice of page P in the open book of Theorem M, for any fixed choice of $k\geq 1$, for sufficiently small μ (depending on k), for energy c below the first critical value $H(L_1(\mu))$, along the bounded components of the Hill region, and for every $l\leq k$, there exist*

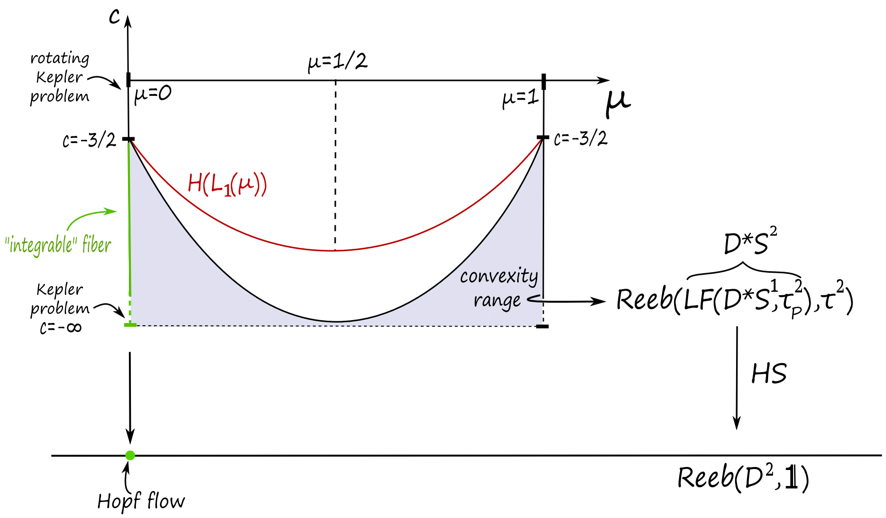


Figure 20. An abstract sketch of the convexity range in the SCR3BP (shaded), for which the holomorphic shadow is well defined. We should disclaim that the above is not a plot; the convexity range is not yet fully understood, although it contains (perhaps strictly) a region which qualitatively looks like the above, cf. [5,6]

infinitely many points in $int(P)$ *which are leaf-wise* l*-recurrent with respect to* $\mathcal{M}_{int}$.

In simpler words, the spatial three-body problem admits an abundance of leaf-wise recurrent points, at least in the perturbative regime.

Remark 7.9. The same conclusion holds for arbitrary $\mu \in [0,1]$, but sufficiently negative $c \ll 0$ (depending on μ and k).

In fact, the conclusion of the Theorem T holds whenever the relevant return map is sufficiently close to a return map which preserves the leaves of the holomorphic foliation of Theorem Q (i.e., which coincides with its holomorphic shadow on M). It may then be interpreted as a symplectic version of the main theorem in [101], for two-dimensional symplectic leaves.

8. Conclusion and further discussion

In the above account, we have tried to paint a picture of the relevance of the three-body problem in the modern mathematical discourse, in the hope to convince the reader of the richness of material that has ensued from this concrete problem alone. It has been more than a 100 years since Poincaré's work, and this problem is still a benchmark for modern developments.

Concerning the spatial problem, several of its aspects remain vastly unexplored and poorly understood. We have chosen to focus on the search of closed orbits as a starting point, for historical and heuristic reasons, as well as the fact that we have available techniques in the form of Floer theory. However, even this part of the story of is far from over, although we seem to be closing in. On the one hand, Theorem M provides the underlying geometric structure, and Theorem Q goes further and provides an adapted foliation which is compatible with the dynamics, and which is intimately related to the dynamics of the integrable limit case, as stated in Proposition 7.8. The general guiding question is how we use these underlying structures to extract dynamical applications. Moreover, one can also write down global hypersurfaces of section explicitly (see Theorem A in [98]), and this allows to use numerical methods in a hands-on way, which will certainly shed light on the problem. We will pursue this in further work.

Inspired by the Poincaré two-step approach, we have obtained a very general fixed-point theorem in the form of Theorem O. One may attempt to generalize it in several directions, although at this point, it is perhaps worth it to do so once one knows it applies to the problem by which it was inspired. So far, the Hamiltonian twist condition, while simple to state and rather appealing (specially from the perspective of Floer theory), seems hard to check in practice and rather restrictive.

The alternative holomorphic approach that we discussed above is also very appealing from a theoretical perspective, since in principle it allows to relate a dynamical system on a 5-fold which we wish to understand, to a dynamical system on the 3-fold S^3 of a type which has been studied much more extensively. The hope is to "lift" knowledge from the holomorphic shadow to the original dynamics (entropy, invariant subsets, invariant measures...). The main difficulty is that the shadow alters the dynamics, perhaps significantly, as it involves projecting the vector field to the tangent space of the moduli space. It is the "best approximation" of the original flow with the property that it maps a holomorphic annulus to another holomorphic annulus. It also has the disadvantage that it forgets dynamical information in the vertical directions, i.e., those tangent to the annuli, as well as most of the interesting dynamical information at the binding B (it is adapted to study spatial problems rather than planar ones). Observe that, in dimension 3, the shadow, when seen as a flow on B, is just a reparametrization of the original one. How much control we may obtain on the difference between the flow and its shadow, is unclear at the moment. More importantly, the relationship between closed orbits of the two flows is also not apparent.

On the other hand, one can follow an orbit and keep track of all the holomorphic annuli it intersects; this gives a path in S^3 which is tranverse to the contact structure and all the pages, and is in fact an orbit of what we called the shadowing cone. We call the collection of all such paths the *transverse* shadow. While no longer a flow, it remembers the original dynamics in a much more reliable way. In [100], we have used this idea (in combination with Brouwer's translation theorem) to extract Theorem T above, and perhaps may be exploited further.

Acknowledgements

The author wishes to thank all of those present (virtually and physically) during the lectures: Helmut Hofer (who also provided more accurate historical background), Chris Wendl (who also provided useful background, and corrected my pronunciation on Eastern-European names), Otto van Koert (who also provided amazing pictures and videos, and stayed up until 2am to watch the lectures), Urs Frauenfelder, Ezequiel Maderna, Paolo Ghiggini (*Povero Paolo*!), Alex Takeda, Jagna Wiśniewska, Fabio Gironella . . . to name a few whom I remember seeing on the screen; sorry if I missed you, and thank you. I am also very grateful to my uruguayan colleagues Alejandro Passeggi and Rafael Potrie for helping me organize this, to Gabriele Benedetti for pointing out a mathematical flaw in the work of the author with Otto van Koert, and to all the students in Uruguay, Sweden, and abroad (e.g., Turkey, France, US, South Korea, . . .) who showed interest in the project. This material is based on work supported by the Swedish Research Council under Grant No. 2016-06596, while the author was in residence at Institut Mittag–Leffler in Djursholm, Sweden during the Winter Semester, 2020.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

[1] Abbondandolo, A., Schwarz, M.: On the Floer homology of cotangent bundles. Commun. Pure Appl. Math. **59**(2), 254–316 (2006)

[2] Abouzaid, M., Blumberg, A.J.: Arnold conjecture and Morava K-theory. arXiv:2103.01507

[3] Acu, B.: The Weinstein conjecture for iterated planar contact structures. arXiv:1710.07724

[4] Acu, B., Etnyre, J.B., Ozbagci, B.: Generalizations of planar contact manifolds to higher dimensions. arXiv:2006.02940

[5] Albers, P., Fish, J.W., Frauenfelder, U., Hofer, H., van Koert, O.: Global surfaces of section in the planar restricted 3-body problem. Arch. Ration. Mech. Anal. **204**(1), 273–284 (2012)

[6] Albers, P., Fish, J.W., Frauenfelder, U., van Koert, O.: The Conley–Zehnder indices of the rotating Kepler problem. Math. Proc. Camb. Philos. Soc. **154**(2), 243–260 (2013)

[7] Albers, P., Frauenfelder, U., van Koert, O., Paternain, G.P.: Contact geometry of the restricted three-body problem. Commun. Pure Appl. Math. **65**(2), 229–263 (2012)

[8] Albers, P., Geiges, H., Zehmisch, K.: Reeb dynamics inspired by Katok's example in Finsler geometry. Math. Ann. **370**(3–4), 1883–1907 (2018)

[9] Albers, P., Hofer, H.: On the Weinstein conjecture in higher dimensions. Comment. Math. Helv. **84**(2), 429–436 (2009)

[10] Arnold, V.: Sur une propriete topologique des applications globalement canoniques de la mecanique classique (French). C. R. Acad. Sci. Paris **261**, 3719–3722 (1965)

[11] Arnol'd, V.I.: Some remarks on symplectic monodromy of Milnor fibrations. The Floer memorial volume, Progr. Math., vol. 133, pp. 99–103. Birkhäuser, Basel (1995)

[12] Ballmann, W.: Der Satz von Lusternik und Schnirelmann. Bonn. Math. Schr. **102**, 1–25 (1978)

[13] Bangert, V.: On the existence of closed geodesics on two-spheres. Int. J. Math. **4**(1), 1–10 (1993)

[14] Barth, K., Geiges, H., Zehmisch, K.: The diffeomorphism type of symplectic fillings. J. Symplectic Geom. **17**(4), 929–971 (2019)

[15] Benedetti, G., Ritter, A.F.: Invariance of symplectic cohomology and twisted cotangent bundles over surfaces. Int. J. Math. **31**(9), 2050070 (2020) **(p. 56)**

[16] Birkhoff, G.D.: Proof of Poincaré's last geometric theorem. Trans. AMS **14**, 14–22 (1913)

[17] Bott, R.: On the iteration of closed geodesics and the sturm intersection theory. Commun. Pure Appl. Math. **9**(2), 171–206 (1956)

[18] Bourgeois, F., Oancea, A.: Symplectic homology, autonomous Hamiltonians, and Morse–Bott moduli spaces. Duke Math. J. **146**(1), 71–174 (2009)

[19] Bowden, J.: Exactly fillable contact structures without Stein fillings. Algebraic Geom. Topol. **12**(3), 1803–1810 (2012)

[20] Bowden, J., Crowley, D., Stipsicz, A.I.: The topology of Stein fillable manifolds in high dimensions I. Proc. Lond. Math. Soc. **109**(6), 1363–1401 (2014)

[21] Bowden, J., Gironella, F., Moreno, A.: Bourgeois contact structures: tightness, fillability and applications. arXiv:1908.05749

[22] Burns, K., Matveev, V.S.: Open problems and questions about geodesics. arXiv:1308.5417

[23] Cannas da Silva, A.: Lectures on symplectic geometry. Lecture Notes in Mathematics, vol. 1764, pp. xii+217. Springer, Berlin (2001)

[24] Cardona, R., Miranda, E., Peralta-Salas, D., Presas, F.: Universality of Euler flows and flexibility of Reeb embeddings. arXiv:1911.01963

[25] Chenciner, A.: Poincaré and the three-body problem. Henri Poincaré, 1912–2012, Prog. Math. Phys., vol. 67, pp. 51–149. Birkhäuser/Springer, Basel (2015)

[26] Cho, W., Jung, H., Kim, G.: The contact geometry of the spatial circular restricted 3-body problem (2018). arXiv:1810.05796

[27] Cieliebak, K., Eliashberg, Y.: From Stein to Weinstein and back. Symplectic geometry of affine complex manifolds. American Mathematical Society Colloquium Publications, vol. 59, pp. xii+364. American Mathematical Society, Providence (2012)

[28] Cieliebak, K., Oancea, A.: Symplectic homology and the Eilenberg-Steenrod axioms. Appendix written jointly with Peter Albers. Algebraic Geom. Topol. **18**(4), 1953–2130 (2018)

[29] Conley, C.C.: On some new long periodic solutions of the plane restricted three body problem. In: Internat. Sympos. Monlinear Differential Equations and Nonlinear Mechanics, pp. 86–90. Academic Press, New York (1963)

[30] Cristofaro-Gardiner, D., Hutchings, M., Pomerleano, D.: Torsion contact forms in three dimensions have two or infinitely many Reeb orbits. Geom. Topol. **23**(7), 3601–3645 (2019)

[31] Cristofaro-Gardiner, D., Hutchings, M., Ramos, V.G.B.: The asymptotics of ECH capacities. Invent. Math. **199**, 187–214 (2015)

[32] Croke, C.B.: Poincaré's problem and the length of the shortest closed geodesic on a convex hypersurface. J. Differ. Geom. **17**(4), 595–634 (1983)

[33] Donaldson, S.K., Thomas, C.B.: Geometry of low-dimensional manifolds. 2. Symplectic manifolds and Jones-Witten theory. In: Proceedings of the symposium held in Durham, July 1989. London Mathematical Society Lecture Note Series, vol. 151, pp. xiv+242. Cambridge University Press, Cambridge (1990)

[34] Donaldson, S.K.: Lefschetz pencils on symplectic manifolds. J. Differ. Geom. **53**(2), 205–236 (1999)

[35] Eliashberg, Y.: Unique holomorphically fillable contact structure on the 3-torus. Int. Math. Res. Not. **1996**(2), 77–82 (1996)

[36] Etnyre, J.B.: Lectures on open book decompositions and contact structures. Floer homology, gauge theory, and low-dimensional topology, Clay Math. Proc., vol. 5, pp. 103–141. Amer. Math. Soc., Providence (2006)

[37] Fet, A.I.: Variational problems on closed manifolds. Mat. Sbornik 30 (1952). English translation in Amer. Math. Society, Translation No. 90 (1953)

[38] Floer, A.: Proof of the Arnold conjecture for surfaces and generalizations to certain Kähler manifolds. Duke Math. J. **53**(1), 1–32 (1986)

[39] Floer, A.: Morse theory for Lagrangian intersections. J. Differ. Geom. **28**(3), 513–547 (1988)

[40] Floer, A.: The unregularized gradient flow of the symplectic action. Commun. Pure Appl. Math. **41**(6), 775–813 (1988)

[41] Floer, A.: Cup length estimates on Lagrangian intersections. Commun. Pure Appl. Math. **42**(4), 335–356 (1989)

[42] Floer, A.: Symplectic fixed points and holomorphic spheres. Commun. Math. Phys. **120**(4), 575–611 (1989)

[43] Floer, A.: Witten's complex and infinite-dimensional Morse theory. J. Differ. Geom. **30**, 207–221 (1989)

[44] Floer, A., Hofer, H.: Symplectic homology. I. Open sets in Cn. Math. Z. **215**(1), 37–88 (1994)

[45] Floer, A., Hofer, H., Viterbo, C.: The Weinstein conjecture in $P \times C^l$. Math. Z. **203**(3), 469–482 (1990)

[46] Franks, J.: Geodesics on S^2 and periodic points of annulus homeomorphisms. Invent. Math. **108**(2), 403–418 (1992)

[47] Frauenfelder, U., van Koert, O.: The restricted three-body problem and holomorphic curves. Pathways in Mathematics, pp. xi+374. Birkhäuser/Springer, Cham (2018)

[48] Fukaya, K., Ono, K.: Arnold conjecture and Gromov-Witten invariant. Topology **38**(5), 933–1048 (1999)

[49] Geiges, H.: An introduction to contact topology. Cambridge Studies in Advanced Mathematics, vol. 109, pp. xvi+440. Cambridge University Press, Cambridge (2008)

[50] Geiges, H.: A brief history of contact geometry and topology. Expos. Math. **19**(1), 25–53 (2001)

[51] Geiges, H., Kwon, M., Zehmisch, K.: Diffeomorphism type of symplectic fillings of unit cotangent bundles. arXiv:1909.13586

[52] Ghiggini, P.: Strongly fillable contact 3-manifolds without Stein fillings. Geom. Topol. **9**(3), 1677–1687 (2005)

[53] Ginzburg, V.L.: The Conley conjecture. Ann. Math. (2) **172**, 1127–1180 (2010)

[54] Ginzburg, V.L., Gürel, B.Z.: The Conley conjecture and beyond. Arnold Math. J. **1**(3), 299–337 (2015)

[55] Ginzburg, V.L., Gürel, B.Z.: Conley conjecture revisited. Int. Math. Res. Not. IMRN (3), pp. 761–798 (2019)

[56] Giroux, E.: Géométrie de contact: de la dimension trois vers les dimensions supérieures. (French) [Contact geometry: from dimension three to higher dimensions]. In: Proceedings of the International Congress of Mathematicians, Vol. II (Beijing, 2002), pp. 405–414. Higher Ed. Press, Beijing (2002)

[57] Griffiths, P., Harris, J.: Principles of algebraic geometry. Pure and Applied Mathematics, pp. xii+813. Wiley-Interscience [John Wiley & Sons], New York (1978)

[58] Gromoll, D., Meyer, W.: Periodic geodesics on compact riemannian manifolds. J. Differ. Geom. **3**, 493–510 (1969)

[59] Gromov, M.: Pseudoholomorphic curves in symplectic manifolds. Invent. Math. **82**, 307–347 (1985)

[60] Gromov, M.: Homotopical effects of dilatation. J. Differ. Geom. **13**(3), 303–310 (1978)

[61] Hein, D.: The Conley conjecture for the cotangent bundle. Arch. Math. (Basel) **96**(1), 85–100 (2011)

[62] Hill, G.: On the part of the Motion of the Lunar Perigee which is a Function of the Mean Motions of the Sun and the Moon. John Wilson & Son, Cambridge (1877), Reprinted in Acta 8, 1–36 (1886)

[63] Hill, G.: Researches in the lunar theory. Am. J. Math. **1**, 5–26, 129–147, 245–260 (1878)

[64] Hingston, N.: Equivariant Morse theory and closed geodesics. J. Differ. Geom. **19**(1), 85–116 (1984)

[65] Hingston, N.: On the growth of the number of closed geodesics on the two-sphere. Int. Math. Res. Not. **9**, 253–262 (1993)

[66] Hingston, N.: On the lengths of closed geodesics on a two-sphere. Proc. Am. Math. Soc. **125**(10), 3099–3106 (1997)

[67] Hitchin, N.J., Segal, G.B., Ward, R.S.: Integrabke systems, twistors, loop groups, and Riemann surfaces. Oxford Graduate Texts in Mathematics, vol. 4, p. 136. Clarendon Press, Oxford (1999)

[68] Hofer, H.: Symplectic capacities Geometry of low-dimensional manifolds, 2 (Durham, 1989), London Math. Soc. Lecture Note Ser., vol. 151, pp. 15–34. Cambridge Univ. Press, Cambridge (1990)

[69] Hofer, H.: Pseudoholomorphic curves in symplectizations with applications to the Weinstein conjecture in dimension three. Invent. Math. **114**(3), 515–563 (1993)

[70] Hofer, H., Salamon, D.A.: Floer homology and Novikov rings. The Floer memorial volume, Progr. Math., vol. 133, pp. 483–524. Birkhäuser, Basel (1995)

[71] Hofer, H., Viterbo, C.: The Weinstein conjecture in cotangent bundles and related results. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **15**(3), 411–445 (1989)

[72] Hofer, H., Viterbo, C.: The Weinstein conjecture in the presence of holomorphic spheres. Commun. Pure Appl. Math. **45**(5), 583–622 (1992)

[73] Hofer, H., Wysocki, K., Zehnder, E.: The dynamics on three-dimensional strictly convex energy surfaces. Ann. Math. (2) **148**(1), 197–289 (1998)

[74] Hryniewicz, U.: Fast finite-energy planes in symplectizations and applications. Trans. Am. Math. Soc. **364**, 1859–1931 (2012)

[75] Hryniewicz, U.: Systems of global surfaces of section for dynamically convex Reeb flows on the 3-sphere. J. Symplectic Geom. **12**(4), 791–862 (2014)

[76] Hryniewicz, P.S.: On the existence of disk-like global sections for Reeb flows on the tight 3-sphere. Duke Math. J. **160**(3), 415–465 (2011)

[77] Hryniewicz, U., Salomão, Pedro A.S., Wysocki, K.: Genus zero global surfaces of section for Reeb flows and a result of Birkhoff. arXiv:1912.01078

[78] Hryniewicz, U., Salomão, P.A.S.: Global surfaces of section for Reeb flows in dimension three and beyond. In: Proceedings of the International Congress of Mathematicians-Rio de Janeiro 2018. Vol. II. Invited Lectures, pp. 941–967, World Sci. Publ., Hackensack (2018)

[79] Huybrechts, D.: Complex geometry. An introduction. Universitext, pp. xii+309. Springer, Berlin (2005)

[80] Irie, K.: Dense existence of periodic Reeb orbits and ECH spectral invariants. J. Mod. Dyn. **9**, 357–363 (2015)

[81] Irie, K.: Equidistributed periodic orbits of C^∞-generic three-dimensional Reeb flows (2018). arXiv:1812.01869

[82] Katok, A.B.: Ergodic perturbations of degenerate integrable Hamiltonian systems (Russian). Izv. Akad. Nauk SSSR Ser. Mat. **37**, 539–576 (1973)

[83] Katok, A., Hasselblatt, B.: Introduction to the modern theory of dynamical systems. With a supplementary chapter by Katok and Leonardo Mendoza. Encyclopedia of Mathematics and its Applications, vol. 54, pp. xviii+802. Cambridge University Press, Cambridge (1995)

[84] Klingenberg, W.: Lectures on closed geodesics. Grundlehren der Mathematischen Wissenschaften, vol. 230. Springer, Berlin (1978)

[85] Klingenberg, W., Takens, F.: Generic properties of geodesic flows. Math. Ann. **197**, 323–334 (1972)

[86] Kwon, M., van Koert, O.: Brieskorn manifolds in contact topology. Bull. Lond. Math. Soc. **48**(2), 173–241 (2016)

[87] Li, T.-J., Mak, C.Y., Yasui, K.: Calabi-Yau caps, uniruled caps and symplectic fillings. Proc. Lond. Math. Soc. (3) **114**(1), 159–187 (2017)

[88] Li, Y., Ozbagci, B.: Fillings of unit cotangent bundles of nonorientable surfaces. Bull. Lond. Math. Soc. **50**(1), 7–16 (2018)

[89] Liu, G., Tian, G.: Floer homology and Arnold conjecture. J. Differ. Geom. **49**(1), 1–74 (1998)

[90] Lyusternik, L.A., Fet, A.I.: Variational problems on closed manifolds. Dokl. Akad. Nauk SSSR **81**, 17–18 (1951)

[91] Lusternik, L., Schnirelmann, L.: Existence de trois géodésiques fermées sur toute surface de genre 0. C. R. Acad Sci. Paris **188**, 269–271 (1929)

[92] Lusternik, L., Schnirelmann, L.: Sur le probléme de trois géodésiques fermées sur toute surface de genre 0. C. R. Acad. Sci. Paris **189**, 534–536 (1929)

[93] Massot, P., Niederkrüger, K., Wendl, C.: Weak and strong fillability of higher dimensional contact manifolds. Invent. Math. **192**(2), 287–373 (2013)

[94] McDuff, D.: Symplectic manifolds with contact type boundaries. Invent. Math. **103**(3), 651–671 (1991)

[95] McDuff, D., Salamon, D.: Introduction to symplectic topology. Oxford Graduate Texts in Mathematics, 3rd edn, pp. xi+623. Oxford University Press, Oxford (2017)

[96] McGehee, R.P.: Some homoclinic orbits for the restricted three-body problem. Thesis (Ph.D.). The University of Wisconsin-Madison, p. 63. ProQuest LLC, Ann Arbor (1969)

[97] Milnor, J.: Singular points of complex hypersurfaces. Annals of Mathematics Studies, No. 61, pp. iii+122. Princeton University Press, Princeton, University of Tokyo Press, Tokyo (1968)

[98] Moreno, A., van Koert, O.: Global hypersurfaces of section in the spatial restricted three-body problem. Nonlinearity. Preprint arXiv:2011.10386

[99] Moreno, A., van Koert, O.: A generalized Poincaré-Birkhoff theorem. J. Fixed Point Theory Appl. Preprint arXiv:2011.06562

[100] Moreno, A.: Holomorphic dynamics in the restricted three-body problem. Preprint arXiv:2011.06568

[101] Moser, J.: A fixed point theorem in symplectic geometry. Acta Math. **141**(1–2), 17–34 (1978)

[102] Moser, J.: Monotone twist mappings and the calculus of variations. Ergodic Theory Dyn. Syst. **6**(3), 401–413 (1986)

[103] Neumann, W.D.: Generalizations of the Poincaré Birkhoff fixed point theorem. Bull. Austral. Math. Soc. **17**(3), 375–389 (1977)

[104] Nicholls, R.: Second species orbits of negative action and contact forms in the circular restricted three-body problem. arXiv:2108.05741

[105] Oancea, A.: A survey of Floer homology for manifolds with contact type boundary or symplectic homology. Symplectic geometry and Floer homology. A survey of the Floer homology for manifolds with contact type boundary or symplectic homology, Ensaios Mat., vol. 7, pp. 51–91. Soc. Brasil. Mat., Rio de Janeiro (2004)

[106] Oancea, A.: Morse theory, closed geodesics, and the homology of free loop spaces. With an appendix by Umberto Hryniewicz. IRMA Lect. Math. Theor. Phys., vol. 24, Free loop spaces in geometry and topology, pp. 67–109, Eur. Math. Soc., Zürich (2015)

[107] Oba, T.: Lefschetz-Bott fibrations on line bundles over symplectic manifolds. arXiv:1904.00369

[108] Ono, K.: On the Arnold conjecture for weakly monotone symplectic manifolds. Invent. Math. **119**(3), 519–537 (1995)

[109] Poincaré, H.: Les Méthodes Nouvelles de la Mécanique Céleste, Tome I, Paris, Gauthier-Viltars (1892). Republished by Blanchard, Paris (1987)

[110] Poincaré, H.: Sur un théorème de géométrie. Rend. Circ. Mat. Palermo **33**, 375–407 (1912)

[111] Poincaré, H.: Sur les lignes geodesiques des surfaces convexes. Trans. Am. Math. Soc. **6**, 237–274 (1905)

[112] Rademacher, H.-B.: On the average indices of closed geodesics. J. Differ. Geom. **29**(1), 65–83 (1989)

[113] Robbin, J., Salamon, D.: The Maslov index for paths. Topology **32**(4), 827–844 (1993)

[114] Salamon, D.A., Weber, J.: Floer homology and the heat flow. Geom. Funct. Anal. **16**(5), 1050–1138 (2006)

[115] Seidel, P.: A long exact sequence for symplectic Floer cohomology. Topology **42**(5), 1003–1063 (2003)

[116] Siefring, R.: Intersection theory of punctured pseudoholomorphic curves. Geom. Topol. **15**(4), 2351–2457 (2011)

[117] Sivek, S., Van Horn-Morris, J.: Fillings of unit cotangent bundles. Math. Ann. **368**(3–4), 1063–1080 (2017)

[118] Sorrentino, A.: Lecture notes on Mather's theory for Lagrangian systems. http://pmu.uy/pmu16/pmu16-0169.pdf

[119] Taubes, C.H.: The Seiberg-Witten equations and the Weinstein conjecture. Geom. Topol. **11**, 2117–2202 (2007)

[120] Ustilovsky, I.: Contact homology and contact structures on S^{4m+1}. PhD thesis, Stanford University (1999)

[121] Vigué-Poirrier, M., Sullivan, D.: The homology theory of the closed geodesic problem. J. Differ. Geom. **11**(4), 633–644 (1976)

[122] Viterbo, C.: Functors and computations in Floer homology with applications. I. Geom. Funct. Anal. **9**(5), 985–1033 (1999)

[123] Viterbo, C.: Functors and Computations in Floer homology with Applications Part II. arXiv:1805.01316

[124] Viterbo, C.: A proof of Weinstein's conjecture in $\mathbb{R}^{2n}$. Ann. Inst. H. Poincaré Anal. Non Linéaire **4**(4), 337–356 (1987)

[125] Wayne, C.E.: An introduction to KAM theory, lecture notes. http://math.bu.edu/INDIVIDUAL/cew/preprints/introkam.pdf

[126] Wendl, C.: Strongly fillable contact manifolds and J-holomorphic foliations. Duke Math. J. **151**(3), 337–384 (2010)

[127] Wendl, C.: Holomorphic curves in low dimensions. From symplectic ruled surfaces to planar contact manifolds. Lecture Notes in Mathematics, vol. 2216, pp. xiii+292. Springer, Cham (2018)

[128] Zhou, Z.: $(\mathbb{R}P^{2n-1}, \xi_{std})$ is not exactly fillable for $n \neq 2k$. arXiv:2001.09718

[129] Ziller, W.: The free loop space of globally symmetric spaces. Invent. Math. **41**(1), 1–22 (1977)

[130] Ziller, W.: Geometry of the Katok examples. Ergodic Theory Dyn. Syst. **3**(1), 135–157 (1983)

Agustin Moreno
School of Mathematics
Institute for Advanced Study
Princeton NJ
USA
e-mail: agustin.moreno2191@gmail.com

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:32
https://doi.org/10.1007/s11784-022-00957-6
Published online April 8, 2022

Journal of Fixed Point Theory and Applications

A generalized Poincaré–Birkhoff theorem

Agustin Moreno and Otto van Koert

To H. Poincaré, who taught us much;
To A. Floer, who followed suit;
To C. Viterbo, now on his 60th birthday, who took the cue;
and to all those who stand on the Shoulders of Giants.

Abstract. We prove a generalization of the classical Poincaré–Birkhoff theorem for Liouville domains, in arbitrary even dimensions. This is inspired by the existence of global hypersurfaces of section for the spatial case of the restricted three-body problem (Moreno and van Koert in Global hypersurfaces of section in the spatial restricted three-body problem, 2020).

Mathematics Subject Classification. 37C25, 37E40, 58C30, 57R58.

Contents

This article is part of the topical collection "Symplectic geometry - A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

1. Introduction

1.1. Poincaré–Birkhoff theorem, and the planar restricted three-body problem

The problem of finding closed orbits in the planar case of the restricted three-body problem goes back to ground-breaking work in celestial mechanics of Poincaré [31,32], building on work of Hill on the lunar problem [22]. The basic scheme for his approach may be reduced to:

(1) Finding a *global surface of section* for the dynamics;
(2) Proving a fixed-point theorem for the resulting first return map.

This is the setting for the celebrated Poincaré–Birkhoff theorem, proposed and confirmed in special cases by Poincaré and later proved in full generality by Birkhoff in [8]. The statement can be summarized as: if $\tau : A \to A$ is an area-preserving homeomorphism of the annulus $A = [-1, 1] \times S^1$ that satisfies a *twist* condition at the boundary, then it admits infinitely many periodic points of arbitrary large period.

In [30], the authors proved the existence of S^1-families of global hypersurfaces of section for the *spatial* restricted three-body problem (in the low-energy range, i.e., below and slightly above the first critical value, and independent of mass ratio), fully and non-perturbatively generalizing step (1) in the above approach to the spatial situation. The relevant return map τ is a Hamiltonian diffeomorphism defined on the interior of the global hypersurface of section, which is symplectomorphic to the interior of a Liouville domain $(\mathbb{D}^*S^2, \omega)$, where ω is deformation equivalent to the standard symplectic form. Furthermore, τ extends smoothly to the boundary of the global hypersurface of section, and gives rise to a homeomorphism of $(\mathbb{D}^*S^2, \omega)$ that is a Hamiltonian diffeomorphism on the interior. Drawing inspiration from this situation, in this paper, we propose a general fixed-point theorem for Liouville domains, as an attempt to address step (2) for the spatial case.

Fixed-point theory of Hamiltonian twist maps

The periodic points of τ are either boundary periodic points, which give planar orbits, or interior periodic points which are in 1:1 correspondence with spatial orbits. We are interested in finding *interior* periodic points.

1.2. The Hamiltonian twist condition

We propose a generalization of the twist condition introduced by Poincaré, for the Hamiltonian case and for arbitrary Liouville domains. Let $(W, \omega = \mathrm{d}\lambda)$ be a $2n$-dimensional Liouville domain, and consider a Hamiltonian symplectomorphism τ of W. Let $(B, \xi) = (\partial W, \ker \alpha)$ be the contact manifold at the boundary where $\alpha = \lambda|_B$, and R_α the Reeb vector field of α (uniquely determined via the equations $\mathrm{d}\alpha(R_\alpha, \cdot) = 0, \alpha(R_\alpha) = 1$). Recall that τ is Hamiltonian if $\tau = \phi_H^1$, where ϕ_H^t is the isotopy of W defined by $\phi_H^0 = id$, $\frac{\mathrm{d}}{\mathrm{d}t}\phi_H^t = X_{H_t} \circ \phi_H^t$, where we write $H_t = H(t, \cdot)$, and X_{H_t} is the Hamiltonian vector field of H_t defined via $i_{X_{H_t}}\omega = -dH_t$. The Liouville vector field V_λ is defined via $i_{V_\lambda}\omega = \lambda$.

Definition 1.1. (*Hamiltonian twist map*) We say that τ is a *Hamiltonian twist map* (with respect to α), if τ is generated by a *smooth* Hamiltonian $H : W \times \mathbb{R} \to \mathbb{R}$ which satisfies $X_{H_t}|_B = h_t R_\alpha$ for some *positive* and smooth function $h : B \times \mathbb{R} \to \mathbb{R}^+$.

Remark 1.2. For the purposes of this article, one may relax the smoothness assumption on H to C^2 regularity.

In particular, $H_t|_B \equiv const$ on B, and $\tau(B) \subset B$. We have $h_t = dH_t(V_\lambda)|_B$ is the derivative of H_t in the Liouville direction V_λ along B, which we assume strictly positive. Also, $\tau|_B$ is the time-1 map of a positive reparametrization of the Reeb flow on B. But note that, while the latter condition is only localized at B, the twist condition is of a *global* nature, as it requires global smoothness of the generating Hamiltonian (cf. [30, Remark 1.4]).

Here is a simple example illustrating why the smoothness of the Hamiltonian is relevant for the purposes of fixed points:

Example 1.3. (**Integrable twist maps**) Consider $M = S^n$, $n \geq 1$ with its round metric and its cotangent bundle $T^*M = \{(q,p) \in \mathbb{R}^{2n+2} : \langle q, p\rangle = 0,\ |q| = 1\}$. Let $H : T^*M \to \mathbb{R}$, $H(q,p) = 2\pi|p|$ (*not* smooth at the zero section); ϕ_H^1 extends to all of $\mathbb{D}^*M$ as the identity. It is a positive reparametrization of the Reeb flow at S^*M, generating a full turn of the geodesic flow, and all orbits are fixed points with fixed period. If we smoothen H near $|p| = 0$ to $K(q,p) = 2\pi g(|p|)$, with $g(0) = g'(0) = 0$, then $\tau = \phi_K^1 : \mathbb{D}^*M \to \mathbb{D}^*M$, $\tau(q,p) = \phi_H^{2\pi g'(|p|)}(q,p)$, is now a Hamiltonian twist map. If $g'(|p|) = l/k \in \mathbb{Q}$ with l, k coprime, then τ has a simple k-periodic orbit; therefore, τ has simple interior orbits of arbitrary large period (cf. [26, p. 350], [29], for the case $M = S^1$).

The Hamiltonian twist condition will be used to extend the Hamiltonian to a Hamiltonian that is admissible for computing symplectic homology. The extended Hamiltonian can have additional 1-periodic orbits and these, as well as 1-periodic orbits on the boundary, need be distinguished from the interior periodic points of τ. We impose the following conditions to do so.

1.3. Index growth

We consider a suitable index growth condition on the dynamics on the boundary, which is satisfied in the restricted three-body problem whenever the *planar* dynamics is strictly convex (see Theorem D.1). This assumption will allow us to separate boundary and extension orbits from interior ones via the index.

We call a strict contact manifold $(Y, \xi = \ker \alpha)$ *strongly index-definite* if the contact structure $(\xi, \mathrm{d}\alpha)$ admits a symplectic trivialization ϵ with the property that

- There are constants $c > 0$ and $d \in \mathbb{R}$, such that for every Reeb arc[1] $\gamma : [0, T] \to Y$ of Reeb action $T = \int_0^T \gamma^*\alpha$, we have
$$|\mu_{\mathrm{RS}}(\gamma; \epsilon)| \geq cT + d,$$
where μ_{RS} is the Robbin–Salamon index [33].

Index-positivity is defined similarly, where we drop the absolute value. A variation of this notion was explored in Ustilovsky's thesis [37]. He imposed the additional condition $\pi_1(Y) = 0$. With this extra assumption, the concept of index-positivity becomes independent of the choice of trivialization, although the exact constants c and d still depend on the trivialization ϵ. The global trivialization will be important when considering extensions of our Hamiltonians, as it will allow us to measure the index growth of potential new orbits.

1.4. Fixed-point theorems

We propose the following generalization of the Poincaré–Birkhoff theorem:

Theorem A. (Generalized Poincaré–Birkhoff theorem) *Suppose that τ is an exact symplectomorphism of a connected Liouville domain (W, λ), and let $\alpha = \lambda|_B$. Assume the following:*

- ***(Hamiltonian twist map)*** *τ is a Hamiltonian twist map, where the generating Hamiltonian is at least C^2. In addition, assume that all fixed points of τ are isolated;*
- ***(Index-definiteness)*** *If $\dim W \geq 4$, then assume $c_1(W)|_{\pi_2(W)} = 0$, and $(\partial W, \alpha)$ is strongly index-definite;*
- ***(Symplectic homology)*** *$SH_\bullet(W)$ is infinite dimensional.*

Then τ has simple interior periodic points of arbitrarily large (integer) period.

Remark 1.4. Let us discuss some aspects of the theorem:

[1] We will refer to the restriction of a Reeb orbit or Hamiltonian orbit to a finite interval as a Reeb arc or Hamiltonian arc.

(1) **(Grading)** We impose the assumptions $c_1(W)|_{\pi_2(W)} = 0$ (i.e. W is symplectic Calabi–Yau) to have a well-defined integer grading on symplectic homology.
(2) **(Surfaces)** If $\dim W = 2$, then the condition that $SH_\bullet(W)$ is infinite dimensional just means that W is not D^2 (see Appendix B); for D^2, we have $SH_\bullet(D^2) = 0$, and a rotation on D^2 gives an obvious counterexample to the conclusion. In the surface case, the argument simplifies, and one can simply work with homotopy classes of loops rather than the grading on symplectic homology. The Hamiltonian twist condition implies the classical twist condition for $W = \mathbb{D}^*S^1$, due to orientations.
(3) **(Cotangent bundles)** The symplectic homology of the cotangent bundle of a closed manifold with finite fundamental group is well known to be infinite dimensional, due to a result of Viterbo [38,39] (see also [3, 34]), combined, e.g., with a theorem of Gromov [19, Sect. 1.4]. We have $c_1(T^*M) = 0$ whenever M is orientable. As for the existence of a global trivialization of the contact structure $(\xi, d\lambda_{can})$, we note the following:
   - if Σ is an oriented surface, then $S^*\Sigma$ admits such a global symplectic trivialization;
   - if M^3 is an orientable 3-manifold, then S^*M^3 also admits such a global symplectic trivialization;
   - In addition, we know that symplectic trivializations of the contact structure on (S^*S^2, λ_{can}) are unique up to homotopy, since $[S^*S^2, Sp(2)] \cong H^1(S^*S^2; \mathbb{Z}) = 0$.
(4) **(Fixed points)** If fixed points are non-isolated, then we vacuously obtain infinitely many of them, although we cannot conclude that their periods are unbounded; "generically", one expects finitely many fixed points.
(5) **(Long orbits)** If W is a global hypersurface of section for some Reeb dynamics, with return map τ, interior periodic points with long (integer) period for τ translate into spatial Reeb orbits with long (real) period; see Lemma C.1.
(6) **(Katok examples)** There are well-known examples due to Katok [25] of Finsler metrics on spheres with only finitely many simple geodesics, which are arbitrarily close to the round metric (we review them in Appendix A.2); they admit global hypersurfaces of section with Hamiltonian return maps, for which the index-definiteness and the condition on symplectic homology hold. It follows that the return map does not satisfy the twist condition for any choice of Hamiltonians.
(7) **(Spatial restricted three-body problem)** From the above discussion and [30], we gather: the only standing obstruction for applying the above result to the spatial restricted three-body problem, in case where the planar problem is strictly convex, is the Hamiltonian twist condition. Here, note that symplectic homology is invariant under deformations of Liouville domains; see, e.g., [9] for a paper with detailed proofs. This would give a proof of existence of *spatial* long orbits in the spirit of Conley [13], which could in principle be collision orbits. Since the geodesic flow on S^2 arises as a limit case (i.e., the Kepler problem), it should be clear from the discussion on Katok examples that this is a subtle

condition. In [30], we have computed a generating Hamiltonian for the integrable case of the rotating Kepler problem; it does *not* satisfy the twist condition in the spatial case (in the planar case, a Hamiltonian twist map was essentially found by Poincaré). This does not mean a priori that there is not *another* generating Hamiltonian which does, but this seems rather unlikely.

As a particular case of Theorem A, we state the above result for star-shaped domains in cotangent bundles, as of independent interest (cf. [21]):

Theorem B. *Suppose that W is a fiber-wise star-shaped domain in the Liouville manifold (T^*M, λ_{can}), where M is simply connected, orientable and closed, and assume that $\tau : W \to W$ is a Hamiltonian twist map. If the Reeb flow on ∂W is strongly index-positive, and if all fixed points of τ are isolated, then τ has simple interior periodic points of arbitrarily large period.*

The above holds in particular for $M = S^2$, as explained in Remark 1.4 (3). One difference with [21] is that we work with compact domains in cotangent bundles and conclude that periodic points are interior, at the expense of imposing index-positivity.

1.5. Sketch of the proof

The proof is fairly simple: due to the twist condition, we can extend the map τ to a Hamiltonian diffeomorphism $\widehat{\tau}$ that is generated by a weakly admissible Hamiltonian (defined in Sect. 4). This allows us to appeal to symplectic homology. In particular, we will show $\varinjlim_k HF_\bullet(\widehat{\tau}^k) = SH_\bullet(W)$ (Lemma 4.1). Using an index filtration (via index-definiteness *and* the twist condition), we can show that all generators contributing to homology are actually fixed points of some τ^k, rather than fixed points of the extension. The crucial technical input is Lemma 4.5. If the minimal periods of periodic points of τ are bounded, then we can show using a spectral sequence, involving local Floer homology groups, that the rank of the resulting symplectic homology should also be bounded, leading to a contradiction. Alternatively, one could use the methods used for the proof of the Conley conjecture [17,21] to finish the proof.

1.6. Remarks on the twist condition and generalizations

If the Liouville domain is a surface, this definition of the Hamiltonian twist condition is not restrictive, and implements the idea sketched above in a simple way. In higher dimensions, the Hamiltonian twist condition is much more restrictive. Some examples illustrating the nature of the twist condition and applications of the above theorem will be presented in Appendix A. Given the above sketch of the proof, there is obviously some freedom in Definition 1.1 that allows the same methods to work. For example, if the vector field X_{H_t} is sufficiently C^1-close to a positive reparametrization of the Reeb vector field, then the methods will still go through. However, we will not pursue this generalization, because its depends on details that make the formulation awkward and difficult to check. We list some other generalizations, whose proofs will not be worked out in detail:

- (**Action positivity**) One can impose constraints on the functions h_t in the Hamiltonian twist condition that force the periodic orbits in the extension to have large action under iterates. In the setting of cotangent bundles, one can then use a theorem of Gromov [19, Sect. 1.4] cited below, to construct infinitely many interior periodic points.
- (**Isolated sets**) The assumption that the fixed points are isolated can be replaced by the weaker assumption that the fixed point set consists of a finite union of submanifolds. This is based on a slight generalization of local Floer homology, and is useful when studying integrable systems and their perturbations.
- (**Non-vanishing symplectic homology**) The condition $\dim SH_\bullet(W) = \infty$ can be replaced by the condition $SH_\bullet(W) \neq 0$. The key point here is that non-vanishing symplectic homology implies its unit is non-trivial. Then, the methods of the proof of the Conley conjecture [17,21] can be applied to conclude the existence of infinitely many simple periodic points. Strong index-definiteness is needed to show that these periodic points do not correspond to boundary and extension orbits, and so are interior.

Remark 1.5. Concerning the last generalization, we remark that we do not know a single example of a Liouville domain (W, λ) with $c_1(W) = 0$, $SH_\bullet(W) \neq 0$, and $\dim SH_\bullet(W) < \infty$.

2. Motivation and background

Hypersurfaces of section, return maps, and open books

Definition 2.1. Suppose that Y is a compact, oriented, smooth manifold with a non-singular autonomous flow ϕ_t. We call an oriented, compact hypersurface Σ in Y a *global hypersurface of section* for ϕ_t if

- the set $\partial\Sigma$ is an invariant set for the flow ϕ_t (if non-empty);
- the flow ϕ_t is positively transverse to the interior of Σ;
- for all $x \in Y \backslash \partial\Sigma$ there are $t_+ > 0$ and $t_- < 0$, such that $\phi_{t_+}(x) \in \Sigma$ and $\phi_{t-}(x) \in \Sigma$.

Given a global hypersurface of section, we can define a return map τ as follows: for each $x \in \mathrm{int}(\Sigma)$, we choose a minimal $t_+(x) > 0$ as in the definition above. Then, we put $\tau(x) = \phi_{t_+(x)}(x)$. Periodic points of τ then correspond to closed orbits of ϕ_t. In general, there is no continuous extension to the boundary, although it is unique whenever exists. Although global hypersurfaces of section do not have good stability properties in higher dimensions, we found that they can be constructed in certain classes of Hamiltonian dynamical systems that admit an involution. This class includes the restricted three-body problem and several variations (e.g., suitable Stark–Zeeman systems [30]).

This notion is also closely related to the notion of an open book decomposition. This consists of a fiber bundle $\pi : Y \backslash B \to S^1$, where $B \subset Y$ is a codimension 2 submanifold with trivial normal bundle (called the *binding*), such that π coincides with the angular coordinate along some choice of collar neighborhood $B \times \mathbb{D}^2$ of B. The *pages* of the open book are the closure of the fibers of π, all having B as boundary. Whenever ϕ_t is a Reeb dynamics of a contact form α on Y which is adapted to the open book (i.e., $\alpha|_B$ is also contact, and $d\alpha$ is symplectic on the pages), each page is a global hypersurface of section, and the return map preserves the symplectic form $d\alpha$. This is precisely the situation in [30].

In Appendix C, we will collect some standard facts which apply for return maps arising from Reeb dynamics, as described here, for which Theorem A may be applied.

3. Preliminaries on symplectic homology

3.1. Liouville domains and Hamiltonian dynamics

There are various forms of Hamiltonian Floer homology for Liouville domains: these are all referred to as *symplectic homology*. The first version was due to Floer-Hofer, [15]. See also Sect. 5 of [23] for an even earlier version, called symplectology. However, we will review the version due to Viterbo, [38,39]. Roughly speaking, this is a ring with unit that encodes both topological and dynamical data; it is the homology of a chain complex that is freely generated by 1-periodic Hamiltonian orbits.

We now fix conventions. Consider a *Liouville domain* (W, λ), i.e., $(W, d\lambda)$ is a compact symplectic manifold with boundary, and the vector field X defined by the equation $\iota_X d\lambda = \lambda$ is outward pointing along each boundary component of W. This vector field is the *Liouville vector field*. The 1-form λ is the *Liouville form*, and its restriction to ∂W, which we denote by α, is a contact form.

Given a Liouville domain (W, λ), we build its completion to a *Liouville manifold* by attaching a cylindrical end

$$(\widehat{W}, \widehat{\lambda}) := (W, \lambda) \cup_\partial ([1, \infty) \times \partial W, r\alpha).$$

Throughout the paper, we will consider smooth functions of the form $H : W \times S^1 \to \mathbb{R}$, a (time-dependent) *Hamiltonian* on W. Given such a Hamiltonian, we define its Hamiltonian vector field X_H via

$$\iota_{X_H} \mathrm{d}\lambda = -\mathrm{d}H.$$

We denote the set of 1-periodic orbits of X_H by $\mathcal{P}(H)$. For the purpose of Floer theory on non-compact manifolds, we will need a suitable class of Hamiltonians to work with. First, we recall the spectrum of a contact form α. If $\mathcal{P}(\alpha)$ denotes the set of all periodic Reeb orbits (including covers and without period bound), then

$$\mathrm{spec}(\alpha) = \{a \in \mathbb{R} \mid \text{there is } \gamma \in \mathcal{P}(\alpha), \text{ such that } a = \mathcal{A}(\gamma)\},$$

where the *action* is defined as $\mathcal{A}(\gamma) = \int_\gamma \alpha$.

Definition 3.1. We recall some standard terminology.

- A 1-periodic orbit $\gamma \in \mathcal{P}(H)$ is *non-degenerate* if $dFl_1^{X_H}(\gamma(0)) - \mathrm{id}$ invertible.

- The Hamiltonian H is *non-degenerate* if all $\gamma \in \mathcal{P}(H)$ are non-degenerate.
- A Hamiltonian H on $\widehat{W}$ is *linear at infinity* if at the cylindrical end H has the form $H(r,b,t) = cr + d$ for some constants $c > 0$ and d. In this case, we write $\operatorname{slope}(H) := c$.
- A Hamiltonian H that is non-degenerate and linear at infinity with $\operatorname{slope}(H) \notin \operatorname{spec}(\alpha)$ will be called *admissible*.

We call an S^1-family of almost complex structures $J = J_t$ on a Liouville manifold $(\widehat{W}, \widehat{\lambda})$ *SFT-like* if

- it is compatible with $(T\widehat{W}, d\widehat{\lambda})$; and
- on the cylindrical end it satisfies $\mathcal{L}_{r\partial r} J = 0$, $J\xi = \xi$, and $Jr\partial_r = R_\alpha$.

We denote by $\mathcal{J}$ the space of such families of almost complex structures.

3.2. Conley–Zehnder index, Robbin–Salamon index, and mean index

We will also need invariants of Hamiltonian orbits, i.e., the Conley–Zehnder index, or more generally, the Robbin–Salamon index, and the mean index. Assume that $x : \mathbb{R} \to \widehat{W}$ is an orbit of X_H. Take a symplectic trivialization $\epsilon : \mathbb{R} \times \mathbb{R}^{2n} \to x^*T\widehat{W}$, $(t, v) \mapsto \epsilon_t(v) \in T_{x(t)}\widehat{W}$. Then, we get a path of symplectic matrices associated with x, namely, $\psi_t = \epsilon_t^{-1} \circ dFl_t^{X_H} \circ \epsilon_0$. We can then define the Robbin–Salamon index of x as $\mu_{\mathrm{RS}}(x|_{[0,T]}, \epsilon) := \mu_{\mathrm{RS}}(\psi|_{[0,T]})$. If $\psi_T - \mathrm{id}$ is invertible, then the Robbin–Salamon index reduces to the Conley-Zehnder index. The case of Reeb flows is done similarly; we simply restrict the linearized Reeb flow to the symplectic vector bundle $(\xi, d\alpha)$. Similarly, we define the mean index of a 1-periodic orbit x as $\Delta(x, \epsilon) := \Delta(\psi)$, where $\Delta(\psi)$ is the mean index of the symplectic path ψ.[2] We have the following properties (see, e.g., Sect. 3.1.1 of [18]):

$$
\begin{aligned}
&(1)\quad |\mu_{\mathrm{RS}}(x|_{[0,T]}, \epsilon) - \Delta(x, \epsilon)| \le \tfrac{\dim W}{2}, \text{ for all } T;\\
&(2)\quad \lim_{T \to +\infty} \frac{\mu_{\mathrm{RS}}(\psi|_{[0,T]}, \epsilon)}{T} = \Delta(x, \epsilon);\\
&(3)\quad \Delta(x^{(k)}, \epsilon) = k\Delta(x, \epsilon),
\end{aligned}
$$

where we interpret the k-fold catenation $x^{(k)}$, a k-periodic orbit of H, as a 1-periodic orbit of the iterated Hamiltonian $H^{\# k}$.

Definition 3.2. We will call a Hamiltonian flow on W *strongly index-definite* if there is a symplectic trivialization $\epsilon_W : W \times \mathbb{R}^{2n} \to TW$, and constants $c > 0, d$, such that for every orbit of X_H, we have

$$|\mu_{\mathrm{RS}}(x|_{[0,T]}, \epsilon)| \ge cT + d.$$

The notion of *strong index-positivity* is obtained by dropping the absolute value in the above definition, and similarly for *strong index-negativity.* As in Introduction, we can also define it for Reeb flows. Here are some examples:

Lemma 3.3. *Suppose that (M, g) is a closed Riemannian manifold with positive sectional curvature. Assume in addition that the contact structure $(S^*M, (\xi, d\alpha))$ admits a global symplectic trivialization. Then, $(S^*M, d\alpha)$ is strongly index-positive.*

---

[2] A description of the mean index can be found on page 1318 of of [35].

Other examples are complements of Donaldson hypersurfaces in monotone symplectic manifolds provided that the degree is sufficiently high and symplectically trivial: these manifolds are index negative.

3.3. Hamiltonian Floer homology and symplectic homology

Given *Floer data* (J, H) of an SFT-like J and an admissible H, we note the following:

- There are no 1-periodic orbits of X_H on the cylindrical end, because of the spectrum assumption.
- Non-degenerate 1-periodic orbits of X_H are isolated.

Then $\mathcal{P}(H)$ consists of finitely many 1-periodic orbits. Informally speaking, we think of Floer homology as "Morse homology" of the following action functional:

$$\mathcal{A}_H : W^{1,2}(S^1 = \mathbb{R}/\mathbb{Z}, \widehat{W}) \longrightarrow \mathbb{R}, \quad \gamma \longmapsto \int_{S^1} \gamma^* \widehat{\lambda} - \int_0^1 H(\gamma(t), t)\mathrm{d}t.$$

This functional has the property $\mathcal{A}_{H^{\#k}}(x^{(k)}) = k\mathcal{A}_H(x)$ for iterates. A computation shows that $\operatorname{crit} \mathcal{A}_H = \mathcal{P}(H)$, and we define the Floer chain complex as

$$CF_\bullet(\widehat{W}, \widehat{\lambda}, H, J) := \bigoplus_{\gamma \in \mathcal{P}(H)} \mathbb{Z}_2\langle \gamma \rangle.$$

We grade this chain complex by the Conley–Zehnder index, so $\deg \gamma := \mu_{CZ}(\gamma, \epsilon)$. We make a couple of comments:

- in the standard procedure, we choose a capping disk $\tilde{\gamma}$ of a contractible 1-periodic orbit γ, and a symplectic trivialization $\epsilon_{\tilde{\gamma}}$ of $\tilde{\gamma}^* T\widehat{W}$. This gives a trivialization $e_t : (\mathbb{R}^{2n}, \omega_0) \to T_{\gamma(t)}\widehat{W}$ by restriction. We then define the Conley–Zehnder index of an orbit as in Sect. 3.2. Once the capping disk $\tilde{\gamma}$ is fixed, this index is independent of the choice of trivialization on a *fixed* capping disk. The index does depend on the choice of capping disk but not if $c_1(W)|_{\pi_2(W)} = 0$.
- for non-contractible orbits, one needs to choose a reference loop c and a reference symplectic trivialization ϵ_c for each free homotopy $[c] \in \tilde{\pi}_1(W)$. Given a 1-periodic orbit x in the same free homotopy class as c we choose a connecting cylinder S; the trivialization extends over S, and we can then define the Conley–Zehnder index as before.
- we choose the simpler, but more restrictive approach to use a global symplectic trivialization on some subdomain $\widetilde{W}$. The existence of such a trivialization implies that $c_1(T\widetilde{W}) = 0$. This approach obviously reduces to the previous approach provided that capping disks or connecting cylinders can be chosen to lie in the domain of definition of the global symplectic trivialization.

If we define an L^2-metric on $W^{1,2}(S^1, x^* T\widehat{W})$ by

$$\langle X, Y \rangle = \int_0^1 \omega(X(t), J_t(x(t))Y(t))\mathrm{d}t,$$

then the Floer equation is the L^2-gradient "flow"[3] of the above functional: for a cylinder $u : Z = \mathbb{R} \times S^1 \to \widehat{W}$, this is

$$(\mathrm{d}u - X_H \otimes \mathrm{d}t)^{0,1} = 0, \quad \lim_{s \to \pm\infty} u(s,t) = x_\pm(t). \tag{3.1}$$

Solutions to this equation are called *Floer trajectories.* Given 1-periodic orbits $x_+, x_- \in \mathcal{P}(H)$, the moduli space of Floer trajectories is

$$\mathcal{M}(x_+, x_-) := \{u : Z \to \widehat{W} \mid u \text{ satisfies } (3.1)\}.$$

In general, this space does not need to have a manifold structure. To obtain this extra structure, we first interpret Eq. (3.1) as a section of a vector bundle, via

$$\begin{aligned} \bar{\partial}_F : \mathcal{P}(x_+, x_-) &\longrightarrow \mathcal{E}(x_+, x_-), \quad u \longmapsto (\mathrm{d}u - X_H \otimes \mathrm{d}t)^{0,1} \\ &\in L^p(Z, \Omega^{0,1}(u^* T\widehat{W})\,). \end{aligned}$$

Here, $\mathcal{P}(x_+, x_-)$ is a Banach manifold of cylinders of class $W^{1,p}$ that are $W^{1,p}$-pushoffs of smooth cylinders that exponentially converge to the given asymptotes x_+ and x_-, and $\mathcal{E}(x_+, x_-)$ is a Banach bundle over $\mathcal{P}(x_+, x_-)$ whose fiber over $u \in \mathcal{P}(x_+, x_-)$ is $L^p(Z, \Omega^{0,1}(u^* T\widehat{W})\,)$. For details, see Chapter 8 in [7]. We will denote the linearization of $\bar{\partial}_F$ at $u \in \mathcal{P}(x_+, x_-)$ by $D_u \bar{\partial}_F$.

Proposition 3.4. *For Floer data (J, H) and $u \in \mathcal{M}(x_+, x_-)$, $D_u \bar{\partial}_F$ is a Fredholm operator of index*

$$\operatorname{ind} D_u \bar{\partial}_F = \mu_{CZ}(x_+, \epsilon) - \mu_{CZ}(x_-, \epsilon),$$

where ϵ is a symplectic trivialization of $u^* T\widehat{W}$.

In addition, we can always choose suitable Floer data close to initial Floer data such that all moduli spaces are transversely cut out:

Proposition 3.5. *There is a dense set $\mathcal{J}_{reg} \subset \mathcal{J}$ with the property for all $J \in \mathcal{J}_{reg}$, the linearized operator $D_u \bar{\partial}_F$ is surjective for all $u \in \mathcal{M}(x_+, x_-)$, and so $\mathcal{M}(x_+, x_-)$ is a smooth manifold of dimension $\mu_{CZ}(x_+, \epsilon) - \mu_{CZ}(x_-, \epsilon)$.*

Floer data (J, H) as in Proposition 3.5 will be called *regular Floer data.* We now have all the basic ingredients in place: choose regular Floer data (J, H), and define the boundary operator for the chain complex $CF_\bullet(\widehat{W}, \widehat{\lambda}, H, J)$ via

$$\partial x_+ = \sum_{x_- \in \mathcal{P}(H),\ \deg(x_-) = \deg(x_+) - 1} \#_{\mathbb{Z}_2} \left(\mathcal{M}(x_+, x_-)/\mathbb{R} \right) \cdot x_-.$$

Here, we have modded out $\mathcal{M}(x_+, x_-)$ by the reparametrization action in the domain, and the resulting quotient spaces can be compactified, so the coefficients in the above sum are actually finite.

Lemma 3.6. *This linear map is a differential: $\partial \circ \partial = 0$.*

The Floer homology of $(\widehat{W}, \widehat{\lambda}, J, H)$ is then defined as the homology

$$HF_\bullet(\widehat{W}, \widehat{\lambda}, J, H) := H_\bullet(CF_\bullet(\widehat{W}, \widehat{\lambda}, J, H), \partial).$$

[3] The flow is strictly speaking not defined, since it leads to an ill-posed initial value problem.

Remark 3.7. In the case of closed symplectic manifolds, Floer homology is independent of the choice of Floer data. This is not the case for Liouville domains, and this is the next topic we will deal with.

3.4. Continuation maps and symplectic homology

Assume that H_1 and H_2 are admissible Hamiltonians on a Liouville manifold $\widehat{W}$ with $\text{slope}(H_1) \leq \text{slope}(H_2)$. We interpolate between them via

$$K : \widehat{W} \times S^1 \times \mathbb{R} \longrightarrow \mathbb{R}, \quad (w,t,s) \longmapsto K_s(w,t),$$

where we impose the monotonicity condition $\partial_s K \leq 0$, [4] and

$$K_s(w,t) = \begin{cases} H_1(w,t), & \text{if } s \gg 0 \\ H_2(w,t), & \text{if } s \ll 0. \end{cases}$$

We then consider the parametrized Floer equation for $u : Z \to \widehat{W}$:

$$(\mathrm{d}u - X_K \otimes \mathrm{d}t)^{0,1} = 0, \quad \lim_{s\to\infty} u(s,t) = x_+(t) \in \mathcal{P}(H_1),$$
$$\lim_{s\to-\infty} u(s,t) = x_-(t) \in \mathcal{P}(H_2).$$

The results of the Fredholm theory mentioned in the previous section also apply in this setup, and we can define a *continuation map* as

$$c_{12} : CF_\bullet(\widehat{W}, \widehat{\lambda}, J, H_1) \longrightarrow CF_\bullet(\widehat{W}, \widehat{\lambda}, J, H_2),$$
$$x_+ \longmapsto \sum_{x_- \in \mathcal{P}(H_2),\ \deg(x_-)=\deg(x_+)} \#_{\mathbb{Z}_2}\mathcal{M}(x_+, x_-, J, K) \cdot x_-,$$

where $\mathcal{M}(x_+, x_-, J, K)$ is the moduli space of Floer trajectories of the parametrized Hamiltonian K.

Lemma 3.8. *The map c_{12} is a chain map, and the induced map on homology is independent of J, K.*

We also write c_{12} for the induced map on Floer homology

$$c_{12} : HF_\bullet(\widehat{W}, \widehat{\lambda}, J, H_1) \longrightarrow HF_\bullet(\widehat{W}, \widehat{\lambda}, J, H_2).$$

Symplectic homology is then defined as the direct limit over a direct system $\{H_i\}_i$ of admissible Hamiltonians for whose slopes $\text{slope}(H_i)$ increase to ∞

$$SH_\bullet(W, \lambda, J, \{H_i\}_i) := \varinjlim_{c_{ij},\ j>i} HF_\bullet(\widehat{W}, \widehat{\lambda}, J, H_i). \tag{3.2}$$

Remark 3.9. Symplectic homology is independent of J, and the sequence of Hamiltonians $\{H_i\}_i$. We will henceforth write $SH_\bullet(W,\lambda)$, or $SH_\bullet(W)$ (omitting the dependence on λ for notational simplicity), for symplectic homology. We similarly use the notation $CF_\bullet(H)$ when (W,λ) is fixed.

[4] This is the monotonicity condition for the continuation map with our conventions for the Floer equation. Our conventions agree with those in [14].

3.5. Degenerate Hamiltonians and local Floer homology

In case there is a 1-periodic orbit of H that is degenerate, we perturb H to a non-degenerate Hamiltonian $\widetilde{H}$ with the same slope as H, choose regular Floer data $(\widetilde{J}, \widetilde{H})$, and define

$$HF_\bullet(\widehat{W}, \widehat{\lambda}, H) := HF_\bullet(\widehat{W}, \widehat{\lambda}, \widetilde{J}, \widetilde{H}).$$

Lemma 3.10. *This is well defined, i.e., it is independent of the choice of perturbation, and of $\widetilde{J}$.*

Instead of choosing explicit perturbed Hamiltonians, we package them in local Floer homology, which we now review. Suppose H is a Hamiltonian and assume that $x \in \mathcal{P}(H)$ is isolated.[5] We need the following lemma, which we adapt from [12]:

Lemma 3.11. *Suppose that γ is an isolated 1-periodic orbit of X_H with an isolating neighborhood U. Then, for every neighborhood V of γ with $V \subset U$, there is a C^2-small perturbation $\widetilde{H}$ of H with the following properties:*

- *All 1-periodic orbits of $X_{\widetilde{H}}$ contained in U are already contained in V;*
- *For a compatible almost complex structure $\widetilde{J}$, all Floer trajectories contained in U are already contained in V.*

Take a C^2-small perturbation $\widetilde{H}$ as in the lemma, so that 1-periodic orbits in U are non-degenerate (via [35, Theorem 9.1]). As in [12], we define the local Floer homology $HF_\bullet^{loc}(\gamma, H)$ of γ as the homology of the complex $CF_\bullet^{loc}(U, \widetilde{H}, \widetilde{J})$ generated by 1-periodic orbits of $\widetilde{H}$, with differential counting Floer solutions lying in U. This is well defined and independent of the isolating neighborhood U, and the perturbed Floer data $(\widetilde{J}, \widetilde{H})$.

We have the following (see, e.g., formula (3.1) in [18]):

$$\mathrm{supp} HF_\bullet^{loc}(\gamma, H) \subset [\Delta(\gamma) - n, \Delta(\gamma) + n], \tag{3.3}$$

where $\mathrm{supp} HF_\bullet^{loc}(\gamma, H) = \{i : HF_i^{loc}(\gamma, H) \neq 0\}$, and $n = \frac{\dim(W)}{2}$.

Remark 3.12. We observe that the perturbation in Lemma 3.11 can chosen, such that the 1-periodic orbits of the perturbed Hamiltonian $\widetilde{H}$ have the same free homotopy class as γ.

3.6. Spectral sequence

Suppose now that H is a Hamiltonian that is linear at infinity with slope$(H) \notin$ spec(α). We assume furthermore that the 1-periodic orbits of H are all isolated. Hence, there are finitely many 1-periodic orbits with finite action spectrum $\mathcal{A}_H(\mathcal{P}(H))$. We order the action values in a strictly increasing sequence $\{a_i\}_{i=1}^k$. Choose a strictly increasing function $f : \mathbb{N}_0 \to \mathbb{R}$, such that $f(i) < a_{i+1} < f(i+1)$.

[5]In general, we can define local Floer homology for an isolated invariant set.

Proposition 3.13. *There is a spectral sequence converging to the Floer homology $HF_\bullet(W,\lambda,H)$, whose E^1-page is given by*

$$E^1_{pq}(H) = \bigoplus_{\substack{\gamma\in\mathcal{P}(H)\\ f(p-1)<\mathcal{A}_H(\gamma)<f(p)}} HF^{loc}_{p+q}(\gamma,H).$$

We will not give a detailed proof here, but refer to Appendix B of [27] for an almost identical setup. The spectral sequence is the spectral sequence associated with the action filtration given by f.

Remark 3.14. This description allows us to define Floer homology for Hamiltonians with isolated possibly *degenerate* orbits. In addition, we can directly use the free homotopy class of a degenerate periodic orbit, since sufficiently small perturbations cannot change this class as we already observed in Remark 3.12. This means that we can decompose Floer homology also in this degenerate setting into free homotopy classes.

A difficulty of this degenerate setup is that a single degenerate orbit can be responsible for several generators in Floer homology. Formula (3.3) can be used to retain some control.

This point of view is not new, and has been used extensively in for instance [17] and [21]. The spectral sequence, although not used in [17] and [21], just gives a convenient packaging, and only serves to make some arguments shorter. The idea of perturbation in Floer theory to get statements of degenerate orbits is even older, and was for example already used in [35].

3.7. Index-definiteness and grading

We shall need the following:

Lemma 3.15. *Suppose that $SH_\bullet(W,\lambda)$ is infinite dimensional, and assume that $\lambda|_{\partial W}$ is an index-definite contact form. Then, $\#\{i \mid SH_i(W,\lambda)\neq 0\} = \infty$.*

Proof. To prove this, choose a family $\{H_N\}_N$ of admissible Hamiltonians with increasing slopes, such that H_N is independent of N on W, and so that $CF_\bullet(H_N)$ injects into $CF_\bullet(H_M)$ for $M > N$. By non-degeneracy, each $CF_\bullet(H_N)$ is finitely generated, so the chain complexes get more generators with increasing N (since $\dim SH_\bullet(W,\lambda)=\infty$). By the index-definiteness assumption, these new generators have a degree whose absolute value is strictly increasing if N increases sufficiently. This settles the claim. □

4. Proof of the Generalized Poincaré–Birkhoff Theorem

Let (W,λ) be a Liouville domain with completion $(\widehat{W},\widehat{\lambda})$, r the coordinate in the cylindrical end, $B=\partial W$, $\alpha=\lambda|_B$, and τ a Hamiltonian twist map generated by $H=H_t$. The symplectic form on the cylindrical end is $d(r\alpha)$, so by the Hamiltonian twist condition, we get $h_t : B\to\mathbb{R}^+$, such that $X_{H_t}|_B = h_tR_\alpha$. This means that $H_t|_{r=1}\equiv C_t>0$, with $\partial_rH_t|_{r=1}=h_t$. The family of Hamiltonians H_t is not necessarily linear at infinity, and might hence be unsuitable to compute symplectic homology. To deal with this we will construct an extension $\widehat{H}$ to the cylindrical end of $\widehat{W}$ that is linear at infinity. By assumption, we have a time-dependent Hamiltonian H defined on W. In a collar neighborhood $\nu(B)$ of the boundary, we will write $H : (1-\epsilon, 1] \times B \times S^1 \to \mathbb{R}$, where r is the collar neighborhood parameter. We extend H to $\widehat{H}$ on $\widehat{W}$ using the following procedure. First of all, define $H_0(b,t) := H(r=1,b,t)$ and $H_1(b,t) := \frac{\partial H}{\partial r}|_{r=1,b,t}$. We put the remainder in the function $\frac{(r-1)^2}{2!}H_2$, so in short, we define

$$H_2 = (H - (H_0 + (r-1)H_1))\frac{2}{(r-1)^2}$$

on the collar neighborhood $\nu(B)$. By construction H_2 is a smooth function on a halfspace. The functions H_0 and H_1 are r-independent, so admit obvious extensions to $r>1$, but the function H_2 is r-dependent, so we will appeal to [36], which is based on reflection, to extend H_2 to $r>1$. We call this extension $\overline{H}_2$. Now, choose $\delta_1 > \delta_0 > 0$ and choose a decreasing cut-off function ρ with $\rho|_{[1,1+\delta_0]} = 1$ and $\rho(r) = 0$ for $r > 1+\delta_1$;

- put $\widehat{H}_2(r,b,t) = \overline{H}_2(r,b,t) \cdot \rho(r)$;
- put $\widehat{H}_0(r,b,t) = C \geq \max_t(C_t)$, $\widehat{H}_1(r,b,t) = A \geq \max_{t,b}(h_t(b))$ for $r \geq 1+\delta_1$;
- and put $\widehat{H}_j(r,b,t) = H_j(b,t)\cdot\rho(r) + (1-\rho(r))\widehat{H}_j(1+\delta_1,b,t)$, for $j=0,1$.

The extension is then defined as

$$\widehat{H} := \widehat{H}_0(r,b,t) + (r-1)\widehat{H}_1(r,b,t) + \frac{(r-1)^2}{2!}\widehat{H}_2(r,b,t). \tag{4.4}$$

By the above, we see that $H_0 = C_t$ and $H_1 = h_t$, so with our choices, we conclude that $\widehat{H} = A(r-1) + C$ for large r. The extension $\widehat{H}$ is therefore linear at infinity, and by perturbing A, we can assume that $A \notin \mathrm{spec}(\alpha)$. The same can be arranged for all iterates $\widehat{H}^{\#k}$ by possibly changing the slope A. The resulting Hamiltonians are then all linear at infinity, but they may have 1-periodic orbits that are degenerate. If all the degenerate 1-periodic orbits are isolated, then we can still define the Floer homology $HF_\bullet(\widehat{W}, \widehat{\lambda}, \widehat{H}^{\#k})$ using Remark 3.14. Let us call a Hamiltonian for which all 1-periodic orbits are isolated, and that is linear at infinity with slope not in the spectrum *weakly admissible.*

Since we will focus on return maps, it will be convenient to have some shorthand notation. Define $\widehat{\tau}^k := Fl_1^{X_{\widehat{H}^{\#k}}}$, and with the above remark in mind, we put $HF_\bullet(\widehat{\tau}^k) := HF_\bullet(\widehat{W}, \widehat{\lambda}, \widehat{H}^{\#k})$. We summarize this discussion in the following lemma:

Lemma 4.1. *The extended Hamiltonians $\widehat{H}^{\#k}$ are linear at infinity. Furthermore, if there is an increasing sequence $\{k_i\}_i \subset \mathbb{N}$, such that each Hamiltonian $\widehat{H}^{\#k_i}$ is weakly admissible, then we have the following isomorphism:*

$$SH_\bullet(W,\lambda) \cong \varinjlim_{k_i} HF_\bullet(\widehat{\tau}^{k_i}),$$

where $\widehat{\tau}^k = Fl_1^{X_{\widehat{H}^{\#k}}}$. □

For later purposes, we need the explicit form of $X_{\widehat{H}_t}$. This is given by

$$X_{\widehat{H}_t} = \left(\partial_r \widehat{H}_0 + \widehat{H}_1 + (r-1)\partial_r \widehat{H}_1 + (r-1)\widehat{H}_2 + \frac{(r-1)^2}{2}\partial_r \widehat{H}_2\right) R_\alpha$$
$$+ \frac{r-1}{r}\left(X^\xi_{\widehat{H}_1} + \frac{r-1}{2} X^\xi_{\widehat{H}_2} - \left(d\widehat{H}_1(R_\alpha) + \frac{r-1}{2} d\widehat{H}_2(R_\alpha)\right) Y\right). \tag{4.5}$$

Here, $Y = r\partial_r$ is the Liouville vector field, and $X^\xi_h \in \xi$ is the ξ-component of the contact Hamiltonian vector field $X_h = hR_\alpha + X^\xi_h$ of a Hamiltonian $h : B \to \mathbb{R}$, defined implicitly by the equation $d\alpha(X^\xi_h, \cdot) = -dh|_\xi$. Due to our choice of interpolation, the second term will be smaller in C^0-norm if we choose δ_1 smaller. We denote the coefficient of R_α by

$$F = \partial_r \widehat{H}_0 + \widehat{H}_1 + (r-1)\partial_r \widehat{H}_1 + (r-1)\widehat{H}_2 + \frac{(r-1)^2}{2}\partial_r \widehat{H}_2.$$

Lemma 4.2. *If δ_1 is chosen to be sufficiently small, then F is positive.*

Proof. To see this, we note that the first three terms are non-negative, and the second term is at least $\min_{t,b} h_t(b) > 0$. To see that the last two terms can be made sufficiently small, note that $\widehat{H}_2$ has a bound independent of δ_1, and $\partial_r \widehat{H}_2$ is bounded by C_2/δ_1, where C_2 is independent of δ_1. Because this term is multiplied by a factor $(r-1)^2$, which is bounded by δ_1^2, the claim follows. $\square$

As a result, we see that $X_{\widehat{H}}$ is mostly following the positive Reeb direction if we choose δ_1 sufficiently small. In the proof of Lemma 4.5, we will investigate the linearization of $X_{\widehat{H}}$, which ideally would require closeness to a reparametrized Reeb flow in C^1-norm rather than C^0-norm. However, C^1-closeness does not hold, but we will perform a finer analysis with additional assumptions, which will allow us to fix δ_1.

Lemma 4.1 allows us to compute symplectic homology with the extended Hamiltonian, but it does, by itself, not give any control over periodic orbits in the extension. To prove our main theorem, we want to show that all generators of $SH_\bullet(W, \lambda)$ represent periodic points of τ (i.e., lie in W). To do so, we need to show that the additional periodic points of $\widehat{\tau}$ do not contribute to the symplectic homology. Depending on the situation, we will use a filtration by homotopy classes or a filtration by index. More specifically, for $p \in \mathrm{Fix}(\widehat{\tau}^k)$, consider the loop $\gamma_p(t) = Fl_t^{X_{\widehat{H}_t}}(p)$. Then

- If $\dim W = 2$, the free homotopy class of γ_p in $\tilde{\pi}_1(W)$ can be used to see that the additional periodic orbits do not contribute homologically;
- If $\dim W > 2$, the CZ-index and the index-definiteness assumption will be used to arrive at the same conclusion.

4.1. Filtration by homotopy class

Assume $\dim W = 2$. Let $\mathrm{Fix}_\partial(\widehat{\tau}^k) := \mathrm{Fix}(\widehat{\tau}^k) \cap ([1, +\infty) \times B)$. Given $p \in \mathrm{Fix}_\partial(\widehat{\tau}^k)$, let $[\gamma_p]$ be the free homotopy class in $\tilde{\pi}_1(\partial_p W) \cong \mathbb{Z}$, where $\partial_p W$ is the connected component of ∂W containing p. We denote the absolute value of this integer by $|[\gamma_p]|$.

Lemma 4.3. *Assume the hypothesis of Theorem A, and that* $\dim W = 2$. *Then, there is* $A > 0$, *independent of* k, *such that for all* $p \in Fix_\partial(\widehat{\tau}^k)$, *we have* $|[\gamma_p]| \geq Ak$.

Proof. On each circle component of B, choose an angular coordinate ϕ, such that $R_\alpha = \partial_\phi$. From Eq. (4.5) and Lemma 4.2, we see that the component of $X_{\widehat{H}}$ in the ∂_ϕ-direction is bounded from below by some constant $A > 0$, e.g., $A = \inf_B F$. Iterating, we get a bound of the claimed form Ak. □

Corollary 4.4. *Suppose* W *and* τ *are as in the assumptions of Theorem A, with* $\dim W = 2$. *Then, Theorem A holds.*

Proof. To prove the statement, we will argue by contradiction, so we assume that the periods of τ are bounded: denote the minimal periods by $m_0 = 1 < m_1 < \ldots < m_M$; we include m_0 even if τ has no fixed points.

Fix a positive integer N and let A be as in Lemma 4.3. Let δ denote a free homotopy class in $\tilde{\pi}_1(W)$ that is represented by a simple Reeb orbit (a boundary parallel simple loop). For $i \in \{1, \ldots, N\}$ and the iterate $i\delta$, from Corollary B.3, we have $\operatorname{rk} SH^{i\delta}_\bullet(W) = 2$ (here, we use the notation from Appendix B).

We now use the assumption that all fixed points of τ are isolated, and choose $k > m_M$, such that k is not divisible by $m_1, \ldots, m_M$ (for example choose a large prime). This choice of k forces the 1-periodic orbits of $\widehat{H}^{\#k}$ to be isolated on the interior of W. By construction, the Hamiltonian $\widehat{H}^{\#k}$ is linear at infinity, so we find r_∞, such that $\widehat{H}^{\#k}$ is linear on $[r_\infty, \infty) \times B$. If there are non-isolated 1-periodic points of $\widehat{H}^{\#k}$ on the cylindrical part $[1, r_\infty] \times B$, then we use Lemma 4.6 below to perturb the Hamiltonian $\widehat{H}^{\#k}$; this perturbation makes all orbits on $[1, r_\infty] \times B$ non-degenerate, and does not affect 1-periodic orbits on the interior of W. Hence, we obtain a weakly admissible, possibly degenerate Hamiltonian, which we continue to denote by $\widehat{H}^{\#k}$. For this Hamiltonian, we can define Floer homology using Remark 3.14.

By choosing an increasing sequence $\{k\}$ of primes, we can then define $SH^{i\delta}_\bullet(W) = \varinjlim_k HF^{i\delta}_\bullet(\widehat{H}^{\#k})$. Hence, we find a sufficiently large k that

$$2 \leq \operatorname{rk} HF^{i\delta}_\bullet(\widehat{H}^{\#k}) = \sum_{p,q} E^\infty_{pq}(\widehat{H}^{\#k}) \leq \sum_{p,q} E^1_{pq}(\widehat{H}^{\#k}) = \sum_{p,q} \operatorname{rk} HF^{loc,i\delta}_{p+q}(\gamma, \widehat{H}^{\#k}).$$

All of these sums are finite by the assumption that the fixed points are isolated. We conclude that there is a 1-periodic orbit $\gamma_{k,i\delta}$ of $\widehat{H}^{\#k}$ whose free homotopy class equals $i\delta$. From Lemma 4.3, every $p \in \operatorname{Fix}_\partial(\widehat{\tau}^k)$ has $[\gamma_p] = j\delta$ with $j \geq Ak$. If we choose $k > N/A$, we see that $j > N$, so the 1-periodic orbit $\gamma_{k,i\delta}$ is represented by a fixed point of τ^k.

This works for all N, so by sending k to infinity, we get infinitely many periodic points of τ. To see that these are geometrically distinct, note that if $p \in \operatorname{Fix}(\tau^k)$ and $a := [\gamma_p] = i\delta$ is boundary parallel, then γ_p^ℓ has homotopy class ℓa, so another orbit must represent the free homotopy class a. Taking the limit in k, we see that new generators in the homotopy class a need to appear to generate $SH^a(W)$. This gives infinitely many geometrically distinct interior periodic points (in different boundary parallel homotopy classes). □

4.2. Filtration by index

We now deal with the second case, so we assume now that $\dim W > 2$, $c_1(W)|_{\pi_2(W)} = 0$, and that the Reeb flow is strongly index-definite. To set up the argument, we first need to establish that index-definiteness of the linearized Reeb flow equation at the boundary (in the sense of Definition E.1 in Appendix E) implies index-definiteness of the linearized Hamiltonian equation along the cylindrical end:

Lemma 4.5. *Assume that $(\xi|_B, d\alpha|_B)$ is symplectically trivial, and that the linearized Reeb flow equation $\dot{\psi} = \nabla_\psi R_\alpha$ along $B = \partial W$ is strongly index-definite. Then, the linearized Hamiltonian flow equation $\dot{\psi} = \nabla_\psi X_{\widehat{H}}$ of the extension $\widehat{H}$ given by Eq. (4.4) is also strongly index-definite along the cylindrical end $[1, +\infty) \times B$.*

Proof. We prove this using a matrix representation. To do this, we need to symplectically trivialize the full tangent bundle on the cylindrical ends. Given a symplectic trivialization of $(\xi|_B, d\alpha|_B)$, we only need to trivialize the symplectic complement of ξ. We do this using the trivialization $L = \langle Y = r\partial_r, R\rangle$, where $R = R_\alpha/r$ is the Reeb vector field at the r-slice.

We will work with the usual formalism of time-dependent Hamiltonians, and we do not include this time-dependence in the notation. Exterior and covariant derivatives are computed using the base manifold only, and do not involve time derivatives. We will also use the following notation:

$$X_\xi := X^\xi_{\widehat{H}_1} + \frac{r-1}{2} X^\xi_{\widehat{H}_2},$$
$$G := d\widehat{H}_1(R_\alpha) + \frac{r-1}{2} d\widehat{H}_2(R_\alpha).$$

To compute the linearization, we choose a convenient connection ∇, namely the Levi–Civita connection for the metric $1/r^2 \cdot dr \otimes dr + \alpha \otimes \alpha + d\alpha(\cdot, J\cdot)$. This connection has the following properties:

- $\nabla Y = 0$. Keep in mind that Y is the Liouville vector field $r\partial_r$;
- $\nabla_{R_\alpha} R_\alpha = 0$ and $\nabla_Y R_\alpha = 0$;
- $\nabla_X R_\alpha \in \xi$ for all $X \in \xi$.

With respect to this connection, we compute the linearization as

$$\begin{aligned}\nabla X_{\widehat{H}} = F\nabla R_\alpha + dF \otimes R_\alpha + \frac{1}{r^2} dr \\ \otimes (X_\xi - GY) + \frac{r-1}{r}(\nabla X_\xi - dG \otimes Y).\end{aligned} \tag{4.6}$$

Before we continue our analysis of the linearization, we first need to discuss the behaviour of the Hamiltonians $\widehat{H}_j$ and their derivatives under rescaling the interpolation parameter δ_1. We will write the terms in the expression (4.4) as $\widehat{H}'_j$ if we use δ'_1 as interpolation parameter. For $\delta'_1 < \delta_1$, we have the following:

- derivatives in the B-direction (denoted ∂_b) admit a uniform bound, independent of δ_1, that is
$$\max_{[1,+\infty)\times B} |\partial_b^k \widehat{H}'_j| \leq \max_{[1,+\infty)\times B} |\partial_b^k \widehat{H}_j| \text{ for all } k \geq 0;$$
- derivatives in the r-direction scale as follows:
$$\max_{[1,+\infty)\times B} |\partial_r^k \widehat{H}'_j| \leq \left(\frac{\delta_1}{\delta'_1}\right)^k \max_{[1,+\infty)\times B} |\partial_r^k \widehat{H}_j| \text{ for all } k \geq 0.$$

Keeping this scaling behaviour in mind, we regroup terms in Eq. (4.6) to obtain the following representation:
$$\nabla X_{\widehat{H}} = L_0 + L_1,$$
where
$$L_0 = F\nabla R_\alpha + dF \otimes R_\alpha + \frac{1}{r^2} dr \otimes (X_\xi - GY) - \frac{r-1}{r^2} dG(Y) dr \otimes Y$$
$$+ \frac{r-1}{r^2} dr \otimes \nabla_Y X_\xi$$
and
$$L_1 = \frac{r-1}{r}\left(\nabla^\xi X_\xi + \alpha \otimes \nabla_{R_\alpha} X_\xi - R_\alpha(G)\alpha \otimes Y - d^\xi G \otimes Y\right).$$
Here, $\nabla^\xi = P_\xi \nabla|_\xi$, where P_ξ is the orthogonal projection to ξ, and $d^\xi = d|_\xi$. We will explain below that the matrices L_0 and L_1 have the following matrix representations, with respect to the decomposition $T\widehat{W} = \xi \oplus \langle Y, R\rangle$:
$$L_0 = \left(\begin{array}{c|c} F \cdot \nabla^\xi R_\alpha & \begin{matrix} U & 0 \\ V & 0 \end{matrix} \\ \hline \begin{matrix} 0 & 0 \\ W & Z \end{matrix} & \begin{matrix} a & 0 \\ b & c \end{matrix} \end{array}\right), \quad L_1 = \frac{r-1}{r}\left(\begin{array}{c|c} \nabla^\xi X_\xi & \begin{matrix} 0 & U' \\ 0 & V' \end{matrix} \\ \hline \begin{matrix} W' & Z' \\ 0 & 0 \end{matrix} & \begin{matrix} 0 & a' \\ 0 & 0 \end{matrix} \end{array}\right).$$
This is clear for L_0. We further want to show that $L_0 \in \mathfrak{sp}(2n)$, which will constrain the entries more. Since we know that $L_0 + L_1 \in \mathfrak{sp}(2n)$, we will show that $L_1 \in \mathfrak{sp}(2n)$, since the latter contains fewer terms. For this we note the following:

- the matrix representation for $\nabla^\xi X_\xi$ is in $\mathfrak{sp}(2n-2)$. This is because these entries come from the ξ-part of a contact Hamiltonian;
- the matrix representation for $R_\alpha(G)\alpha \otimes Y$ is in $\mathfrak{sp}(2)$: the non-trivial entry corresponds to the element a';
- non-trivial entries in the matrix representation of $-d^\xi G \otimes Y$ appear only on the first row of the lower left block. These correspond to the elements W', Z';
- non-trivial entries in the matrix representation of $\alpha \otimes \nabla_{R_\alpha} X_\xi$ appear as the last column. We will show that these correspond to the elements U' and V'. We claim that $\langle \nabla_{R_\alpha} X_\xi, R_\alpha \rangle = 0$. Indeed, since the contact structure is orthogonal to the Reeb vector field with our choice of metric, we have
$$0 = R_\alpha \langle X_\xi, R_\alpha \rangle = \langle \nabla_{R_\alpha} X_\xi, R_\alpha \rangle + \langle X_\xi, \nabla_{R_\alpha} R_\alpha \rangle = \langle \nabla_{R_\alpha} X_\xi, R_\alpha \rangle.$$

Similarly, we obtain $\langle \nabla_{R_\alpha} X_\xi, Y\rangle = 0$. This means that the L-entries in the matrix representation of $\alpha \otimes \nabla_{R_\alpha} X_\xi$ are zero.

- we have $(W', Z')^T = J \cdot (U', V')^T = (-V', U')^T$. This follows, since $-d^\xi G$ is dual to $\nabla_{R_\alpha} X_\xi$, i.e., $d\alpha(\nabla_{R_\alpha} X_\xi, \cdot) = -d^\xi G$.

We conclude that $L_1 \in \mathfrak{sp}(2n)$, and hence, L_0 is, too. Observe also that for all $\epsilon > 0$, we can choose $\delta_1 > 0$, such that $\|L_1\| < \epsilon$ due to the scaling behaviour we discussed earlier: this can be done in a way that is compatible with Lemma 4.2, i.e., δ_1 getting smaller as ϵ gets smaller.

Since $J_0 L_0$ is symmetric, we can fix the terms of L_0. They must necessarily have the following form:

$$L_0 = \left(\begin{array}{c|c} F \cdot \nabla^\xi R_\alpha & \begin{matrix} U & 0 \\ V & 0 \end{matrix} \\ \hline \begin{matrix} 0 & 0 \\ V & -U \end{matrix} & \begin{matrix} a & 0 \\ b & -a \end{matrix} \end{array}\right) \in \mathfrak{sp}(2n).$$

This matrix has precisely the form that we consider in Appendix E. Moreover, note that strong index-definiteness is invariant under scaling by a positive (possibly time-dependent) function of the generating matrix. Indeed, this scaling has the effect of positively reparametrizing the flow, and so the new flow intersects the Maslov cycle as often as the original one (although the constants in the definition of strong index-definiteness might change). Therefore, since the ODE $\dot\psi = \nabla^\xi_\psi R_\alpha$ is strongly index-definite by assumption and $F > 0$, then so is the ODE $\dot\psi = F \cdot (\nabla^\xi_\psi R_\alpha)$. Lemma E.2 in Appendix E now tells us that the system $\dot\psi = L_0 \psi$ is strongly index-definite. By choosing δ_1 sufficiently small, we can make the matrix L_0 get arbitrarily C^0-close to $L_0 + L_1 = \nabla X_{\widehat{H}}$. Since the system $\dot\psi = L_0\psi$ is strongly index-definite, we can adapt Lemma 2.2.9 from [37] to see that $\dot\psi = \nabla_\psi X_{\widehat{H}}$ is strongly index-definite, too. This concludes the proof of Lemma 4.5. □

We need the following lemma to ensure that our Hamiltonians are weakly admissible.

Lemma 4.6. *Given an extension $\widehat{\tau} : \widehat{W} \to \widehat{W}$ as in the beginning of Sect. 4, there is a Hamiltonian perturbation $\widetilde{\tau} = \phi^1_f \circ \widehat{\tau} : \widehat{W} \to \widehat{W}$ with the following properties:*

(1) *$\phi^1_f(x) = x$ for all x not in a neighborhood of $[1, r_\infty] \times B$, for some fixed $r_\infty > 1$. In particular, all interior fixed points of τ are unaffected by the perturbation;*
(2) *all fixed points of $\widetilde{\tau}|_{[1,r_\infty]\times B}$ are non-degenerate and hence isolated;*
(3) *by the standard composition rule for Hamiltonians $\widetilde{\tau}$ is the time 1-flow of a Hamiltonian $\widetilde{H}$. This Hamiltonian $\widetilde{H}$ is C^2-close to $\widehat{H}$, and its fixed points have Robbin–Salamon index close to that of the unperturbed fixed points.*

Proof. We adapt the argument from [10, Lemma 2] to our setting. Set $V := [1, r_\infty] \times B$, where $r_\infty > 1$ is such that $\widehat{H}$ is linear on $[r_\infty, \infty) \times B$. We need to find a C^2-small function f vanishing on the complement of a neighborhood of V, such that all fixed points of $\phi^1_f \circ \widehat{\tau}$ are non-degenerate on V. Consider the map

$$j : V \longrightarrow V \times V, \quad x \longmapsto (x, \widehat{\tau}(x)),$$

and denote its image by Γ. Define the diagonal $\Delta = \{(v, v) \in V \times V \mid v \in V\}$. Observe that $v \in \mathrm{Fix}(\widehat{\tau})$ is a non-degenerate fixed point of $\widehat{\tau}$ if and only if Γ and Δ intersect transversely at (v, v).

For all points $(v, v) \in \Gamma \cap \Delta$, choose a Darboux ball $B_\epsilon(v) \subset \widehat{W}$, such that v corresponds to 0 in the Darboux ball. To choose ϵ, let $\lambda_{\max}$ denote the maximal (in absolute value) eigenvalue of $d_x\widehat{\tau}$ over all fixed points of $\widehat{\tau}$ in the cylinder. In a formula

$$\lambda_{\max} := \max\{|\lambda| \mid \lambda \text{ eigenvalue of } d_x\widehat{\tau}, x \in \mathrm{Fix}(\widehat{\tau}) \cap [1, r_\infty] \times B\}.$$

Note that $\lambda_{\max} \geq 1$, since $d_x\widehat{\tau}$ is symplectic. Choose ϵ so small, such that the following two properties hold:

- if $x \in B_{\epsilon/2\lambda_{\max}}(v)$, then $\widehat{\tau}(x) \in B_{3\epsilon/4}(v)$.
- for all interior fixed points of τ, i.e., for $x \in \mathrm{Fix}(\tau|_{int(W)})$ we have $d(x, [1, \infty] \times B) > \epsilon$, where d is some fixed reference metric (for example, induced by the Riemannian metric $\widehat{\omega}(\cdot, \widehat{J}\cdot)$).

We give some intuition for these choices, before going into the computation. By the first property, we retain some control after applying $\widehat{\tau}$ to a point that is sufficiently close to a fixed point. Intuitively, if x is close to the fixed point v, then by the definition of $\lambda_{\max}$, the map $\widehat{\tau}$ sends x approximately away by a factor of at most $\lambda_{\max}$, and so we ensure that if $x \in B_{\epsilon/2\lambda_{\max}}(v)$, then $\widehat{\tau}(x) \in B_{3\epsilon/4}(v)$. Below, we will define Hamiltonian functions to perturb the map $\widehat{\tau}$, and this property will ensure that we have maximal control over the value of the Hamiltonian vector fields. This point is actually not essential, but it makes the computation below a little more uniform.

We now come to our Hamiltonian perturbation functions. Choose functions $f_{v,i}$ for $i=1, \ldots, 2n$, such that, in Darboux coordinates $z=(z_1, \ldots, z_{2n})$, we have $f_{v,i}(z) = z_i \cdot \rho_v(z)$, and ρ_v is a cut-off function that equals 1 on $B_{3\epsilon/4}(v)$ and vanishes outside $B_\epsilon(v)$. For the sake of explicitness, note that the Hamiltonian vector field of $f_{v,i}$ is given by

$$X_{f_{v,i}} = \rho_v(z) J_0 \cdot e_i + z_i \cdot X_{\rho_v},$$

where e_i is the ith standard basis vector, and J_0 is the standard complex structure on the Darboux ball $B_\epsilon(v)$. By construction, this vector field vanishes on the complement of $B_\epsilon(v)$. Moreover, for sufficiently small r, the time-1 flow of the Hamiltonian vector field of $r f_{v,i}$ on the smaller ball $B_{\epsilon/2\lambda_{\max}}(v)$ is the map $z \mapsto z + rJ_0 \cdot e_i$. If $x \in \mathrm{Fix}(\widehat{\tau}) \cap B_{\epsilon/2\lambda_{\max}}(v)$, we have

$$\frac{\partial}{\partial r}\bigg|_{r=0} \phi^1_{r f_{v,i}} \circ \widehat{\tau}(x) = \frac{\partial}{\partial r}\bigg|_{r=0} \phi^r_{f_{v,i}} \circ \widehat{\tau}(x) = X_{f_{v,i}}(\widehat{\tau}(x)).$$

For completeness, we observe that each of these functions $f_{v,i}$ is a C^∞ function defined on all of $\widehat{W}$, vanishing outside $B_\epsilon(v)$.

Since $\Gamma \cap \Delta$ is compact, we find a finite cover of its projection to V of the form $\bigcup_{v \in I} B_{\epsilon/2\lambda_{\max}}(v)$. We make the following observation. Consider

$x \in \mathrm{Fix}(\widehat{\tau}) \cap [1, r_\infty] \times B$. Then, there is $v \in I$, such that $x \in B_{\epsilon/2\lambda_{\max}}(v)$. On this small ball, the Hamiltonian vector fields associated with $f_{v,1}, \ldots, f_{v,2n}$ are linearly independent, and form a basis of sections.

Define the finite-dimensional vector space

$$R := \left(\mathbb{R}^{2n}\right)^{\#I} = \mathbb{R}^D,$$

where we have set $D := 2n\#I$. We relabel the functions $f_{v,i}$ using a single index j, and put $f = (f_1, \ldots, f_D)$. Define the projection $p : V \times V \times R \to R$, $(x, y, r) \mapsto r$, and consider the "universal" space

$$\Gamma_R = \left\{ (x, y, r) \in V \times V \times R \;\middle|\; x \in V, r = (r_1 \ldots, r_D) \in R, \right.$$
$$\left. y = \phi^1_{r \cdot f} \circ \widehat{\tau}(x),\ r \cdot f = \sum_j r_j f_j \right\}.$$

Note that the function $r \cdot f$ is a C^∞ function defined on all of $\widehat{W}$: this function vanishes outside a neighborhood of $[1, r_\infty] \times B$.

Claim: The space Γ_R intersects the enlarged diagonal $\Delta_R = \{(v, v, r) \in V \times V \times R\}$ transversely for r that are sufficiently close to 0. In particular, $V_R := \Gamma_R \cap \Delta_R$ is a submanifold. □

Proof of claim. To verify the claim, we compute the derivatives of the map

$$j_R : (x, r) \mapsto (x, \phi^1_{r \cdot f} \circ \widehat{\tau}(x), r)$$

and the corresponding map for Δ_R, $j_\Delta : (x, r) \mapsto (x, x, r)$. For j_R, we find the derivative

$$d_{x,r=0} j_R = \begin{pmatrix} \mathrm{id}_V & 0 \\ d_x \widehat{\tau} & X_{f_1}(\widehat{\tau}(x)), \ldots, X_{f_D}(\widehat{\tau}(x)) \\ 0 & \mathrm{id}_R \end{pmatrix}.$$

For j_Δ, we find the derivative

$$d_{x,r=0} j_\Delta = \begin{pmatrix} \mathrm{id}_V & 0 \\ \mathrm{id}_V & 0 \\ 0 & \mathrm{id}_R \end{pmatrix}.$$

Given a point $(x, x, 0) \in V_R = \Gamma_R \cap \Delta_R$ (so $\widehat{\tau}(x) = x$), there is $v \in I$, such that $x \in B_{\epsilon/2\lambda_{\max}}(v)$. By construction, the vector fields $X_{f_{v,1}} \circ \widehat{\tau}, \ldots, X_{f_{v,2n}} \circ \widehat{\tau}$ are linearly independent on $B_{\epsilon/2\lambda_{\max}}(v)$. This means that, taken together, the matrix representations of $d_{x,r=0} j_R$ and $d_{x,r=0} j_\Delta$ have full rank, namely $2n + 2n + D$. We conclude that j_R is transverse to the enlarged diagonal Δ_R for $r = 0$, and hence, by compactness, also for small r. □

Applying Sard's theorem to the projection $p|_{V_R}$, we find a regular value r_0 of $p|_{V_R}$ close to 0. We see that

$$\Gamma_{r_0} = \{(x, y) \in V \times V \mid y = \phi^1_{r_0 \cdot f} \circ \widehat{\tau}(x)\}$$

intersects Δ transversely. This means that all fixed points of $\phi^1_{f_0} \circ \widehat{\tau}$ in V are non-degenerate, where $f_0 = r_0 \cdot f$, so claim (2) holds. Since we can choose the regular value r_0 arbitrarily small, and since the support of the perturbation f_0 is a small neighborhood of $[1, r_\infty] \times B$, the claim (1) holds. To see that claim (3) also holds, we note that the free homotopy class of a 1-periodic orbit is not affected by this perturbation if it is sufficiently small. For the index, we use the same argument as before. The unperturbed system is strongly index-definite, and the same will be true for small perturbations. This concludes the proof of the lemma. □

Proof of Theorem A ($\dim W > 2$). Write $\tau = \phi_H^1$ for H as in Definition 1.1. Assuming its interior fixed points are isolated, we have finitely many isolated interior 1-periodic orbits of H, say $\gamma_1, \ldots, \gamma_k$. The starting points $\gamma_1(0), \ldots, \gamma_k(0)$ are the fixed points of τ.

Assume by contradiction that the minimal periods of all interior periodic points of τ are, in increasing order, given by $m_0 = 1, m_1, \ldots, m_\ell$. Take an increasing sequence $\{p_i\}_{i=1}^{\infty}$ going to infinity, such that each p_i is indivisible by the $m_1, \ldots, m_\ell$. For instance, one can take the sequence $\{p_i\}$ to be an increasing sequence of primes all of which are larger than $\max_j m_j$.

As in the proof of Corollary 4.4, we can appeal to Remark 3.14 to define Floer homology for a possibly degenerate Hamiltonian. Indeed, due to the choice of p_i's, all fixed points of $\widehat{\tau}^{p_i}$ are isolated, and we can apply Lemma 4.6 if necessary to perturb the Hamiltonian $\widehat{H}^{\# p_i}$ on the cylindrical part $[1, r_\infty] \times B$, for some r_∞. This ensures that the Hamiltonian $\widehat{H}^{\# p_i}$ is weakly admissible, so we can use local Floer homology and the spectral sequence from Proposition 3.13, to define $HF(\widehat{H}^{\# p_i})$. Hence, we can compute symplectic homology as $SH_\bullet(W) = \varinjlim_i HF_\bullet(\widehat{H}^{\# p_i})$. By Lemma 3.15, for all $N > 2nk$, where $\dim(W) = 2n$, we find distinct degrees $i_1, \ldots, i_N$, such that $SH_{i_j}(W) \neq 0$, ordered by increasing absolute value. By Lemma 4.5, we can choose p_i sufficiently large, such that the following holds:

(1) Each 1-periodic orbit of $\widehat{H}^{\# p_i}$ that is contained in $\widehat{W} \setminus \mathrm{int}(W)$ has RS-index whose absolute value is larger than $|i_N| + 2n$;
(2) the Floer homology groups $HF_{i_j}(\widehat{H}^{\# p_i})$ are non-trivial for $j = 1, \ldots, N$.

Now, consider the spectral sequence from Proposition 3.13 for $\widehat{H}^{\# p_i}$. We deduce from (2) that there must be non-trivial summands on $E^1_{pq}(\widehat{H}^{\# p_i})$ with $p + q = i_j$ for $j = 1, \ldots, N$. Since the terms of the spectral sequence are made up from local Floer homology groups, and we know from (1) that no 1-periodic orbit in $\widehat{W} \setminus \mathrm{int}(W)$ can contribute to local Floer homology of degree i_j, we conclude that every term $E^1_{pq}(\widehat{H}^{\# p_i})$ in the spectral sequence with $p + q = i_j$ must come from the local Floer homology of an orbit γ in $\mathrm{int}(W)$.

Because we have assumed that the p_i's are indivisible by $m_1, \ldots, m_\ell$, we conclude that each such orbit γ must be an iterate of one of the orbits $\gamma_1, \ldots, \gamma_k$. Moreover, by (3.3) and Sect. 3.2:

$$\mathrm{supp} HF^{loc}_\bullet(\gamma_j^{p_i}, \widehat{H}^{\# p_i}) \subset [p_i \Delta(\gamma_j) - n, p_i \Delta(\gamma_j) + n].$$

This covers at most $2nk$ different degrees, leaving some of the degree i_j uncovered as we had chosen $N > 2nk$. This is a contradiction. □

Proof of Theorem B. We only need to show that $\dim SH_\bullet(W) = \infty$. Since $W \subset T^*M$ is star-shaped, from Viterbo's theorem [38], we have $SH_\bullet(W) \cong H_\bullet(\mathcal{L}M; \mathbb{Z}_2)$ where $\mathcal{L}M$ is the free loop space of M. The statement is more subtle when using $\mathbb{Z}$ or $\mathbb{Q}$-coefficients, see [4]. Now, we can apply the following theorem due to Gromov: □

Theorem. [19, Sect. 1.4] *Let (M, g) be a closed Riemannian manifold with finite fundamental group. For $a > 0$, let $\mathcal{L}M$ be the free loop space of M, and let $\mathcal{L}^{<a}M \subset \mathcal{L}M$ denote the space of free loops with length less than a. Let $\iota^a : \mathcal{L}^{<a}M \hookrightarrow \mathcal{L}M$ denote the inclusion, and $\iota^a_k : H_k(\mathcal{L}^{<a}M; \mathbb{R}) \to H_k(\mathcal{L}M; \mathbb{R})$ be the map induced in real homology of degree k. Then, there exists a positive constant $C = C(M, g)$, such that*

$$\sum_{k \geq 0} rank(\iota^a_k) \geq Ca.$$

Together with the above, this tells us that $SH_\bullet(W)$ is infinite dimensional. □

Acknowledgements

The authors thank Urs Frauenfelder, for suggesting this problem to the first author, for his generosity with his ideas and for insightful conversations throughout the project; Lei Zhao, Murat Saglam, Alberto Abbondandolo, and Richard Siefring, for further helpful inputs, interest in the project, and discussions. We are also grateful to an anonymous referee for careful comments on earlier versions. The first author has also significantly benefited from several conversations with Kai Cieliebak in Germany and Sweden, as well as with Alejandro Passeggi in Montevideo, Uruguay. This research was started, while the first author was affiliated to Augsburg Universität, Germany. The first author is also indebted to a Research Fellowship funded by the Mittag-Leffler Institute in Djursholm, Sweden, where this manuscript was finalized. The second author was supported by NRF grant NRF-2019R1A2C4070302, which was funded by the Korean Government.

Appendix A: Hamiltonian twist maps: examples and non-examples

We will now discuss some examples that help clarify the nature of the Hamiltonian twist condition.

A.1. Examples

The following construction, an adaptation of a standard one, further illustrates that the Hamiltonian twist condition is *not* localized at B.

Proposition A.1. *For each $k, \ell \in \mathbb{N}$, there are strict contact manifolds $(Y_k, \alpha_{k,\ell})$ carrying adapted open books $(B_k = B, \pi_k)$, $\pi_k : Y_k \backslash B \to S^1$, with fixed page Σ, such that the following holds:*

- *The return maps τ_k all agree in a collar neighborhood of $B = \partial\Sigma$, and are generated by Hamiltonians H_k;*
- *Furthermore, there is a symplectomorphism ϕ_k from $\mathring{\Sigma}$, the interior of the page Σ, to the open subset W_2 of the Liouville completion $\widehat{W}$ of a fixed Liouville domain (W, λ) with $\partial W = B$, where*

$$W_2 = W \cup_\partial ([1,2) \times B, d(r\alpha_B)),$$

and $\alpha_B = \lambda|_B$ is the contact form at B, and $r \in [1,2)$.
- *The return map $\phi_k \circ \tau_k \circ \phi_k^{-1}$ extends to a Hamiltonian diffeomorphism $\bar{\tau}_k$ on the closure $\bar{W}_2$, generated by Hamiltonians $\bar{H}_k$.*
- *The Hamiltonian twist condition holds for $\bar{H}_k$ for $k \leq \ell$, but not for $k > \ell$.*

Proof. Consider a Liouville domain (W, λ) with a 2π-periodic Reeb flow on its boundary (e.g., D^*S^2). We identify a collar neighborhood $\nu_W(B)$ of $B = \partial W$ with $(1/2, 1] \times B$, where $B = \{r = 1\}$, via a diffeomorphism $\varepsilon : (1/2, 1] \times B \longrightarrow \nu_W(B) \subset W$. We assume $\lambda = r\alpha_B$ along $\nu_W(B)$, $\alpha_B = \lambda|_B$. Define the smooth Hamiltonian

$$H(x) = \begin{cases} 0, & \text{if } x \notin \nu_W(B), \\ f(r), & \text{if } x = \varepsilon(r, b) \in \nu_W(B). \end{cases}$$

Here, f is a smooth, decreasing function with the property

- $f(1/2) = 0$;
- $f'(r) \geq -2\pi$ and $f'(r) = -2\pi$ near $r = 1$.

The Hamiltonian vector field of H is given by

$$X_H(x) = \begin{cases} 0 & \text{if } x \notin \nu_W(B), \\ f'(r)R_\alpha & \text{if } x = \varepsilon(r, b) \in \nu_W(B). \end{cases}$$

Define the fibered Dehn twist by $\tau(x) = Fl_1^{X_H}(x)$, where $Fl_t^{X_H}$ is the Hamiltonian flow of H with respect to $d\lambda$. We have $\tau^*\lambda = \lambda - \mathrm{d}U$, where we choose the primitive U to be a negative function: with a computation, we can show that it is possible to choose $U(1) = -2\pi$, and will do so. The iterate τ^k is generated by $H_k = kH$, and $(\tau^k)^*\lambda = \lambda - \mathrm{d}U_k$, with $U_k = \sum_{j=0}^{k-1}(U \circ \tau^j)$.

We consider the associated open book

$$Y_k = OB(W, \tau^k) := B \times D^2 \cup_\partial W_{\tau^k},$$

where $W_{\tau^k} = W \times \mathbb{R}/(x,t) \sim (\tau^k(x), t + U_k(x))$ is the mapping torus. The manifold Y_k carries an adapted contact form $\alpha_{k,\ell}$ which looks like $\alpha_{k,\ell} = \lambda + d\theta$ along W_{τ^k}, and $\alpha_{k,\ell} = h_1(\rho)\alpha_B + h_2(\rho)d\theta$ along $B \times D^2$. Here, $(\rho, \theta) \in D^2$, and h_1 and $h_2 = h_{2,k,\ell}$ are suitable profile functions, which we will fix now. Choose h_1 and h_2, such that

- they do not depend on k for $\rho \leq 1/2$;
- $h_1' \leq 0$ with equality only at $\rho = 0$. We may take $h_1(\rho) = 2 - \rho^2$ near $\rho = 0$ (this is not essential but very convenient);
- near $\rho = 0$, we have $-\frac{h_2'}{h_1'}(\rho) = \ell + \epsilon > 0$ (non-singular) for some small $\epsilon \in (0, 1)$.

- $h_2 \equiv k$, $h_1 = -\rho + 2$ near $\rho = 1$ (so h_2 depends on k on the interval $(1/2, 1]$).

Note that, in the definition of Y_k, the binding model is glued to the mapping torus using the gluing map

$$\Phi_{glue} : B \times D^2_{\rho > 1/2} \longrightarrow W_{\tau^k}$$
$$(b; \rho, \theta) \longmapsto \left(2 - \rho, b; \frac{-U_k(1)\theta}{2\pi}\right) = (2 - \rho, b; k\theta).$$

This pulls back $d\theta + \lambda$ to $kd\theta + (2 - \rho)\alpha_B$. This explains the above choices.

The global hypersurface of section, i.e., a fixed page, is $\Sigma = W \cup_\partial B \times [0, 1]$, with coordinate $\rho \in [0, 1]$, and we can compute the return map τ_k explicitly. We find

$$\tau_k(x) = \begin{cases} \tau^k(x), & \text{if } x \in W, \\ (Fl^R_{-2\pi h_2'(\rho)/h_1'(\rho)}(b), \rho), & \text{if } x = (b, \rho) \in B \times [0, 1], \end{cases}$$

where Fl^R_t is the Reeb flow of α_B at B. The Hamiltonian generating τ_k can be obtained by patching H_k on W to a Hamiltonian that generates τ_k along $B \times [0, 1]$; we need to match the slopes on the boundary, which can be done by rewriting $\tau_k(b, \rho) = (Fl^R_{-2\pi(h_2'(\rho)/h_1'(\rho)+k)}(b), \rho)$. Then, H_k extends to Σ via $\bar{H}_k(r) = -2\pi \int_1^\rho (h_2'(s)/h_1'(s) + k)h_1'(s)ds + f(1)$ along $B \times [0, 1]$. Note that the form $d\lambda$ also extends along $B \times [0, 1]$ via $d\alpha_{k,\ell}|_\Sigma = h_1'(\rho)d\rho \wedge \alpha_B$. Therefore, H_k generates τ_k, and τ_k is independent of k on the collar neighborhood $B \times [0, 1/2]$.

To complete the proof, we first note that the 2-form $d(h_1(\rho)\alpha_B)$ is degenerate on $\partial\Sigma$. However, the map

$$\phi_k : \mathring{\Sigma} \longrightarrow W_2 = W \cup_\partial ([1, 2) \times B, d(r\alpha_B)),$$
$$w \longmapsto \begin{cases} w & w \in W \\ (h_1(\rho), b) & w = (b, \rho) \in B \times (0, 1] \end{cases}$$

is a symplectomorphism, and the closure of W_2 is an actual Liouville domain. Furthermore, due to our explicit choice $h_1(\rho) = 2 - \rho^2$ near $\rho = 0$, we find $\rho = \sqrt{2 - r}$, so we can compute the conjugated return map $\phi_k \circ \tau_k \circ \phi_k^{-1}$ near $r = 0$ as

$$\phi_k \circ \tau_k \circ \phi_k^{-1}(r, b) = (r, Fl^R_{2\pi(\ell+\epsilon)}(b)).$$

This map extends to a symplectomorphism $\bar{\tau}_k : \bar{W}_2 \to \bar{W}_2$. Here, note that ϕ_k^{-1} is not smooth at $r = 2$, but this is resolved by the explicit form of τ_k, which does not contain any ρ dependence in the B-direction near $\rho = 0$. This extended map is still Hamiltonian, and satisfies the twist condition for $k \leq \ell$, but not for $k > \ell$.

Therefore, it satisfies the claim of the proposition. □

Remark A.2. Given a return map τ that is Hamiltonian, we point out that the Hamiltonian family generating τ is not unique, and more importantly, that various dynamical properties depend on the choice of Hamiltonian. For example, on the disk $(D^2, rdr \wedge d\theta)$, the return map $\tau = \text{id}$ is generated by the autonomous Hamiltonians $H_k = k\pi r^2$. For given k, the Robbin–Salamon index of the 1-periodic orbit at ∂D^2 is $2k$, i.e., k-dependent. The associated paths of symplectic matrices have the same endpoints, but are not homotopic rel endpoints. This also illustrates the interpretation of the RS-index as a winding number. Note that D^2 has a Hamiltonian circle action that extends over the whole space. We do not know whether the same type of phenomenon occurs for more general symplectic manifolds (i.e., without a global Hamiltonian circle action).

A.2. Non-examples: Katok examples

In [25], Katok constructed examples of non-reversible Finsler metrics on S^n with only finitely many simple closed geodesics. Here is a description of such examples using Brieskorn manifolds. We consider

$$\Sigma^{2n-1} := \left\{ (z_0, \dots, z_n) \in \mathbb{C}^{n+1} \;\middle|\; \sum_j z_j^2 = 0 \right\} \cap S_1^{2n+1},$$

equipped with the contact form $\alpha = \frac{i}{2} \sum_j z_j d\bar{z}_j - \bar{z}_j dz_j$. These spaces are contactomorphic to S^*S^n with its canonical contact structure. The given contact form is actually the prequantization form.

We describe the setup in detail when $n = 2m+1$ is odd. We group the coordinates in pairs, and make the following unitary coordinate transformation:

$$\begin{aligned} w_0 = z_0, w_1 = z_1, w_{2j} = \frac{\sqrt{2}}{2}(z_{2j} + i z_{2j+1}), w_{2j+1} \\ = \frac{i\sqrt{2}}{2}(z_{2j} - i z_{2j+1}) \text{ for } j = 1, \dots, m. \end{aligned}$$

Because this is a unitary transformation, the form α, expressed in w-coordinates, still has the form

$$\alpha = \frac{i}{2} \sum_j w_j d\bar{w}_j - \bar{w}_j dw_j.$$

For a tuple $\epsilon = (\epsilon_1, \dots, \epsilon_m) \in (-1, 1)^m$, define the function H_ϵ on a neighborhood of Σ^{2n-1} via

$$H_\epsilon(w) = \|w\|^2 + \sum_j \epsilon_j (|w_{2j}|^2 - |w_{2j+1}|^2).$$

For ϵ sufficiently small, this function is positive, so we define a perturbed contact form by

$$\alpha_\epsilon = H_\epsilon^{-1} \cdot \alpha.$$

The Reeb vector field of α_ϵ is

$$R_\epsilon = X_\epsilon + \overline{X}_\epsilon,$$

where

$$X_\epsilon = iw_0 \frac{\partial}{\partial w_0} + iw_1 \frac{\partial}{\partial w_1} + \sum_j \left(i(1+\epsilon_j)\frac{\partial}{\partial w_{2j}} + i(1-\epsilon_j)\frac{\partial}{\partial w_{2j+1}} \right),$$
$$\overline{X}_\epsilon = -i\overline{w}_0 \frac{\partial}{\partial \overline{w}_0} - i\overline{w}_1 \frac{\partial}{\partial \overline{w}_1} - \sum_j \left(i(1+\epsilon_j)\frac{\partial}{\partial \overline{w}_{2j}} + i(1-\epsilon_j)\frac{\partial}{\partial \overline{w}_{2j+1}} \right).$$

The Reeb flow is therefore given by

$$\begin{aligned}(w_0,\dots,w_n) \longmapsto (&e^{2\pi it}w_0, e^{2\pi it}w_1, e^{2\pi it(1+\epsilon_1)}w_2, e^{2\pi it(1-\epsilon_1)}w_3,\\ &\dots, e^{2\pi it(1+\epsilon_m)}w_{n-1}, e^{2\pi it(1-\epsilon_m)}w_n).\end{aligned}$$

This flow has only $n+1$ periodic orbits if all ϵ_j are rationally independent. These are given by

$$\begin{aligned}
\gamma_0(t) &= \left(\frac{1}{\sqrt{2}}e^{2\pi it}, \frac{i}{\sqrt{2}}e^{2\pi it}, 0,\dots,0\right), t\in[0,1]\\
\beta_0(t) &= \left(\frac{1}{\sqrt{2}}e^{2\pi it}, -\frac{i}{\sqrt{2}}e^{2\pi it}, 0,\dots,0\right), t\in[0,1]\\
\gamma_j(t) &= \left(0,0,\dots,e^{2\pi it(1+\epsilon_j)},0,\dots,0,0\right), t\in[0,1/(1+\epsilon_j)]\\
\beta_j(t) &= \left(0,0,\dots,0,e^{2\pi it(1-\epsilon_j)},\dots,0,0\right), t\in[0,1/(1-\epsilon_j)]
\end{aligned}$$

for $j=1,\dots,m$.

Remark A.3. As stated, we see that there are only finitely many periodic orbits. Furthermore, since the unperturbed system, i.e., $\epsilon=0$, describes the geodesic flow on the round sphere, and the perturbation α_ϵ is C^2-small for small ϵ, it follows that the Reeb flow of the contact form α_ϵ corresponds to the geodesic flow of a Finsler metric. In Sect. A.3, we describe how to obtain an explicit relation with the famous Katok examples for S^*S^2.

We construct a supporting open book for the contact form α_ϵ using the map

$$\Theta : \Sigma^{2n-1} \longrightarrow \mathbb{C}, (w_0,w_1,\dots,w_n) \longmapsto w_0.$$

The zero set of Θ defines the binding, the pages are the sets of the form $P_\theta = \{\arg\Theta = \theta\}$, $\theta\in S^1$, which are all copies of $\mathbb{D}^*S^{n-1}$, and the monodromy is τ^2 where τ is the Dehn–Seidel twist. The (boundary extended) return map for the page $P_0 = \Theta^{-1}(\mathbb{R}_{>0}) \cong \mathbb{D}^*S^{n-1}$ is

$$\begin{aligned}
&\Phi : P_0 \longrightarrow P_0,\\
p = (r_0,w_1,w_2,w_3,\dots,w_{n-1},w_n) \longmapsto (&r_0,w_1,e^{2\pi i\epsilon_1}w_2, e^{-2\pi i\epsilon_1}w_3,\dots,\\
&e^{2\pi i\epsilon_m}w_{n-1}, e^{-2\pi i\epsilon_m}w_n).
\end{aligned}$$

Here, $w_0 = r_0 \in \mathbb{R}_{\geq 0}$ is a real non-negative number, and note that the first return time is constant equal to 1 (which follows by looking at the first coordinate). If all ϵ_j are irrational and rationally independent, this map has only two periodic points, both actually fixed, given by

$$p_0 = \left(\frac{1}{\sqrt{2}}, \frac{i}{\sqrt{2}}, 0, \dots, 0 \right)$$
$$q_0 = \left(\frac{1}{\sqrt{2}}, -\frac{i}{\sqrt{2}}, 0, \dots, 0 \right).$$

Note that p_0, q_0 are both interior fixed points, and irrationality of the ϵ_j implies that there are no boundary fixed points. We will explain now why this map is Hamiltonian with boundary preserving Hamiltonian flow. The symplectic form on the interior of the page P_0 is the restriction of $d\alpha_\epsilon$. To manipulate this, let us define

$$H = \|w\|^2, \quad \Delta_\epsilon = \sum_j \epsilon_j (|w_{2j}|^2 - |w_{2j+1}|^2),$$

so $H_\epsilon = H + \Delta_\epsilon$. Observe that the return map Φ is generated by the 2π-flow of the vector field

$$X = i \sum_{j=1}^m \epsilon_j \left(w_{2j} \frac{\partial}{\partial w_{2j}} - \bar{w}_{2j} \frac{\partial}{\partial \bar{w}_{2j}} - w_{2j+1} \frac{\partial}{\partial w_{2j+1}} + \bar{w}_{2j+1} \frac{\partial}{\partial \bar{w}_{2j+1}} \right).$$

This vector field is tangent to the page and preserves H and Δ, and hence also H_ϵ. Plug X in into $d\alpha_\epsilon$. We find

$$\begin{aligned} \iota_X (dH_\epsilon^{-1} \wedge \alpha + H_\epsilon^{-1} d\alpha) &= -\alpha(X) dH_\epsilon^{-1} + H_\epsilon^{-1} \iota_X d\alpha \\ &= -\Delta_\epsilon dH_\epsilon^{-1} - H_\epsilon^{-1} d\Delta_\epsilon = -d(H_\epsilon^{-1}\Delta_\epsilon). \end{aligned}$$

This means that the Hamiltonian generating the return map is $H_\epsilon^{-1}\Delta_\epsilon$. Moreover, index-positivity follows, by observing that it holds for the round metric on S^2 and the fact that it is an open condition. It follows from Theorem B that Φ does *not* satisfy the twist condition for any Liouville structure on $\mathbb{D}^* S^2$.

Remark A.4. The setup for n even is very similar: we drop the w_0-coordinate.

A.3. Relation with the Katok examples

We explain how to see that the above dynamical systems indeed correspond to the Katok examples in case of S^*S^2 (i.e., $n = 2$). More precisely, we will show that the geodesic flow of the Katok examples is conjugated to the Reeb flow of α_ϵ. We need some preparation, which applies to all dimensions, before we specialize to dimension 3. First of all, we fix positive weights $(a_1, \dots, a_n) \in \mathbb{R}^n_{>0}$. Then, we define the 1-forms on the sphere S^{2n-1} given by

$$\beta_0 = \iota^* \left(\frac{\sum_j x_j dy_j - y_j dx_j}{\sum_k a_k (x_k^2 + y_k^2)} \right) = \iota^* \left(\frac{1}{\sum_k a_k |z_j|^2} \frac{i}{2} \sum_j z_j d\bar{z}_j - \bar{z}_j dz_j \right)$$

and

$$\beta_1 = \iota^* \left(\sum_j \frac{1}{a_j} (x_j dy_j - y_j dx_j) \right) = \iota^* \left(\frac{i}{2} \sum_j \frac{1}{a_j} \left(z_j d\bar{z}_j - \bar{z}_j dz_j \right) \right),$$

where ι is the inclusion map $S^{2n-1} \to \mathbb{R}^{2n}$. We will show that the first form is a contact form and that it is strictly contactomorphic to the latter. For this, consider the map

$$\begin{aligned}\psi : S^{2n-1} &\longrightarrow S^{2n-1}\\ (z_1,\ldots,z_n) &\longmapsto \left(\sqrt{\frac{a_1}{\sum_k a_k(x_k^2+y_k^2)}}z_1,\ldots,\sqrt{\frac{a_n}{\sum_k a_k(x_k^2+y_k^2)}}z_n\right).\end{aligned}$$

We find

$$\begin{aligned}\psi^*\beta_1 &= \sum_j \frac{1}{a_j}\sqrt{\frac{a_j}{\sum_k a_k(x_k^2+y_k^2)}}\\ &\quad\times\left(\sqrt{\frac{a_j}{\sum_k a_k(x_k^2+y_k^2)}}(x_jdy_j-y_jdx_j)\right.\\ &\quad\left.+(x_jy_j-y_jx_j)d\left(\sqrt{\frac{a_j}{\sum_k a_k(x_k^2+y_k^2)}}\right)\right)\\ &= \beta_0.\end{aligned}$$

This also shows that β_0 is a contact form, as ψ is a diffeomorphism. We have shown:

Lemma A.5. *The form β_0 is a contact form, and it is strictly contactomorphic to β_1. The Reeb field for β_k for $k=0,1$ is given by*

$$R = \sum_j a_j\left(x_j\frac{\partial}{\partial y_j} - y_j\frac{\partial}{\partial x_j}\right).$$

We now specialize to the 3-dimensional situation, for which in w-coordinates (cf. Remark A.4), we have

$$\Sigma^3 = \left\{(w_1,w_2,w_3)\in\mathbb{C}^3 : w_1^2 - 2iw_2w_3 = 0\right\}\cap S^3.$$

Consider the explicit covering map

$$\pi : S^3 \longrightarrow \Sigma^3, (z_0,z_1)\longmapsto\left(w_1=\sqrt{2}z_0z_1, w_2=z_0^2, w_3=-iz_1^2\right).$$

We quickly verify that this is a covering map

- we have $w_1^2 - 2iw_2w_3 = 2z_0^2z_1^2 - 2z_0^2z_1^2 = 0$;
- we have $|w_1|^2+|w_2|^2+|w_3|^2 = 2|z_0|^2|z_1|^2+|z_0|^4+|z_1|^4 = (|z_0|^2+|z_1|^2)^2 = 1$;
- the map is two to one, since all entries are quadratic.

We compute the pullback $\pi^*\alpha_\epsilon$

$$\begin{aligned}\pi^*\alpha_\epsilon &= \frac{1}{2|z_0|^2|z_1|^2+|z_0|^4+|z_1|^4+\epsilon|z_0|^4-\epsilon|z_1|^4}2(|z_0|^2+|z_1|^2)\frac{i}{2}\sum_j(z_jd\bar z_j - \bar z_jdz_j)\\ &= \frac{2(|z_0|^2+|z_1|^2)}{(|z_0|^2+|z_1|^2)((1+\epsilon)|z_0|^2+(1-\epsilon)|z_1|^2)}\frac{i}{2}\sum_j(z_jd\bar z_j-\bar z_jdz_j)\\ &= \frac{2}{((1+\epsilon)|z_0|^2+(1-\epsilon)|z_1|^2)}\frac{i}{2}\sum_j(z_jd\bar z_j-\bar z_jdz_j).\end{aligned}$$

By Lemma A.5, the form $\pi^*\alpha_\epsilon$ to strictly contactomorphic to the contact form β_1 with weights $a_1 = 1+\epsilon$ and $a_2 = 1-\epsilon$, which is just the ellipsoid model for the contact 3-sphere. To complete the argument, we use a result due to Harris and Paternain [20, Sect. 5], which relates the ellipsoids to the Katok examples.

Appendix B: Symplectic homology of surfaces

Let us consider connected Liouville domains in dimension 2. The simplest such Liouville domain is D^2, which has vanishing symplectic homology. For all other surfaces, note:

Lemma B.1. *Let (W,λ) be a connected Liouville domain of dimension 2. Assume that W is not diffeomorphic to D^2. Take a periodic Reeb orbit δ on one of the boundary components of W. Then, $[\delta]\in\tilde{\pi}_1(W)$ is non-trivial. Furthermore, if δ_1 and δ_2 are periodic Reeb orbits on different boundary components, then $[\delta_1]\neq[\delta_2]$ as free homotopy classes.* □

Assume $W\neq D^2$, and denote the completion by $\widehat{W}$. Then, the chain complex for an admissible Hamiltonian $\widehat{H}$ that is both negative and C^2-small on W has the form

$$CF_\bullet(\widehat{H})=\bigoplus_{\delta\in\tilde{\pi}_1(W)}CF_\bullet^\delta(\widehat{H}),$$

where $CF_\bullet^\delta(\widehat{H})$ is generated by 1-periodic orbits in the free homotopy class δ. The direct summand corresponding to contractible orbits needs as least as many generators as $\operatorname{rk}H_\bullet(W)$ by the Morse inequalities.

Lemma B.2. *For each class δ, the direct summand $CF_\bullet^\delta(\widehat{H})$ forms a subcomplex, and so, we have a splitting*

$$HF_\bullet(\widehat{H})=\bigoplus_{\delta\in\tilde{\pi}_1(W)}HF_\bullet^\delta(\widehat{H}).$$

In addition, as ungraded modules, we have

$$HF_\bullet^\delta(\widehat{H})\cong\begin{cases}\mathbb{Z}^2 & \textit{if } \delta \textit{ is a positive boundary class, and slope}(\widehat{H})\textit{is sufficiently large},\\ H_\bullet(W) & \textit{if } \delta \textit{ is the trivial class},\\ 0 & \textit{otherwise}.\end{cases}$$

Here, a positive boundary class just means a homotopy class of a positive multiple of a boundary component (oriented according the positive boundary orientation).

Proof. The first assertion follows from the fact that Floer cylinders do not change the free homotopy class. For the second claim, we use:

- The Floer differential of a C^2-small Hamiltonian between critical points is the Morse differential, which implies the second case.

- After a suitable Morse perturbation (a Morse function on S^1 with precisely two critical points) breaking the S^1-symmetry given by time-shifts, each positive boundary class gives two generators, corresponding to the critical points of the Morse function on S^1; as shown in [12], the differential is the Morse differential, which vanishes. Moreover, this symmetry-breaking process preserves the homotopy classes of periodic orbits, as observed in Remark 3.12. □

Corollary B.3. *Suppose that W is a connected Liouville domain of dimension 2. Assume that W is not diffeomorphic to D^2. Then, as an ungraded module, we have*

$$SH_\bullet(W) \cong H_\bullet(W) \oplus \bigoplus_{\delta \text{ positive boundary class}} \mathbb{Z}^2.$$

Appendix C: On symplectic return maps

In this appendix, for convenience of the reader, we collect some standard facts concerning return maps arising from a given Reeb dynamics on some contact manifold (cf. the construction of the *Calabi homomorphism*, e.g., in [28, Sect. 10.3], or [2, Sect. 3.3] for the case of the 2-disk). In particular, we show that long Hamiltonian orbits on a global hypersurface of section correspond to long Reeb orbits on the ambient contact manifold.

Consider a map $\tau : \mathrm{int}(\Sigma) \to \mathrm{int}(\Sigma)$ defined on the interior of a $2n$-dimensional Liouville domain Σ. We assume that Σ arises as a (connected) global hypersurface of section for some Reeb dynamics on a $2n+1$-dimensional contact manifold (M, α), and τ is the associated return map. Let R_α be the Reeb vector field of α. Denote by $B = \partial\Sigma$, which we assume to be a contact submanifold of M with induced contact form $\alpha_B = \alpha|_B$, so that $R_\alpha|_B$ is tangent to B. Let $\lambda = \alpha|_\Sigma$, which is a Liouville form on $\mathrm{int}(\Sigma)$, since R_α is assumed to be positively transverse to the interior of Σ. That is, the two-form $\omega = d\lambda$ is symplectic on $\mathrm{int}(\Sigma)$. The 1-form $\lambda_B = \lambda|_B$ coincides with the contact form α_B. Note that it is degenerate along B. By Stokes' theorem, the symplectic volume of Σ then coincides with the contact volume of B

$$\mathrm{vol}(\Sigma, \omega) = \int_\Sigma \omega^n = \int_\Sigma \mathrm{d}(\lambda \wedge \mathrm{d}\lambda^{n-1}) = \int_B \alpha_B \wedge \mathrm{d}\alpha_B^{n-1} = \mathrm{vol}(B, \alpha_B).$$

Note that τ is automatically a symplectomorphism with respect to ω. Indeed, denote the time-t Reeb flow by φ_t, and let $T : \mathrm{int}(\Sigma) \to \mathbb{R}^+$

$$T(x) = \min\{t > 0 : \varphi_t(x) \in \mathrm{int}(\Sigma)\}$$

denote the first return time function. Then, $\tau(x) = \varphi_{T(x)}(x)$, and so, for $x \in \mathrm{int}(\Sigma)$, $v \in T_x\Sigma$, we have

$$d_x\tau(v) = d_xT(v)R_\alpha(\tau(x)) + d_x\varphi_{T(x)}(v).$$

Using that φ_t satisfies $\varphi_t^*\alpha = \alpha$, we obtain

$$\begin{aligned}(\tau^*\lambda)_x(v) &= \alpha_{\tau(x)}(d_x\tau(v)) \\ &= d_xT(v) + (\varphi_{T(x)}^*\alpha)_x(v) \\ &= d_xT(v) + \lambda_x(v).\end{aligned} \tag{C.7}$$

Therefore

$$\tau^*\lambda = \mathrm{d}T + \lambda, \tag{C.8}$$

which in particular implies that $\tau^*\omega = \omega$.

Moreover, the average of the return time function gives the contact volume of M, i.e., we have the identity

$$\int_\Sigma T\omega^n = \mathrm{vol}(M, \alpha). \tag{C.9}$$

This may be proved as follows. We have a smooth embedding

$$\psi\colon \mathbb{R}/\mathbb{Z} \times \mathrm{int}(\Sigma) \to M,$$

given by $\psi(s, x) = \varphi_{sT(x)}(x)$, which is a diffeomorphism onto $M\backslash B$. It satisfies

$$(\psi^*\alpha)(\partial_s) = \alpha(TR_\alpha) = T,$$

and, for $v \in T\mathrm{int}(\Sigma)$,

$$(\psi^*\alpha)(v) = \alpha(s\mathrm{d}T(v)R_\alpha + \mathrm{d}\varphi_{sT}(v)) = s\mathrm{d}T(v) + \alpha(v).$$

Then

$$\psi^*\alpha = T\mathrm{d}s + s\mathrm{d}T + \lambda = \mathrm{d}(sT) + \lambda,$$

and so

$$\psi^*(\alpha \wedge d\alpha^n) = (\mathrm{d}(sT) + \lambda) \wedge \mathrm{d}\lambda^n = T\mathrm{d}s \wedge \omega^n.$$

Integrating, and using the fact that B is codimension 2 in M, we obtain

$$\begin{aligned}\mathrm{vol}(M, \alpha) &= \int_{M\backslash B} \alpha \wedge \mathrm{d}\alpha^n = \int_{\mathbb{R}/\mathbb{Z}\times\mathrm{int}(\Sigma)} \psi^*(\alpha \wedge d\alpha^n) \\ &= \int_{\mathbb{R}/\mathbb{Z}\times\mathrm{int}(\Sigma)} T\mathrm{d}s \wedge \omega^n = \int_{\mathrm{int}(\Sigma)} T\omega^n = \int_\Sigma T\omega^n,\end{aligned}$$

where we have used that $\omega^n|_B \equiv 0$, and the claim follows. In case where τ is Hamiltonian, we want to relate the Hamiltonian action of a periodic orbit of τ to the Reeb action of the corresponding Reeb orbit in the ambient contact manifold.

Let $H : S^1 \times \Sigma \to \mathbb{R}^+$ be a Hamiltonian generating τ, i.e., the isotopy ϕ_t defined by $\phi_0 = id$, $\frac{\mathrm{d}}{\mathrm{d}t}\phi_t = X_{H_t} \circ \phi_t$ satisfies $\phi_1 = \tau$. The sign convention for the Hamiltonian vector field is $i_{X_{H_t}}\omega = -dH_t$. We usually view this Hamiltonian isotopy as defining an element $\phi = \phi_H = [\{\phi_t\}]$ in the universal cover $\widetilde{\mathrm{Diff}}(\Sigma, \omega)$ of the space of symplectomorphisms $\mathrm{Diff}(\Sigma, \omega)$. By Cartan's formula, we have

$$\partial_t\phi_t^*\lambda = \phi_t^*\mathcal{L}_{X_{H_t}}\lambda = \phi_t^*(i_{X_{H_t}}\omega + d(i_{X_{H_t}}\lambda)) = \phi_t^*d(i_{X_{H_t}}\lambda - H_t),$$

and so integrating, we obtain

$$\tau^*\lambda - \lambda = dF_H, \tag{C.10}$$

where

$$F_H = \int_0^1 (i_{X_{H_t}}\lambda - H_t) \circ \phi_t \, \mathrm{d}t. \tag{C.11}$$

Combining (C.8) and (C.10), we deduce that

$$\tau = F_H + C \tag{C.12}$$

for some constant C (assuming Σ is connected).

We determine the constant C under a suitable assumption, which we assume holds in all what follows. Namely, assume that τ extends to Σ with the same formula, i.e., via an extension of the return time function T to Σ. Assume also that $H_t|_B \equiv const := C_t > 0$ for some H generating τ. Equivalently, $X_{H_t}|_B = h_t R_B$ for some (not necessarily positive) smooth function h_t on B, satisfying $h_t = dH_t(V_\lambda)|_B$ where V_λ is the Liouville vector field associated with λ. In this case, denoting $\gamma_x(t) = \phi_t(x)$ for $x \in B$ and $t \in [0,1]$, we get

$$F_H(x) = \int_{\gamma_x} \lambda_B - \int_0^1 C_t \mathrm{d}t = \int_0^1 (h_t(\phi_t(x)) - C_t)\mathrm{d}t, \tag{C.13}$$

On the other hand, let $\beta_x(t) = \varphi_t(x)$ be the Reeb orbit through x ending at $\beta_x(1) = \tau(x)$, for $t \in [0,1]$, which we assume parametrized, so that $\dot{\beta}_x = T(x)R_B(\beta_x)$. Note that β_x is a reparametrization of γ_x, and so we obtain

$$\tau(x) = \int_{\beta_x} \lambda_B = \int_{\gamma_x} \lambda_B.$$

This means that T is the unique primitive of $\tau^*\lambda - \lambda$ satisfying $T(x) = \int_{\gamma_x} \lambda_B$ for $x \in B$. Combining (C.12) and (C.13), we conclude that

$$C = \int_0^1 C_t \mathrm{d}t > 0,$$

a positive constant.

By the above computation, T is what is usually called the action of $\phi = \phi_H$ with respect to λ, and is independent of the isotopy class (with fixed endpoints) of the path ϕ_H. The Calabi invariant is then by definition the average action $CAL(\phi_H, \omega) = \int T\omega^n$, which is independent of λ; cf. [2,28]. Combining with (C.9), we obtain

$$CAL(\phi_H, \omega) = \mathrm{vol}(M, \alpha).$$

Let $\gamma : S^1 = \mathbb{R}/k\mathbb{Z} \to \Sigma$, defined by $\gamma(t) = \phi_t(x)$, be a k-periodic Hamiltonian orbit associated to the k-periodic point x of τ. That is, we have $x = \gamma(0)$, $\gamma(1) = \tau(x), \ldots, \gamma(k) = \tau^k(x) = x$, and assume that k is the minimal period of x. We then get

$$\sum_{i=1}^k F_H(\tau^i(x)) = \mathcal{A}_{H^{\#k}}(\gamma)$$

is precisely the Hamiltonian action of γ with respect to the Hamiltonian

$$H_t^{\#k} = \sum_{i=1}^{k} H_t \circ \phi_t^{-i}$$

generating τ^k. If $\beta : S^1 = \mathbb{R}/\mathbb{Z} \to M$ is the Reeb orbit corresponding to γ, (C.12) implies that its period is

$$\int_{S^1} \beta^*\alpha = \sum_{i=1}^{k} T(\tau^i(x)) = \sum_{i=1}^{k} F_H(\tau^i(x)) + kC = \mathcal{A}_{H^{\#k}}(\gamma) + kC.$$

Since $C > 0$, this implies the following: if the Hamiltonian action of every k-periodic orbit γ grows to infinity with k, then the period of the associated Reeb orbits β also. In other words, long Hamiltonian periodic orbits in the global hypersurface of section give long Reeb orbits in the ambient contact manifold.

We summarize the above discussion in the following:

Lemma C.1. *Let (M^{2n+1}, α) be a contact manifold, $(\Sigma^{2n}, \omega = d\alpha|_\Sigma)$ a Liouville domain which is a global hypersurface of section for the Reeb flow, $(B^{2n-1}, \alpha_B) = (\partial\Sigma, \alpha|_B)$, $\tau : int(\Sigma) \to int(\Sigma)$ the Poincaré return map, and $T : int(\Sigma) \to \mathbb{R}^+$ the first return time. Then:*

1. *$vol(\Sigma, \omega) = vol(B, \alpha_B)$.*
2. *$vol(M, \alpha) = \int_\Sigma T\omega^n$.*
3. *τ is an exact symplectomorphism.*
4. *If τ is Hamiltonian with generating isotopy $\phi_H = [\{\phi_t\}] \in \widetilde{Diff}(\Sigma, \omega)$, and extends to Σ as a (not necessarily positive) reparametrization of the Reeb flow at B, then:*
   - (i) *$CAL(\phi_H, \omega) = vol(M, \alpha)$.*
   - (ii) *The period of a Reeb orbit β on M corresponding to a k-periodic Hamiltonian orbit γ on Σ is*

$$\int_{S^1} \beta^*\alpha = \mathcal{A}_{H^{\#k}}(\gamma) + kC$$

for some positive constant $C > 0$, where

$$\mathcal{A}_{H^{\#k}}(\gamma) = \int_{S^1} \gamma^*\lambda - \int_0^1 H_t^{\#k}(\gamma(t))\,dt$$

is the Hamiltonian action of γ with respect to the Hamiltonian

$$H_t^{\#k} = \sum_{i=1}^{k} H_t \circ \phi_t^{-i}$$

generating τ^k. In particular, if γ has large action, then β has large period.

Appendix D: Strong convexity implies strong index-positivity

In this appendix, we give a general condition for index-positivity to hold, which is also relevant for the restricted three-body problem. A connected compact hypersurface $\Sigma \subset \mathbb{R}^4$ is said to *bound a strongly convex domain* $W \subset \mathbb{R}^4$ whenever there exists a smooth function $\phi : \mathbb{R}^4 \to \mathbb{R}$ satisfying:

(i) (Regularity) $\Sigma = \{\phi = 0\}$ is a regular level set;
(ii) (Bounded domain) $W = \{z \in \mathbb{R}^4 : \phi(z) \leq 0\}$ is bounded and contains the origin; and
(iii) (Positive-definite Hessian) $\nabla^2 \phi_z(h, h) > 0$ for $z \in W$ and for each non-zero tangent vector $h \in T\Sigma$.

In this case, the radial vector field is transverse to Σ, and so Σ is a contact-type 3-sphere, inheriting a contact form α induced by the standard Liouville form in $\mathbb{R}^4$.

Lemma D.1. *Suppose that Σ bounds a strongly convex domain. Then, Σ is strongly index-positive.*

Remark D.2. In the planar restricted three-body problem, the values of energy/mass ratio (c, μ) for which the Levi–Civita regularization bounds a strictly convex domain is called the *convexity range*, which in particular implies that the dynamics is *dynamically convex* (cf. [5,6,24]). It follows that index-positivity holds in the convexity range for the quotient $\mathbb{R}P^3$, which is part of the assumptions of Theorem A.

Proof. Write $\Sigma = \phi^{-1}(0)$ as in the definition above. Denote the contact form on Σ by $\alpha := \lambda|_\Sigma$. We will use the standard quaternions I, J, K, where I is chosen to coincide with the standard complex structure.

The tangent space of Σ is spanned by the vectors

$$R = X_\phi/\alpha(X_\phi) = I\nabla\phi/\alpha(X_\phi) = Iw,\ U = Jw - \alpha(Jv)R,\ V = Kw - \alpha(Kv)R.$$

We note that U and V give a symplectic trivialization ϵ of $(\xi = \ker \alpha, d\alpha)$. To see this, we compute

$$d\alpha(U, V) = d\alpha(Jw, Kw) = w^t J^t I^t K w = w^t K^t K w = w^t w = 1.$$

To prove the claim, we investigate the rate of change of a version of the rotation number. See Chapter 10.6 in [16] for a detailed standard description of the Robbin–Salamon index in terms of the rotation number. We will detail the version that we will use below.

We look at the linearization of the Hamiltonian flow:

$$\dot{X} = \nabla_X X_\phi = I\nabla^2\phi \cdot X. \tag{D.14}$$

Starting with $X(0) \in \xi$, we compute how quickly the vector X rotates with respect to the frame. Define the angular form

$$\Theta = \frac{u\mathrm{d}v - v\mathrm{d}u}{u^2 + v^2} = \frac{ud\alpha(U, \cdot) + vd\alpha(V, \cdot)}{u^2 + v^2} = \frac{d\alpha(uU + vV, \cdot)}{u^2 + v^2},$$

where (u, v) are cartesian coordinates on the plane spanned by the frame (U, V), so we may write $X = uU + vV$. We plug in $\dot{X}$ and find

$$\Theta(\dot{X}) = \frac{d\alpha(X, \dot{X})}{u^2+v^2} = \frac{(uU+vV)^t I^t I \nabla^2\phi \cdot (uU+vV)}{u^2+v^2}$$
$$= \frac{\nabla^2\phi(uU+vV, uU+vV)}{u^2+v^2} \geq \lambda_{\min} > 0, \tag{D.15}$$

where $\lambda_{\min}$ is the minimal eigenvalue of $\nabla^2\phi$ over the compact hypersurface Σ. After we have set up some notation, we will see that this is enough to get a lower bound on the growth rate of the Robbin–Salamon index. With our global trivialization ϵ, we can define the matrix

$$\psi(t) = \epsilon \circ dFl_t^R \circ \epsilon^{-1}.$$

By applying Eq. (D.14) to the initial vectors $\epsilon^{-1}(1, 0)$ and $\epsilon^{-1}(0, 1)$, we get a linear evolution equation for the matrix $\psi(t)$

$$\dot{\psi} = A(t)\psi, \tag{D.16}$$

where A is a time-dependent matrix. We will view this ODE as a vector field on $Sp(2)$: the linearized Reeb flow along each Reeb orbit will give rise to such a vector field.

To relate the above angle to the Conley–Zehnder index, we also need to recall the Iwasawa decomposition, also known as KAN decomposition, of $Sp(2)$. Write

$$KAN := \left\{ \left(\begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix}, \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}, \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \right) \;\middle|\; \phi \in [0, 2\pi), a \in \mathbb{R}_{>0}, t \in \mathbb{R} \right\}.$$

And put

$$\mathrm{kan} : KAN \longrightarrow Sp(2), (\phi, a, t) \longmapsto \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

This map has the inverse

$$\mathrm{kan}^{-1} : Sp(2) \longrightarrow KAN$$
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \longmapsto \left(\frac{1}{\sqrt{a^2+c^2}} \begin{pmatrix} a & -c \\ c & a \end{pmatrix}, \begin{pmatrix} \sqrt{a^2+c^2} & 0 \\ 0 & \frac{1}{\sqrt{a^2+c^2}} \end{pmatrix}, \begin{pmatrix} 1 & \frac{ab+cd}{\sqrt{a^2+c^2}} \\ 0 & 1 \end{pmatrix} \right).$$

The KAN angle can locally be determined as

$$\arg(\mathrm{kan}^{-1}(\psi)) = \mathrm{atan}(c/a),$$

so we see that the change in angle equals

$$\frac{\mathrm{d}}{\mathrm{d}t} \arg(\mathrm{kan}^{-1}(\psi)) = \frac{\mathrm{d}}{\mathrm{d}t} \mathrm{atan}(c/a) = \frac{a\dot{c} - c\dot{a}}{a^2+c^2}.$$

On the other hand, the rate of change of the KAN angle equals $\Theta(\dot{X})$. Indeed, the first column of $\psi(t)$ is the vector $Z(t) := \epsilon(X(t)) = \begin{pmatrix} u \\ v \end{pmatrix}$ if we put $X(0) = \epsilon^{-1}(1, 0)$.

By Eq. (D.15), this rate of change is at least $\lambda_{\min}$, where $\lambda_{\min}$ is the minimal eigenvalue of $-IA$ (which we assume to be positive-definite). This means that each slice

$$S_\phi = \text{kan}\left(\left\{\left(\begin{pmatrix}\cos(\phi) & -\sin(\phi)\\ \sin(\phi) & \cos(\phi)\end{pmatrix}, \begin{pmatrix}a & 0\\ 0 & a^{-1}\end{pmatrix}, \begin{pmatrix}1 & t\\ 0 & 1\end{pmatrix}\right) \;\middle|\; a \in \mathbb{R}_{>0}, t \in \mathbb{R}\right\}\right)$$

is a global surface of section for the vector field associated with Eq. (D.16): the maximal return time is $\frac{2\pi}{\lambda_{\min}}$. Now, take a matrix $\psi(0)$ in the slice S_0, and let $\psi(t)$ denote the solution to Eq. (D.16).

Claim: Each crossing is regular and contributes positively.

To see this, recall that the crossing form of a path ψ in $Sp(2)$ at a crossing t is defined as the bilinear form

$$\omega_0(\cdot, \dot\psi(t)\cdot)|_{\ker(\psi(t)-\mathrm{id})}.$$

Since U, V is a symplectic frame, we have with $Z = (u, v)$ (i.e., $X = uU + vV$), the following inequality:

$$\omega_0(Z, \dot\psi Z) = d\alpha(X, \dot X) \geq \lambda_{\min}(u^2 + v^2),$$

by Eq. (D.15). This establishes the claim.

Let t_ℓ denote the ℓ-th return time to S_0 of the path ψ. The symplectic path $\psi|_{[0,t_\ell]}$ is not necessarily a loop, but we can make it into a loop by connecting $\psi(t_\ell)$ to $\psi(0)$ while staying in the slice S_0. We can and will do this by adding at most one crossing, which we make regular. Call the extension to a loop $\tilde\psi$. The additional crossing that we may have inserted can contribute negatively.

Now, use the loop axiom for the Robbin–Salamon index. This tells us that

$$\mu_{\mathrm{RS}}(\tilde\psi) = 2\mu_\ell(\tilde\psi) = 2\ell.$$

By the catenation property of the Robbin–Salamon index and positivity of all but the last (potential) crossing, we have

$$\mu_{\mathrm{RS}}(\psi) \geq 2\ell - 2.$$

Now, consider a symplectic path ψ of length T. We can bound the winding number as

$$\ell \geq \left\lfloor \frac{\lambda_{\min}}{2\pi} T \right\rfloor.$$

With this in mind, we obtain for a Hamiltonian arc γ of length T

$$\mu_{\mathrm{RS}}(\gamma; \epsilon) \geq \frac{2\lambda_{\min}}{2\pi} T - 4.$$

When this Hamiltonian arc γ is viewed as a Reeb arc γ_R with Reeb action T_R, we can rewrite this bound as follows, using:

$$T_R = \int_0^T \alpha(X_\phi)\mathrm{d}t \leq T \cdot \max \alpha(X_\phi).$$

We find

$$\mu_{\mathrm{RS}}(\gamma_R;\epsilon) \geq \frac{\lambda_{\min}}{\pi \max \alpha(X_\phi)} T_R - 4.$$

$\square$

Remark D.3. Observe that the proof actually shows the stronger claim that index-positivity holds when the Hessian of ϕ restricted to the contact structure is positive-definite. Note also that the latter condition is not enough for dynamical convexity.

Finally, we note that the bound obtained can be sharpened, since the index is necessarily positive by observing that $\psi(0) = \mathrm{id}$, so it is a crossing and using that each crossing of the path ψ contributes positively.

Appendix E: Strongly index-definite symplectic paths

In this appendix, we prove a crucial index growth estimate needed to rule out non-relevant boundary orbits via index considerations (needed in Lemma 4.5 in the main body of the paper).

Definition E.1. Consider the linear ODE $\dot{\psi}(t) = A(t)\psi(t)$, where $A : \mathbb{R}_{\geq 0} \to \mathfrak{sp}(2n)$ and $A(0) = 0$. Its solution is a path of symplectic matrices with $\psi(0) = \mathbb{1}$. We say that the ODE is *strongly index-definite* if there exist constants $c > 0, d \in \mathbb{R}$, such that

$$|\mu_{\mathrm{RS}}(\psi|_{[0,t]})| \geq ct + d,$$

where μ_{RS} is the Robbin–Salamon index [33].

Note that we make no non-degeneracy assumptions on the symplectic paths in the above definition.

We now consider the specific family of linear ODEs $\dot{\psi}(t) = A(t)\psi(t)$, where the matrix A has the special form

$$A(t) = \left(\begin{array}{cc|cc} \multicolumn{2}{c|}{\multirow{2}{*}{$R(t)$}} & \overline{X}(t) & 0 \\ & & \overline{Y}(t) & 0 \\ \hline 0 & 0 & a(t) & 0 \\ \overline{Y}(t) & -\overline{X}(t) & b(t) & -a(t) \end{array}\right) \in \mathfrak{sp}(2n).$$

Here, we use the notation $(\overline{X}, \overline{Y}) = (X_1, Y_1, \ldots, X_{n-1}, Y_{n-1})$, and we assume $R(t) \in \mathfrak{sp}(2n-2)$, $A(0) = 0$.

Lemma E.2. *Assume that the linear ODE $\dot{M}(t) = R(t)M(t)$ is strongly index-definite as an ODE in dimension $2n-2$. Then, the same holds for the linear ODE $\dot{\psi}(t) = A(t)\psi(t)$.*

Proof. One may check that

$$\mathfrak{g} = \left\{ \left(\begin{array}{cc|cc} \multicolumn{2}{c|}{\multirow{2}{*}{R}} & \overline{X} & 0 \\ & & \overline{Y} & 0 \\ \hline 0 & 0 & a & 0 \\ \overline{Y} & -\overline{X} & b & -a \end{array}\right) : R \in \mathfrak{sp}(2n-2) \right\}$$

is a Lie subalgebra of $\mathfrak{sp}(2n)$. The corresponding Lie subgroup of $Sp(2n)$ is

$$G=\left\{\left(\begin{array}{c|c} M & \begin{array}{cc}\overline{x} & 0\\ \overline{y} & 0\end{array}\\ \hline \begin{array}{cc}0 & 0\\ \overline{u} & \overline{v}\end{array} & \begin{array}{cc}\alpha & 0\\ \beta & \alpha^{-1}\end{array}\end{array}\right) : M\in Sp(2n-2),\ \alpha>0,\ (-\overline{y},\overline{x})\cdot M+\alpha\cdot(\overline{u},\overline{v})=0\right\}.$$

We deduce that $\psi \in G$. We then write

$$\psi=\left(\begin{array}{c|c} M & \begin{array}{cc}\overline{x} & 0\\ \overline{y} & 0\end{array}\\ \hline \begin{array}{cc}0 & 0\\ \overline{u} & \overline{v}\end{array} & \begin{array}{cc}\alpha & 0\\ \beta & \alpha^{-1}\end{array}\end{array}\right)\in G,$$

where M is a solution to $\dot{M}=RM$, and consider the following homotopy of paths:

$$\psi_s=\left(\begin{array}{c|c} M & \begin{array}{cc}s\overline{x} & 0\\ s\overline{y} & 0\end{array}\\ \hline \begin{array}{cc}0 & 0\\ s\overline{u} & s\overline{v}\end{array} & \begin{array}{cc}\alpha & 0\\ \beta & \alpha^{-1}\end{array}\end{array}\right).$$

Note that ψ_s is a path in $G\subset Sp(2n)$ for every s, and ψ_0 has no off-diagonal terms. For any given t, this gives a homotopy in G relative endpoints of $\psi|_{[0,t]}$ to a concatenated path of the form $\psi_0|_{[0,t]}\#\phi_t$, where $\phi_t(s)=\psi_s(t)$. We therefore have

$$\mu_{\mathrm{RS}}(\psi|_{[0,t]})=\mu_{\mathrm{RS}}(\psi_0|_{[0,t]})+\mu_{\mathrm{RS}}(\phi_t). \tag{E.17}$$

On the other hand, from the block decomposition of ψ_0 and the fact that the lower block can be homotoped to a symplectic shear by joining $\alpha(t)$ to 1, we have

$$\mu_{\mathrm{RS}}(\psi_0|_{[0,t]})=\mu_{\mathrm{RS}}(M|_{[0,t]})\pm\frac{1}{2}\mathrm{sign}(\beta(t)), \tag{E.18}$$

where the sign depends on conventions. Moreover, one may easily check that the characteristic polynomial of an element in G is completely independent of the off-diagonal terms. In particular, we obtain that

$$\det(\psi_s-\mathbb{1})=\det(\psi_0-\mathbb{1})=\det(M-\mathbb{1})(\alpha-1)(\alpha^{-1}-1)$$

is independent of s. In other words, $\psi(t)$ is an intersection point with the Maslov cycle if and only if $\psi_0(t)$ is, and the eigenvalue 1 has the same algebraic multiplicity for both such intersections. Moreover, if $\psi(t)$ is not an intersection, then ϕ_t does not intersect the Maslov cycle at all.

One may check that if $\alpha(t)\neq 1$, then the geometric multiplicity of 1 as an eigenvalue of $\phi_t(s)$ is independent of s (and, therefore, $\mu_{\mathrm{RS}}(\phi_t)=0$ for such t). If $\alpha(t)=1$, this may not necessarily still hold. However, we may appeal to the following general fact, whose proof was provided to the authors by Alberto Abbondandolo:

Lemma E.3. *There exists a universal bound $C=C(n)$ (depending only on dimension), such that, if $\phi:[0,1]\to Sp(2n)$ is a continuous path of symplectic* *matrices for which the algebraic multiplicity of the eigenvalue* 1 *of the matrix* $\phi(t)$ *is independent of* t*, then*

$$|\mu_{RS}(\phi)| \leq C.$$

Proof of Lemma E.3. **Step 1.** We first reduce to the case where ϕ has 1 as the only eigenvalue. We have a continuous symplectic splitting $\mathbb{R}^{2n} = V(t) \oplus W(t)$ where $V(t)$ is the generalized eigenspace of $\phi(t)$ corresponding to 1, and $W(t)$ is the direct sum of the generalized eigenspaces of $\phi(t)$ corresponding to the other eigenvalues (here, the dimensions of $V(t)$ and $W(t)$ are t-independent by assumption), for which $\phi(t) = \phi_V(t) \oplus \phi_W(t)$ splits symplectically. Since ϕ_W does not intersect the Maslov cycle by construction, we have $\mu_{\mathrm{RS}}(\phi) = \mu_{\mathrm{RS}}(\phi_V) + \mu_{\mathrm{RS}}(\phi_W) = \mu_{\mathrm{RS}}(\phi_V)$.

Step 2. A loop ϕ of symplectic matrices having 1 as the only eigenvalue is nullhomotopic in $Sp(2n)$, and hence, $\mu_{\mathrm{RS}}(\phi) = 0$. This follows for instance by the interpretation of the Robbin–Salamon index as the total winding number of the Krein-positive eigenvalues on the unit circle (see, e.g., [1, Lemma 1.3.7]).

Step 3. The identity matrix may be joined to any symplectic matrix M satisfying $\mathrm{spec}(M) = \{1\}$ via a path $M(t)$ satisfying $\mathrm{spec}(M(t)) = \{1\}$, and for which $|\mu_{\mathrm{RS}}(M(t))| \leq C$ for some universal bound C. Indeed, we may write $M = e^{JS}$ where S is a symmetric matrix having 0 as the only eigenvalue, and consider the path $M(t) = e^{tJS}$. This satisfies the required properties, since $M(t)$ changes strata of the Maslov cycle only at $t = 0$, the geometric multiplicity of 1 jumping from $2n$ at $t = 0$ to perhaps a lower one at $t > 0$, and so the contribution of this wall-crossing to $\mu_{\mathrm{RS}}(M)$ is universally bounded.

The proof finishes by combining the previous steps, where we join the endpoints of a path ϕ as in Step 1 to the identity as in Step 3, use the concatenation property of μ_{RS}, and appeal to Step 2. □

Combining Eqs. (E.17) and (E.18) with Lemma E.3, we conclude that

$$|\mu_{\mathrm{RS}}(\psi|_{[0,t]}) - \mu_{\mathrm{RS}}(M|_{[0,t]})| \leq C$$

for some universal constant $C = C(n)$, from which the conclusion of Lemma E.2 is immediate. □

References

[1] Abbondandolo, A. Morse theory for Hamiltonian systems. Chapman & Hall/CRC Research Notes in Mathematics, 425. Chapman & Hall/CRC, Boca Raton, FL, (2001). xii+189 pp. ISBN: 1-58488-202-6

[2] Abbondandolo, A., Bramham, B., Hryniewicz, U.L., Salomão, P.A.S.: Sharp systolic inequalities for Reeb flows on the three-sphere. Invent. Math. 211(2), 687–778 (2018)

[3] Abbondandolo, A., Schwarz, M.: On the Floer homology of cotangent bundles. Commun. Pure Appl. Math. **59**(2), 254–316 (2006)

[4] Abouzaid, M.: Symplectic cohomology and Viterbo's theorem Free loop spaces in geometry and topology, 271–485, IRMA Lect. Math. Theor. Phys., 24, Eur. Math. Soc., Zürich, (2015)

[5] Albers, P., Fish, J.W., Frauenfelder, U., van Koert, O.: The Conley-Zehnder indices of the rotating Kepler problem. Math. Proc. Cambridge Philos. Soc. **154**(2), 243–260 (2013)

[6] Albers, P., Fish, J.W., Frauenfelder, U., Hofer, H., van Koert, O.: Global surfaces of section in the planar restricted 3-body problem. Arch. Ration. Mech. Anal. **204**(1), 273–284 (2012)

[7] Audin, M., Damian, M.M.: Theory and Floer homology. Translated from the,: French original by Reinie Erné, p. 2014. Universitext, Springer, London; EDP Sciences, Les Ulis (2010)

[8] Birkhoff, G.D.: Proof of Poincaré's last geometric theorem. Trans. AMS **14**, 14–22 (1913)

[9] Benedetti, G., Ritter, A.F.: Invariance of symplectic cohomology and twisted cotangent bundles over surfaces. Int. J. Math. **31**(9), 56 (2020)

[10] Bourgeois, F.: Introduction to Contact Homology, lecture notes for a Summer School in Berder (2003)

[11] Cheeger, J., Ebin, D.G.: Comparison theorems in Riemannian geometry. Revised reprint of the 1975 original. AMS Chelsea Publishing, Providence, RI, (2008). x+168 pp. ISBN: 978-0-8218-4417-5

[12] Cieliebak, K., Floer, A., Hofer, H., Wysocki, K.: Applications of symplectic homology. II. Stability of the action spectrum. Math. Z. **223**(1), 27–45 (1996)

[13] Conley, C.C.: On some new long periodic solutions of the plane restricted three body problem. (1963) Internat. Sympos. Monlinear Differential Equations and Nonlinear Mechanics pp. 86–90 Academic Press, New York

[14] Fauck, A.: On manifolds with infinitely many fillable contact structures. Int. J. Math. **31**(13), 71 (2020)

[15] Floer, A., Hofer, H.: Symplectic homology. I. Open sets in $\mathbb{C}^n$. Math. Z. **215**(1), 37–88 (1994)

[16] Frauenfelder, U., van Koert, O.: The restricted three-body problem and holomorphic curves Pathways in Mathematics. Birkhäuser/Springer, Cham, (2018). xi+374 pp. ISBN: 978-3-319-72277-1

[17] Ginzburg, V.L.: The Conley Conjecture. Ann. Math. **172**, 1127–1180 (2010)

[18] Ginzburg, V.L., Gürel, B.Z.: The Conley conjecture and beyond. Arnold Math. J. **1**(3), 299–337 (2015)

[19] Gromov, M.: Homotopical effects of dilatation. J. Diff. Geom. **13**(3), 303–310 (1978)

[20] Harris, A., Paternain, G.P.: Dynamically convex Finsler metrics and J-holomorphic embedding of asymptotic cylinders. Ann. Global Anal. Geom. **34**(2), 115–134 (2008)

[21] Hein, D.: The Conley conjecture for the cotangent bundle. Arch. Math. (Basel) **96**(1), 85–100 (2011)

[22] Hill, G.W.: Researches in the Lunar Theory. Am. J. Math. **1**(1), 5–26 (1878)

[23] Hofer, H.: Symplectic capacities Geometry of low-dimensional manifolds, 2 (Durham, 1989), 15–34, London Math. Soc. Lecture Note Ser., 151, Cambridge Univ. Press, Cambridge, (1990)

[24] Hofer, H., Wysocki, K., Zehnder, E.: The dynamics on three-dimensional strictly convex energy surfaces. Ann. Math. (2) **148**(1), 197–289 (1998)

[25] Katok, A.B.: Ergodic perturbations of degenerate integrable Hamiltonian systems. (Russian) Izv. Akad. Nauk SSSR Ser. Mat. 37, 539–576 (1973)

[26] Katok, A., Hasselblatt, B.: Introduction to the modern theory of dynamical systems. With a supplementary chapter by Katok and Leonardo Mendoza. Encyclopedia of Mathematics and its Applications, 54. Cambridge University Press, Cambridge, (1995). xviii+802 pp. ISBN: 0-521-34187-6

[27] Kwon, M., van Koert, O.: Brieskorn manifolds in contact topology. Bull. Lond. Math. Soc. **48**(2), 173–241 (2016)

[28] McDuff, D., Salamon, D.: Introduction to symplectic topology. Third edition. Oxford Graduate Texts in Mathematics. Oxford University Press, Oxford, (2017). xi+623 pp. ISBN: 978-0-19-879490-5; 978-0-19-879489-9

[29] Moser, J.: Monotone twist mappings and the calculus of variations. Ergodic Theory Dyn. Syst. **6**(3), 401–413 (1986)

[30] Moreno, A., van Koert, Otto: Global hypersurfaces of section in the spatial restricted three-body problem. (2020). arXiv:2011.10386

[31] Poincaré, H.: Les Méthodes Nouvelles de la Mécanique Céleste, Tome I, Paris, Gauthier-Viltars, 1892. Republished by Blanchard, Paris (1987)

[32] Poincaré, H.: Sur un théorème de géométrie. Rend. Circ. Mat. Palermo **33**, 375–407 (1912)

[33] Robbin, J., Salamon, D.: The Maslov index for paths. Topology **32**(4), 827–844 (1993)

[34] Salamon, D.A., Weber, J.: Floer homology and the heat flow. Geom. Funct. Anal. **16**(5), 1050–1138 (2006)

[35] Salamon, D., Zehnder, E.: Morse theory for periodic solutions of Hamiltonian systems and the Maslov index. Commun. Pure Appl. Math. **45**(10), 1303–1360 (1992)

[36] Seeley, R.T.: Extension of C^∞ functions defined in a half space Proc. Am. Math. Soc. **15**, 625–626 (1964)

[37] Ustilovsky, I.: Contact homology and contact structures on S^{4m+1}. PhD thesis, Stanford University, (1999)

[38] Viterbo, C.: Functors and computations in Floer homology with applications I. Geom. Funct. Anal. **9**(5), 985–1033 (1999)

[39] Viterbo, C.: Functors and computations in Floer homology with applications Part II. arXiv:1805.01316

Agustin Moreno
School of Mathematics
Institute for Advanced Study
Princeton NJ
USA
e-mail: agustin.moreno2191@gmail.com

Otto van Koert
Department of Mathematical Sciences
Seoul National University
Seoul
South Korea
e-mail: okoert@snu.ac.kr

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:52
https://doi.org/10.1007/s11784-022-00967-4
Published online June 15, 2022

Journal of Fixed Point Theory and Applications

Covariant constancy of quantum Steenrod operations

Paul Seidel and Nicholas Wilkins

Abstract. We prove a relationship between quantum Steenrod operations and the quantum connection. In particular, there are operations extending the quantum Steenrod power operations that, when viewed as endomorphisms of equivariant quantum cohomology, are covariantly constant. We demonstrate how this property is used in computations of examples.

Mathematics Subject Classification. 53D45, 14N35, 55S10, 55N91, 57R91.

Keywords. Gromov–Witten invariants, quantum cohomology, quantum Steenrod powers.

1. Introduction

Quantum Steenrod operations, originally introduced by Fukaya [8], have recently appeared in a variety of contexts: their properties have been explored in [22] (which also contains the first nontrivial computations); they can be used to study arithmetic aspects of mirror symmetry [18]; and in Hamiltonian dynamics, they are relevant for the existence of pseudo-rotations [3, 20, 21]. Nevertheless, computing quantum Steenrod operations remains a challenging problem in all but the simplest cases. Using methods similar to [22], this paper establishes a relation between quantum Steenrod operations and the quantum connection. As a consequence, the contribution of rational curves of low degree (very roughly speaking, of degree $< p$ if one is interested in quantum Steenrod operations with $\mathbb{F}_p$-coefficients) can be computed using only ordinary Steenrod operations and Gromov–Witten invariants. This is consonant with other indications that the geometrically most interesting part of quantum Steenrod operations may come from p-fold covered curves. Even though our method does not reach that part, it yields interesting results in many examples (some are carried out here, and there are more in [18]).

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

1a. Throughout this paper, M is a closed symplectic manifold which is weakly monotone [9] (in [14, Definition 6.4.1], this is called semi-positive). Fix an arbitrary coefficient field $\mathbb{F}$. The associated Novikov ring Λ is the ring of series

$$\gamma = \textstyle\sum_A c_A q^A, \tag{1.1}$$

where the exponents are $A \in H_2^{\mathrm{sphere}}(M;\mathbb{Z}) = \mathrm{im}(\pi_2(M) \to H_2(M;\mathbb{Z}))$, such that either $A = 0$ or $\int_A \omega_M > 0$; and among those A, such that $\int_A \omega_M$ is bounded by a given constant, only finitely many c_A may be nonzero. We think of this as a graded ring, where $|q^A| = 2c_1(A)$ (the notation being that $c_1(A)$ is the pairing between $c_1(M)$ and A). Write $I_{\max} \subset \Lambda$ for the ideal generated by q^A for nonzero A, so that $\Lambda/I_{\max} = \mathbb{F}$.

For each $a \in H^2(M;\mathbb{Z})$, there is an $\mathbb{F}$-linear differentiation operation $\partial_a : \Lambda \to \Lambda$

$$\partial_a q^A = (a \cdot A)\, q^A. \tag{1.2}$$

Write $I_{\mathrm{diff}} \subset I_{\max}$ for the ideal generated by q^A, where $A \neq 0$ lies in the kernel of the map $H_2^{\mathrm{sphere}}(M;\mathbb{Z}) \hookrightarrow H_2(M;\mathbb{Z}) \twoheadrightarrow \mathrm{Hom}(H^2(M;\mathbb{Z}),\mathbb{F})$. In other words, the generators are precisely those nontrivial monomials whose derivatives (1.2) are zero. (If $\mathbb{F}$ is of characteristic zero and $H_*^{\mathrm{sphere}}(M;\mathbb{Z})$ is torsion-free, then $I_{\mathrm{diff}} = 0$; but that is not the case we will be interested in.)

Remark 1.1. Clearly, ∂_a only depends on $a \otimes 1 \in H^2(M;\mathbb{Z}) \otimes \mathbb{F}$. One could define such operations for all elements in $H^2(M;\mathbb{F})$, and prove a version of our results in that context. We have refrained from doing so, since it adds a technical wrinkle (having to represent classes in $H^2(M;\mathbb{F})$ geometrically) without giving any striking additional applications.

1b. We will exclusively consider genus zero Gromov–Witten invariants. The three-pointed Gromov–Witten invariant in a class $A \in H_2^{\mathrm{sphere}}(M;\mathbb{Z})$ can be written as a bilinear operation

$$\begin{aligned} &*_A : H^*(M;\mathbb{F})^{\otimes 2} \longrightarrow H^{*-2c_1(A)}(M;\mathbb{F}), \\ &\int_M (c_1 *_A c_2)\, c_3 = \langle c_1, c_2, c_3 \rangle_A. \end{aligned} \tag{1.3}$$

One extends this to $H^*(M;\Lambda)$, and then packages all the $*_A$ into the small quantum product

$$\gamma_1 * \gamma_2 = \sum_A (\gamma_1 *_A \gamma_2)\, q^A. \tag{1.4}$$

Let t be another formal variable, of degree 2. The quantum connection on $H^*(M;\Lambda)[[t]]$ consists of the operations

$$\nabla_a \gamma = t\partial_a \gamma + a * \gamma, \tag{1.5}$$

where $*$ has been extended t-linearly. By the divisor axiom in Gromov–Witten theory, we have that for any $a_1, a_2 \in H^2(M;\mathbb{Z})$ and $c_1, c_2 \in H^*(M;\mathbb{F})$

$$(a_1 \cdot A) \int_M (a_2 *_A c_1)\, c_2 = \langle a_1, a_2, c_1, c_2 \rangle_A = (a_2 \cdot A) \int_M (a_1 *_A c_1)\, c_2. \tag{1.6}$$

This implies that the operations (1.5) for different a commute: the connection is flat.

We will consider endomorphisms Σ of $H^*(M;\Lambda)[[t]]$ which are $\Lambda[[t]]$-linear and covariantly constant, which means that they satisfy

$$\nabla_a \Sigma - \Sigma \nabla_a = 0. \tag{1.7}$$

This is a system of linear first-order differential equations. By looking at the equations for each q^A coefficient of Σ, one sees that:

Lemma 1.2. *For covariantly constant endomorphisms, the constant term determines the behaviour modulo I_{diff}. More formally, if Σ satisfies (1.7), then we have*

$$\Sigma \in \mathrm{End}(H^*(M;\mathbb{F})) \otimes I_{\max}[[t]] \implies \Sigma \in \mathrm{End}(H^*(M;\mathbb{F})) \otimes I_{\mathrm{diff}}[[t]]. \tag{1.8}$$

1c. From now on, we restrict to coefficient fields $\mathbb{F} = \mathbb{F}_p$, for a prime p. Our arguments involve $(\mathbb{Z}/p)$-equivariant cohomology with $\mathbb{F}_p$-coefficients. For a point, that is

$$H^*_{\mathbb{Z}/p}(point;\mathbb{F}_p) = H^*(B\mathbb{Z}/p;\mathbb{F}_p) = \mathbb{F}_p[[t,\theta]], \;\; |t| = 2,\, |\theta| = 1. \tag{1.9}$$

The notation requires some explanation. For $p = 2$, we have $\theta^2 = t$, so $\mathbb{F}_2[[t,\theta]]$ is actually a ring of power series in a single variable θ. For $p > 2$, we have $t\theta = \theta t$ and $\theta^2 = 0$, so that $\mathbb{F}_p[[t,\theta]]$ is a ring of power series in two supercommuting variables.

For any $A \in H_2^{\mathrm{sphere}}(M;\mathbb{Z})$ and any class $b \in H^*(M;\mathbb{F}_p)$, one can use $(\mathbb{Z}/p)$-equivariant Gromov–Witten theory to define an operation

$$Q\Sigma_{b,A} : H^*(M;\mathbb{F}_p) \longrightarrow (H^*(M;\mathbb{F}_p)[[t,\theta]])^{*+p|b|-2c_1(A)}. \tag{1.10}$$

For the trivial class $A = 0$, this is a form of the classical Steenrod operation St, more precisely

$$Q\Sigma_{b,0}(c) = \mathrm{St}(b)c. \tag{1.11}$$

Remark 1.3. Our notational and sign conventions follow [18] (except that we suppress the prime p), which differ from the classical conventions for Steenrod operations. In particular, for $p > 2$

$$\mathrm{St}(b) = (-1)^{\frac{|b|(|b|-1)}{2}\frac{p-1}{2}} \left(\tfrac{p-1}{2}!\right)^{|b|} t^{\frac{p-1}{2}|b|}\, b + \cdots, \tag{1.12}$$

where $\cdots$ is the part involving cohomology classes of degree $> |b|$. For $|b|$ even, this simplifies to

$$\mathrm{St}(b) = (-1)^{\frac{|b|}{2}} t^{\frac{p-1}{2}|b|} b + \cdots \tag{1.13}$$

At the other extreme, setting $t = \theta = 0$ in $\mathrm{St}(b)$ still yields the p-fold (cup) power b^p. The Cartan relation says that

$$\mathrm{St}(\tilde{b})\,\mathrm{St}(b) = (-1)^{|b|\,|\tilde{b}|\,\frac{p(p-1)}{2}}\,\mathrm{St}(\tilde{b}b). \tag{1.14}$$

Note that many coefficients of $\mathrm{St}(b)$ vanish, because this operation comes from the cohomology of the symmetric group. Concretely, if $|b|$ is even, all the potentially nonzero terms in $\mathrm{St}(b)$ are of the form $t^{k(p-1)}$ or $t^{k(p-1)-1}\theta$; and if $|b|$ is odd, of the form $t^{(k+1/2)(p-1)}$ or $t^{(k+1/2)(p-1)-1}\theta$. That is no longer true for quantum operations.

As usual, one adds up (1.10) over all A with weights q^A. The outcome is denoted by

$$Q\Sigma_b : H^*(M;\mathbb{F}_p) \longrightarrow (H^*(M;\Lambda)[[t,\theta]])^{*+p|b|}. \tag{1.15}$$

The non-equivariant ($t = \theta = 0$) part is the p-fold quantum product with b

$$Q\Sigma_b(c) = \overbrace{b * \cdots * b}^{p} * c + (\mathit{terms\,involving}\, t,\theta). \tag{1.16}$$

The case $b = 1$ is trivial

$$Q\Sigma_1 = \mathrm{id}. \tag{1.17}$$

The relation with the more standard formulation of the quantum Steenrod operation is that

$$\mathrm{QSt}(b) = Q\Sigma_b(1). \tag{1.18}$$

It is convenient to formally extend (1.15). First, turn it into an endomorphism of $H^*(M;\Lambda)[[t,\theta]]$, linearly in the variables q^A and (t,θ) (with appropriate Koszul signs). Next, extend the b-variable to $\beta \in H^*(M;\Lambda)$, by setting

$$Q\Sigma_\beta = \textstyle\sum_A q^{pA}\, Q\Sigma_{b_A} \quad \text{for}\, \beta = \textstyle\sum_A b_A q^A. \tag{1.19}$$

Then, the composition of these operations is described by

$$Q\Sigma_{\tilde{b}} \circ Q\Sigma_b = (-1)^{|b|\,|\tilde{b}|\,\frac{p(p-1)}{2}} Q\Sigma_{\tilde{b}\,*\,b}. \tag{1.20}$$

Note that for $b = 1$, (1.16) implies that $Q\Sigma_1$ is an automorphism of $H^*(M;\Lambda)$ $[[t,\theta]]$, and (1.20) that it is idempotent. Hence, it must be the identity, so those two properties imply (1.17).

1d. The quantum connection can be extended to $H^*(M;\Lambda)[[t,\theta]]$ by making it θ-linear. Our main result is:

Theorem 1.4. *For any $b \in H^*(M;\mathbb{F}_p)$, the operation $Q\Sigma_b$ is a covariantly constant endomorphism (of degree $p|b|$), meaning that it satisfies* (1.7).

Lemma 1.2 still applies (the presence of the additional θ-variable makes no difference). Hence, the classical part (1.11), together with the quantum connection, determines $Q\Sigma_b$ modulo I_{diff}.

Remark 1.5. Covariant constancy also means that $Q\Sigma_b$ is related to the fundamental solution of the quantum differential equation (see, e.g., [15]). To explain this, let us temporarily switch coefficients to $\mathbb{Q}$, and write $\tilde{\Lambda}$ for the associated Novikov ring. The fundamental solution is a trivialization of the quantum connection

$$\nabla\tilde{\Psi} = 0, \tag{1.21}$$

whose constant (in the q variables) term is the identity endomorphism. $\tilde{\Psi}$ is multivalued (has $\log(q^A)$ terms), and is also a series in t^{-1}. It is uniquely determined by those conditions, and one can write down an explicit formula in terms of Gromov–Witten invariants with gravitational descendants. Given $\beta \in H^*(M;\mathbb{Z})$, write

$$\tilde{\Xi}_\beta(\gamma) = \tilde{\Psi}(\beta\,\tilde{\Psi}^{-1}(\gamma)). \tag{1.22}$$

By construction, this is a covariantly constant endomorphism, whose constant term is cup product with β. It is single valued; more precisely

$$\tilde{\Xi}_\beta \in \mathrm{End}(H^*(M;\tilde{\Lambda}))[[t^{-1}]]. \tag{1.23}$$

For simplicity, suppose that $H^*(M;\mathbb{Z})$ is torsion-free. One can look at the denominators in $\tilde{\Xi}_\beta$, order by order in the covariant constancy equation. The upshot is that factors of $1/p$ appear for the first time in terms q^A, $A \in pH_2^{\mathrm{sphere}}(M;\mathbb{Z})$. As a consequence, $\tilde{\Xi}_\beta$ has a well-defined partial reduction mod p, which we denote by

$$\Xi_\beta \in \mathrm{End}(H^*(M;\Lambda/I_{\mathrm{diff}}))[[t^{-1}]], \tag{1.24}$$

and which only depends on $\beta \in H^*(M;\mathbb{F}_p)$. Let us extend (1.24) linearly to $\beta \in H^*(M;\mathbb{F}_p)[t,\theta]$, in which case Ξ_β can have both positive and negative powers of t. The case we are interested in is $\beta = \mathrm{St}(b)$. Because of the uniqueness property from Lemma 1.2, we then have

$$\Xi_{\mathrm{St(b)}} = Q\Sigma_b \quad \text{modulo}\, I_{\mathrm{diff}}. \tag{1.25}$$

Example 1.6. Consider $M = S^2$, with the standard basis $\{1, h\}$ of cohomology. Take $p > 2$ (the case $p = 2$ is straightforward, but requires slightly different notation). Using Theorem 1.4, one can compute that $Q\Sigma_h = -t^{p-1}\Sigma$, where

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}, \quad \begin{cases} \sigma_{11} = -\sum_{k=1}^{(p-1)/2} \frac{(2k-1)!}{(k!)^2(k-1)!^2} q^k t^{1-2k}, \\ \sigma_{12} = -\sum_{k=2}^{(p+1)/2} \frac{(2k-2)!}{(k-2)!\,(k-1)!^2\,k!} q^k t^{2-2k}, \\ \sigma_{21} = \sum_{k=0}^{(p-1)/2} \frac{(2k)!}{(k!)^4} q^k t^{-2k}, \\ \sigma_{22} = -\sigma_{11}. \end{cases} \tag{1.26}$$

In particular

$$\mathrm{QSt}(h) = -t^{p-1}\sigma_{11}\, 1 - t^{p-1}\sigma_{21}\, h. \tag{1.27}$$

Note that after multiplying with t^{p-1}, all the powers of t in (1.26) become nonnegative. More precisely

$$-\, t^{p-1}\Sigma = \begin{pmatrix} 0 & q^{(p+1)/2} \\ q^{(p-1)/2} & 0 \end{pmatrix} + (\text{terms involving}\, t), \tag{1.28}$$

in agreement with (1.16) and the fact that the pth quantum power of h is $q^{(p-1)/2}h$. This is proved in Sect. 6.

Example 1.7. Let M be a cubic surface in $\mathbb{C}P^3$ (this is $\mathbb{C}P^2$ blown up at 6 points, with its monotone symplectic form). Take $p = 2$, and let $h \in H^4(M;\mathbb{F}_2)$ be the Poincaré dual of a point. Then

$$\mathrm{QSt}(h) = \mathrm{St}(h) = t^2 h. \tag{1.29}$$

This is interesting because of its implications for Hamiltonian dynamics: by the criterion from [3,21], it means that M cannot admit a pseudo-rotation. We refer to Sect. 6 for further discussion.

The proof of Theorem 1.4 goes roughly as follows. We introduce another operation, depending on $a \in H^2(M;\mathbb{Z})$ as well as $b \in H^*(M;\mathbb{F}_p)$

$$Q\Pi_{a,b} : H^*(M;\mathbb{F}_p) \longrightarrow (H^*(M;\Lambda)[[t,\theta]])^{*+|a|-2+p|b|}. \tag{1.30}$$

Geometrically, this is obtained from (1.15) by equipping the underlying Riemann surface with an additional marked point, which can move around (we insert an incidence constraint dual to a at that point). A localisation-type argument yields

$$t\, Q\Pi_{a,b}(c) = Q\Sigma_b(a * c) - a * Q\Sigma_b(c). \tag{1.31}$$

We also have an analogue of the divisor equation

$$Q\Pi_{a,b}(c) = \partial_a Q\Sigma_b(c). \tag{1.32}$$

Theorem 1.4 follows immediately by combining (1.31) and (1.32).

Remark 1.8. Even though we have no immediate need for it here, it is worth while noting that $Q\Pi_{a,b}$ can be defined more generally for $a \in H^*(M;\mathbb{F}_p)$, and still satisfies (1.31), with suitable added Koszul signs (see Remark 4.11).

Remark 1.9. The argument above is closely related to the Cartan relation for quantum Steenrod squares. Namely, let us set $a = \mathrm{QSt}(b_1)$, $b = b_2$, $c = 1$ in (1.31). Then, using (1.20), one sees that

$$\begin{aligned} &t\, Q\Pi_{\mathrm{QSt(b_1),b_2}}(1) = (-1)^{|b_1|\,|b_2|} Q\Sigma_{b_2}(\mathrm{QSt}(b_1)) - \mathrm{QSt}(b_1) * \mathrm{QSt}(b_2) \\ &= (-1)^{|b_1|\,|b_2|\,(p(p-1)/2+1)} Q\Sigma_{b_2 * b_1}(1) - \mathrm{QSt}(b_1) * \mathrm{QSt}(b_2) \\ &= (-1)^{|b_1|\,|b_2|\,p(p-1)/2} \mathrm{QSt}(b_1 * b_2) - \mathrm{QSt}(b_1) * \mathrm{QSt}(b_2). \end{aligned} \tag{1.33}$$

In view of that, it is not surprising that in applications, computations based on covariant constancy closely resemble those from [22], where the Cartan relation was the main tool.

2. A bit of equivariant (co)homology

This section introduces some of the algebra and topology underlying our construction. Even though this is elementary, it is helpful as a guiding model for the later discussion.

2a. Write

$$\begin{aligned} &S^\infty = \{w = (w_0, w_1, w_2, \dots) \in \mathbb{C}^\infty \;:\; w_k = 0 \text{ for } k \gg 0, \\ &\|w\|^2 = |w_0|^2 + |w_1|^2 + \cdots = 1\}. \end{aligned} \tag{2.1}$$

Fix a prime p, and consider the $\mathbb{Z}/p$-action on S^∞ generated by

$$\tau(w_0, w_1, \dots) = (\zeta w_0, \zeta w_1, \dots), \;\; \zeta = e^{2\pi i/p}. \tag{2.2}$$

Take the following subsets:

$$\Delta_{2k} = \{w \in S^\infty \;:\; w_k \geq 0,\; w_{k+1} = w_{k+2} = \cdots = 0\}, \tag{2.3}$$

$$\begin{aligned} &\Delta_{2k+1} = \{w \in S^\infty \;:\; \mathrm{e}^{-i\theta} w_k \geq 0 \text{ for some } \theta \in [0, 2\pi/p], \\ &w_{k+1} = w_{k+2} = \cdots = 0\}. \end{aligned} \tag{2.4}$$

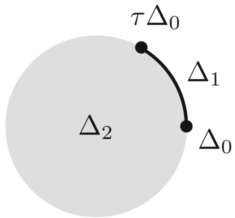


FIGURE 1. The first cells from (2.3), (2.4)

Each of them is homeomorphic to a disc, of the dimension indicated by the subscript (Fig. 1). More precisely, Δ_{2k} is a submanifold with boundary

$$\partial\Delta_{2k} = \{w_k = w_{k+1} = \cdots = 0\} \cong S^{2k-1}, \tag{2.5}$$

and Δ_{2k+1} a submanifold with two boundary faces, whose intersection forms a corner stratum

$$\begin{aligned}\partial\Delta_{2k+1} = \{w_k \geq 0,\, w_{k+1} = w_{k+2} = \cdots = 0\} \cup \{e^{-2\pi i/p} w_k \geq 0, \\ w_{k+1} = w_{k+2} = \cdots = 0\}.\end{aligned} \tag{2.6}$$

The subsets (2.3), (2.4) and their images under the $\mathbb{Z}/p$-action form an equivariant (and regular) cell decomposition of S^∞. The tangent space of Δ_{2k} at the point where $w_k = 1$ (and where all the other coordinates are therefore zero) can be identified with $\mathbb{C}^k$ by projecting to the first k coordinates; we use the resulting orientation. The tangent space of Δ_{2k+1} at the same point can be similarly identified with $\mathbb{C}^k \times i\mathbb{R}$; we use the orientation coming from the complex orientation of $\mathbb{C}^k$, followed by the positive vertical orientation of $i\mathbb{R}$. For those orientations, the differential in the cellular chain complex is

$$\partial\Delta_{2k} = \Delta_{2k-1} + \tau\Delta_{2k-1} + \cdots + \tau^{p-1}\Delta_{2k-1}, \tag{2.7}$$

$$\partial\Delta_{2k+1} = \tau\Delta_{2k} - \Delta_{2k}. \tag{2.8}$$

Here and below, the convention is to ignore terms with negative subscripts.

We adopt the quotient $S^\infty/(\mathbb{Z}/p)$ as our model for the classifying space $B(\mathbb{Z}/p)$. If we use $\mathbb{F}_p$-coefficients, the Δ_i become cycles on the quotient, and their homology classes form a basis for $H_*^{eq}(point; \mathbb{F}_p) = H_*(S^\infty/(\mathbb{Z}/p); \mathbb{F}_p)$. (Moreover, from (2.7), one sees that the Bockstein sends Δ_{2k} to Δ_{2k-1}.)

2b. Consider the diagonal embedding δ on $S^\infty/(\mathbb{Z}/p)$, and the induced map

$$\delta_* : H_*(S^\infty/(\mathbb{Z}/p); \mathbb{F}_p) \longrightarrow (H_*(S^\infty/(\mathbb{Z}/p); \mathbb{F}_p))^{\otimes 2}. \tag{2.9}$$

Lemma 2.1. *In homology with $\mathbb{F}_p$-coefficients*

$$\delta_*\Delta_i = \begin{cases} \displaystyle\sum_{i_1+i_2=i} \Delta_{i_1} \otimes \Delta_{i_2} & \textit{if } i \textit{ is odd or } p = 2, \\ \displaystyle\sum_{\substack{i_1+i_2=i \\ i_k \text{ even}}} \Delta_{i_1} \otimes \Delta_{i_2} & \textit{if } i \textit{ is even and } p > 2. \end{cases} \tag{2.10}$$

Proof. For $p = 2$, this is clear: from the relation between diagonal map and cup product, and the ring structure on the cohomology of $\mathbb{R}P^\infty = S^\infty/(\mathbb{Z}/2)$, we can see that $\delta_* \Delta_i$ must have nonzero components in all groups $H^{i_1} \otimes H^{i_2}$, and each of those is a copy of $\mathbb{F}_2$.

For $p > 2$, the same argument shows that exactly the terms in (2.10) must occur, but possibly with some nonzero $\mathbb{F}_p$-coefficients, which have to be determined by looking a little more carefully. Choose generators $\theta \in H^1(S^\infty/(\mathbb{Z}/p); \mathbb{F}_p)$ and $t \in H^2(S^\infty/(\mathbb{Z}/p); \mathbb{F}_p)$, so that

$$\langle \theta, \Delta_1 \rangle = 1, \quad \langle t, \Delta_2 \rangle = -1. \tag{2.11}$$

Because Δ_2 was defined using the complex orientation, this means that t is the pullback of the (mod p) Chern class of the tautological line bundle $S^\infty \to \mathbb{C}P^\infty$ under the quotient map $S^\infty/(\mathbb{Z}/p) \to S^\infty/S^1 = \mathbb{C}P^\infty$. Looking at the orientations of the higher dimensional cells yields

$$\langle t^k \theta, \Delta_{2k+1} \rangle = \langle t^k, \Delta_{2k} \rangle = (-1)^k. \tag{2.12}$$

For $k = k_1 + k_2$, we have

$$\langle t^{k_1} \otimes t^{k_2}, \delta_* \Delta_{2k} \rangle = \langle \delta^*(t^{k_1} \otimes t^{k_2}), \Delta_{2k} \rangle = \langle t^k, \Delta_{2k} \rangle, \tag{2.13}$$

$$\langle t^{k_1}\theta \otimes t^{k_2}, \delta_* \Delta_{2k+1} \rangle = \langle \delta^*(t^{k_1}\theta \otimes t^{k_2}), \Delta_{2k+1} \rangle = \langle t^k\theta, \Delta_{2k+1} \rangle, \tag{2.14}$$

and that implies that the coefficients in (2.10) are all 1, as desired. □

What does this mean on the cochain level? For each k, take a smooth triangulation of $S^{2k-1}/(\mathbb{Z}/p)$. Pull that back (taking preimages of the simplices) to a triangulation of $\partial\Delta_{2k}$, and then extend that to a triangulation of Δ_{2k}. The outcome is an explicit smooth singular chain in $S^\infty/(\mathbb{Z}/p)$, denoted by $\tilde{\Delta}_{2k}$, which becomes a singular cycle when the coefficients are reduced modulo p, and which represents the homology class of Δ_{2k} in $H_*(S^\infty/(\mathbb{Z}/p); \mathbb{F}_p)$. A version of the same process produces corresponding singular chains $\tilde{\Delta}_{2k-1}$. With that in mind, let us look at the relations underlying (2.10)

$$\delta\tilde{\Delta}_i \sim \begin{cases} \displaystyle\sum_{i_1+i_2=i} \tilde{\Delta}_{i_1} \times \tilde{\Delta}_{i_2} & \text{if } i \text{ is odd or } p=2, \\ \displaystyle\sum_{\substack{i_1+i_2=i \\ i_k \text{ even}}} \tilde{\Delta}_{i_1} \times \tilde{\Delta}_{i_2} & \text{if } i \text{ is even and } p>2. \end{cases} \tag{2.15}$$

On the right-hand side, one decomposes the products into simplices. After that, the relation means that there is a singular chain whose boundary (mod p) equals the difference between the two sides. That chain can again be chosen to be smooth. One could in principle try to spell all of this out using explicit chains, but that is not necessary for our purpose.

2c. Consider the two-sphere $S = \bar{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$, again with a $\mathbb{Z}/p$-action $\sigma(v) = \zeta v$, and the subsets (shown in Fig. 2)

$$P_0 = \{v = 0\}, \quad Q_0 = \{v = \infty\}, \tag{2.16}$$

$$L_1 = \{v \geq 0\} \cup \{v = \infty\}, \tag{2.17}$$

$$B_2 = \{e^{-i\theta} v \geq 0 \text{ for some } \theta \in [0, 2\pi/p]\} \cup \{v = \infty\}. \tag{2.18}$$

We use the real orientation of L_1, and the complex orientation of B_2. Let us denote the associated cellular chain complex simply by $C_*(S)$. Its differential is

$$\partial P_0 = \partial Q_0 = 0, \tag{2.19}$$
$$\partial L_1 = Q_0 - P_0, \tag{2.20}$$
$$\partial B_2 = L_1 - \sigma L_1. \tag{2.21}$$

Now look at $S^\infty \times_{\mathbb{Z}/p} S$, which means identifying

$$(w, \sigma v) \sim (\tau w, v). \tag{2.22}$$

This inherits a cell decomposition. The associated differential, which we denote by ∂^{eq}, is

$$\partial^{\mathrm{eq}}(\Delta_{2k} \times P_0) = 0, \ \ \partial^{\mathrm{eq}}(\Delta_{2k+1} \times P_0) = 0, \tag{2.23}$$
$$\partial^{\mathrm{eq}}(\Delta_{2k} \times Q_0) = 0, \ \ \partial^{\mathrm{eq}}(\Delta_{2k+1} \times Q_0) = 0, \tag{2.24}$$
$$\partial^{eq}(\Delta_{2k} \times \sigma^j L_1) = \Delta_{2k} \times (Q_0 - P_0) + \Delta_{2k-1} \times (L_1 + \sigma L_1 + \cdots + \sigma^{p-1} L_1), \tag{2.25}$$
$$\partial^{eq}(\Delta_{2k+1} \times \sigma^j L_1) = -\Delta_{2k+1} \times (Q_0 - P_0) + \Delta_{2k} \times (\sigma^{j+1} L_1 - \sigma^j L_1), \tag{2.26}$$
$$\partial^{eq}(\Delta_{2k} \times \sigma^j B_2) = -\Delta_{2k} \times (\sigma^{j+1} L_1 - \sigma^j L_1) + \Delta_{2k-1} \times (B_2 + \cdots + \sigma^{p-1} B_2), \tag{2.27}$$
$$\partial^{\mathrm{eq}}(\Delta_{2k+1} \times \sigma^j B_2) = \Delta_{2k+1} \times (\sigma^{j+1} L_1 - \sigma^j L_1) + \Delta_{2k} \times (\sigma^{j+1} B_2 - \sigma^j B_2). \tag{2.28}$$

Lemma 2.2. *Take coefficients in $\mathbb{F}_p$. In the cellular complex of $S^\infty \times_{\mathbb{Z}/p} S$, the following homology relationships hold:*

$$\Delta_{2k} \times (Q_0 - P_0) \sim \Delta_{2k-2} \times (B_2 + \sigma B_2 + \cdots + \sigma^{p-1} B_2), \tag{2.29}$$
$$\Delta_{2k+1} \times (Q_0 - P_0) \sim \Delta_{2k-1} \times (B_2 + \sigma B_2 + \cdots + \sigma^{p-1} B_2). \tag{2.30}$$

Proof. (2.29) is obtained by subtracting (2.25) from the following, which comes from (2.28):

$$\begin{aligned} &\partial^{\mathrm{eq}}\big(\Delta_{2k+1} \times (\sigma B_2 + 2\sigma^2 B_2 + \cdots + (p-1)\sigma^{p-1} B_2)\big) \\ &\qquad = -\Delta_{2k+1} \times (L_1 + \cdots + \sigma^{p-1} L_1) - \Delta_{2k} \times (B_2 + \cdots + \sigma^{p-1} B_2). \end{aligned} \tag{2.31}$$

The second relation (2.30) is a combination of (2.26), (2.27). ☐

To fit this into the general framework of equivariant homology, note that as an application of the localisation theorem, the map induced by inclusion of the fixed point set

$$H_*^{\mathrm{eq}}(\mathrm{point}; \mathbb{F}_p) \otimes P_0 \oplus H_*^{\mathrm{eq}}(\mathrm{point}; \mathbb{F}_p) \otimes Q_0 \longrightarrow H_*^{\mathrm{eq}}(S; \mathbb{F}_p) = H_*(S^\infty \times_{\mathbb{Z}/p} S; \mathbb{F}_p) \tag{2.32}$$

must be an isomorphism in sufficiently high degrees. Using the computations above, one can see how that works out concretely: (2.32) is surjective, and it fails to be injective only in degrees 0 and 1, where the kernel is generated by $\Delta_0 \otimes (Q_0 - P_0)$ and $\Delta_1 \otimes (Q_0 - P_0)$, respectively.

More generally, take any (homologically graded) chain complex, carrying a $(\mathbb{Z}/p)$-action. Its equivariant homology is defined by taking the tensor product with the previously considered cellular complex of S^∞, and then passing to coinvariants for the combined action in the same sense as in (2.22). The resulting equivariant differential is

$$\partial^{\mathrm{eq}}(\Delta_{2k} \otimes \xi) = \Delta_{2k-1} \otimes (\xi + \sigma\xi + \cdots + \sigma^{p-1}\xi) + \Delta_{2k} \otimes \partial\xi, \tag{2.33}$$

$$\partial^{\mathrm{eq}}(\Delta_{2k+1} \otimes \xi) = -\Delta_{2k+1} \otimes \partial\xi + \Delta_{2k} \otimes (\sigma\xi - \xi). \tag{2.34}$$

Here, ξ is an element of the original chain complex, and σ is the automorphism which generates its $(\mathbb{Z}/p)$-action. These formulae generalize the ones we have previously written down for $C_*(S)$.

2d. Dually to our previous construction, one can start with a cohomologically graded complex C with a $(\mathbb{Z}/p)$-action, and define an equivariant complex

$$C_{\mathrm{eq}} = C[[t, \theta]], \tag{2.35}$$

where the formal variables are as in (1.9), with differential

$$d_{\mathrm{eq}}(xt^k) = dx\, t^k + (-1)^{|x|}(\sigma x - x)t^k\theta, \tag{2.36}$$

$$d_{\mathrm{eq}}(xt^k\theta) = dx\, t^k\theta + (-1)^{|x|}(x + \sigma x + \cdots + \sigma^{p-1}x)t^{k+1}. \tag{2.37}$$

Write $H^*_{\mathrm{eq}}(C) = H^*(C_{\mathrm{eq}})$ for the resulting cohomology.

Lemma 2.3. *On C_{eq}, the operations t and σt are homotopic.*

Proof. The desired homotopy is $h(xt^k) = 0$, $h(xt^k\theta) = (-1)^{|x|}xt^{k+1}$. □

From now on, we work with $\mathbb{F}_p$-coefficients. In that case, the equivariant complex (2.35) carries a degree 1 endomorphism $\tilde{\theta}$, which one can informally think of as a corrected version of multiplication with θ (acting on the left)

$$\tilde{\theta}(xt^k) = (-1)^{|x|}xt^k\theta, \tag{2.38}$$

$$\tilde{\theta}(xt^k\theta) = (-1)^{|x|}(\sigma x + 2\sigma^2 x + \cdots + (p-1)\sigma^{p-1}x)t^{k+1}. \tag{2.39}$$

The second part (2.39) contains the kind of expression we have seen previously in (2.31). It is helpful to keep in mind that modulo p

$$\mathrm{id} + \sigma + \sigma^2 + \cdots + \sigma^{p-1} = (\sigma - \mathrm{id})^{p-1} = \sigma(\sigma - \mathrm{id})^{p-1} = \cdots, \tag{2.40}$$

$$\sigma + 2\sigma^2 + \cdots + (p-1)\sigma^{p-1} = -\sigma(\sigma - \mathrm{id})^{p-2}. \tag{2.41}$$

Using that, one sees that the map $\tilde{\theta}$ is a chain map (of degree 1) with respect to d_{eq}

$$\begin{aligned} d_{\mathrm{eq}}\tilde{\theta}(xt^k) &= d_{\mathrm{eq}}((-1)^{|x|}xt^k\theta) = (-1)^{|x|}dx\, t^k\theta + (\mathrm{id} + \sigma + \cdots)x\, t^{k+1} \\ &= (-1)^{|x|}\mathrm{d}x\, t^k\theta - (\sigma + 2\sigma^2 + \cdots)(\sigma - \mathrm{id})xt^{k+1} \\ &= \tilde{\theta}(-dx\, t^k - (-1)^{|x|}(\sigma - \mathrm{id})x\, t^k\theta) = -\tilde{\theta}d_{\mathrm{eq}}(xt^k), \end{aligned} \tag{2.42}$$

and similarly

$$\begin{aligned} d_{\mathrm{eq}}\tilde{\theta}(xt^k\theta) &= d_{\mathrm{eq}}((-1)^{|x|}(\sigma + 2\sigma^2 + \cdots)xt^{k+1}) \\ &= (-1)^{|x|}(\sigma + 2\sigma^2 + \cdots)dx\, t^{k+1} - (\mathrm{id} + \sigma + \cdots)xt^{k+1}\theta \\ &= -\tilde{\theta}(dx\, t^k\theta + (-1)^{|x|}(\mathrm{id} + \sigma + \cdots)xt^{k+1}) = -\tilde{\theta}d_{\mathrm{eq}}(xt^k\theta). \end{aligned} \tag{2.43}$$

Lemma 2.4. *Up to homotopy, $\tilde{\theta}^2$ is multiplication by t if $p = 2$, and 0 for $p > 2$.*

Proof. In terms of (2.41), $\tilde{\theta}^2$ is the action of $-\sigma(\sigma-\mathrm{id})^{p-2}t$ on the equivariant complex. However, the action of $(\sigma - \mathrm{id})t$ is nullhomotopic by Lemma 2.3, and that implies the desired statement. □

A classical application of equivariant cohomology (basic to the definition of Steenrod operations) is to start with a general cochain complex C (without any $(\mathbb{Z}/p)$-action), and consider its p-fold tensor product $C^{\otimes p}$ with the action that cyclically permutes the tensor factors. The equivariant complex $(C^{\otimes p})_{\mathrm{eq}}$ is a homotopy invariant of C. We recall the following:

Lemma 2.5. *Taking a cocycle $x \in C$ to $x^{\otimes p} \in (C^{\otimes p})_{\mathrm{eq}}$ yields a map*

$$H^*(C) \longrightarrow H^{p*}_{\mathrm{eq}}(C^{\otimes p}), \tag{2.44}$$

which becomes additive after multiplying by t.

Proof. Since $x^{\otimes p}$ is a $(\mathbb{Z}/p)$-invariant cocycle in $C^{\otimes p}$ (note that the Koszul signs here are always trivial), it is also a d_{eq}-cocycle.

The next step is to show that if we have two cohomologous cocycles, $x_1 - x_2 = dz$, then $x_1^{\otimes p}$ and $x_2^{\otimes p}$ are cohomologous in $(C^{\otimes p})_{\mathrm{eq}}$. It is enough to consider the case where C is three-dimensional, with basis (x_1, x_2, z); the general case then follows by mapping this C into any desired complex. Take a one-dimensional complex D with a single generator y, and the map $C \to D$ which takes both x_k to y (and maps z to zero). This is clearly a quasi-isomorphism, and therefore induces a quasi-isomorphism $(C^{\otimes p})_{\mathrm{eq}} \to (D^{\otimes p})_{\mathrm{eq}}$. Under that quasi-isomorphism, both $x_1^{\otimes p}$ and $x_2^{\otimes p}$ go to $y^{\otimes p}$. Therefore, they must be cohomologous in $(C^{\otimes p})_{\mathrm{eq}}$.

The additivity statement can be proved by an explicit formula: if we take

$$(x_1 + x_2)^{\otimes p} - x_1^{\otimes p} - x_2^{\otimes p} \tag{2.45}$$

and expand it out, we get $2^p - 2$ monomials, which occur in free $(\mathbb{Z}/p)$-orbits. Take one representative for each orbit, add them up, and multiply the outcome by θ. This yields a cochain in $(C^{\otimes p})_{\mathrm{eq}}$ whose boundary is t times (2.45), up to sign. □

Finally, we return to the example of S. Take the cellular chain complex and reverse its grading, to make it cohomological. Then, on $C_{-*}(S)_{\mathrm{eq}}$ we have

$$d_{\mathrm{eq}}(P_0\, t^k) = 0, \;\; d_{\mathrm{eq}}(P_0\, t^k\theta) = 0, \tag{2.46}$$

$$d_{\mathrm{eq}}(P_0\, t^k) = 0, \;\; d_{\mathrm{eq}}(P_0\, t^k\theta) = 0, \tag{2.47}$$

$$d_{\mathrm{eq}}(\sigma^j L_1\, t^k) = (Q_0 - P_0)t^k - (\sigma^{j+1}L_1 - \sigma^j L_1)t^k\theta, \tag{2.48}$$

$$d_{\mathrm{eq}}(\sigma^j L_1\, t^k\theta) = (Q_0 - P_0)t^k\theta - (L_1 + \cdots + \sigma^{p-1}L_1)t^{k+1}, \tag{2.49}$$

$$d_{\mathrm{eq}}(\sigma^j B_2\, t^k) = -(\sigma^{j+1}L_1 - \sigma^j L_1)t^k + (\sigma^{j+1}B_2 - \sigma^j B_2)t^k\theta, \tag{2.50}$$

$$d_{\mathrm{eq}}(\sigma^j B_2\, t^k\theta) = -(\sigma^{j+1}L_1 - \sigma^j L_1)t^k\theta + (B_2 + \cdots + \sigma^{p-1}B_2)t^{k+1}. \tag{2.51}$$

With $\mathbb{F}_p$-coefficients, we have the following analogue of Lemma 2.2, proved in the same way:

Lemma 2.6. *The following cohomology relations hold in $C_{-*}(S)_{\mathrm{eq}}$:*

$$(P_0 - Q_0)t^k \sim (B_2 + \sigma B_2 + \cdots + \sigma^{p-1}B_2)t^{k+1}, \tag{2.52}$$

$$(P_0 - Q_0)t^k\theta \sim (B_2 + \sigma B_2 + \cdots + \sigma^{p-1}B_2)t^{k+1}\theta. \tag{2.53}$$

3. Basic moduli spaces

This section introduces the relevant moduli spaces of pseudo-holomorphic curves, in their most basic form. This means that we look at a version of the small quantum product, and one of its properties, the divisor equation. Like the previous section, this should be considered as a toy model which introduces some ideas that will recur in more complicated form later on.

3a. Let M^{2n} be a weakly monotone closed symplectic manifold. Choose a Morse function f and metric g, so that the associated gradient flow is Morse-Smale. Our terminology for stable and unstable manifolds is that $\dim(W^s(x)) = |x|$ is the Morse index, whereas $\dim(W^u(x)) = 2n - |x|$.

Definition 3.1. Fix some compatible almost complex structure J. A *J-holomorphic chain of length l* is a set of maps

$$u_1, \dots, u_l : \mathbb{C}P^1 \to M$$

, such that $\overline{\partial}_J u = 0$, and such that

$$u_k(\infty) = u_{k+1}(0) \text{ for } k = 1, \dots, l-1. \tag{3.1}$$

We call such a chain *simple* if each of the maps is simple (non-multiply-covered and non-constant) and no two of the maps are reparametrisations of each other. Two simple chains are called equivalent if they are related by reparametrisations $(\phi_1, \dots, \phi_l)$ of each component, such that $\phi_k(0) = 0$ and $\phi_k(\infty) = \infty$. The moduli space of simple chains representing some class $A \in H_2(M;\mathbb{Z})$ is denoted by $\mathcal{M}_A(\mathit{chain}, l)$. It comes with evaluation maps at the "endpoints of the chains", which send $(u_1, \dots, u_l)$ to $u_1(0)$ and $u_l(\infty)$, respectively.

Assumption 3.2. We fix some compatible almost complex structure J with the following properties.

(i) All spaces $\mathcal{M}_A(\mathrm{chain}, l)$ are regular.
(ii) On those spaces, the evaluation maps $(u_1, \dots, u_l) \mapsto u_1(0)$ are transverse to the stable and unstable manifolds of our Morse function.

Assumption 3.2 is satisfied for generic choice of J. The simplest aspect is the $l = 1$ case of (i), which is just generic regularity of simple J-holomorphic spheres (because of the weak monotonicity condition, this also implies the absence of spheres with negative Chern number). The general form of (i) is a version of [14, Definition 6.2.1] (using chains rather than general trees), and is generically satisfied by [14, Theorem 6.2.6]. The transversality theory for evaluation maps developed there also yields the genericity of (ii).

Our main moduli space uses a specific $(p+2)$-marked sphere as the domain. We introduce specific notation for it: taking $\zeta^{1/2} = e^{\pi i/p}$, set

$$\begin{aligned} &C = \mathbb{C}P^1,\ z_{C,0}=0,\ z_{C,1}=\zeta^{1/2}, \\ &z_{C,2}=\zeta^{3/2},\ \ldots,\ z_{C,p}=\zeta^{(2p-1)/2}=\zeta^{-1/2},\ z_{C,\infty}=\infty. \end{aligned} \tag{3.2}$$

An inhomogeneous term is a J-complex anti-linear vector bundle map $\nu_C : TC \to TM$, where both bundles involved have been pulled back to $C \times M$, such that ν_C is zero near the marked points (3.2). The associated inhomogeneous Cauchy–Riemann equation is

$$\begin{aligned} &u : C \longrightarrow M, \\ &(\bar{\partial}_J u)_z = \nu_{C,z,u(z))}. \end{aligned} \tag{3.3}$$

Given critical points $x_0, \ldots, x_p, x_\infty$ of f, we consider solutions of (3.3) with incidence conditions at the (un)stable manifolds

$$u(z_{C,0}) \in W^u(x_0),\ \ldots,\ u(z_{C,p}) \in W^u(x_p),\ u(z_{C,\infty}) \in W^s(x_\infty). \tag{3.4}$$

It is maybe better to think of this as having gradient half-flowlines

$$\begin{aligned} &y_0, \ldots, y_p : (-\infty, 0] \longrightarrow M, && && y_\infty : [0,\infty) \longrightarrow M, \\ &y_k' = \nabla f(y_k), && && y_\infty' = \nabla f(y_\infty), \\ &y_k(0) = u(z_{C,k}), && \text{and} && y_\infty(0) = u(z_{C,\infty}), \\ &\lim_{s\to-\infty} y_k(s) = x_k && && \lim_{s\to\infty} y_\infty(s) = x_\infty. \end{aligned} \tag{3.5}$$

Assumption 3.3. We impose the following requirements:

(i) The moduli space of solutions of (3.3), (3.4) is regular.

(ii) Take an element in the same space, with a simple J-holomorphic bubble attached at an arbitrary point. This means that we have a pair (u, u_0) with u as in (3.3), (3.4), a point $z \in C$, and a simple J-holomorphic $u_0 : \mathbb{C}P^1 \to M$ with $u(z) = u_0(0)$. We want this moduli space to be regular, as well.

(iii) Consider solutions with a simple holomorphic chain attached at each of a subset of the $(p+2)$ marked points, and incidence constraints transferred accordingly. For simplicity, let us spell out what this means only in the case of a single chain, attached at $z_{C,\infty}$. In that case, we have a solution of (3.3), and a simple holomorphic chain $(u_1, \ldots, u_l)$, with the conditions

$$\begin{aligned} &u(z_{C,0}) \in W^u(x_0),\ \ldots,\ u(z_{C,p}) \in W^u(x_p), \\ &u(z_{C,\infty}) = u_1(0),\ u_l(\infty) \in W^s(x_\infty). \end{aligned} \tag{3.6}$$

We require that the resulting moduli space should be regular. In the general case where there are several marked points with a chain attached to each, we transfer the adjacency condition involving (un)stable manifolds to the end of the respective chain.

This assumption are satisfied for a generic choice of inhomogeneous term (where J is assumed chosen as in Assumption 3.2), following the argument from [14, Chapter 8]. A few comments may be appropriate. In (ii), the bubble may be attached at one of the marked points. Let us say that this point is $z_{C,\infty}$, in which case we have

$$u(z_{C,\infty}) = u_0(0) \in W^s(x_\infty). \tag{3.7}$$

Assumption 3.2(ii), for $l = 1$, says that the subspace of maps u_0 satisfying $u_0(0) \in W^s(x_\infty)$ is regular. What we want to achieve is that the evaluation map on that subspace is transverse to $u \mapsto u(z_{C,\infty})$. This is clearly satisfied for generic ν_C. In the same way, genericity of (iii) depends on Assumption 3.2(ii), but this time for arbitrary l.

Given $A \in H_2(M;\mathbb{Z})$, let $\mathcal{M}_A(C, x_0, \dots, x_p, x_\infty)$ be the space of solutions of (3.3), (3.4), such that u represents A. Given our regularity requirement, this is a manifold of dimension

$$\dim \mathcal{M}_A(C, x_0, \dots, x_p, x_\infty) = 2c_1(A) + |x_\infty| - |x_0| - \cdots - |x_p|. \tag{3.8}$$

We denote by $\bar{\mathcal{M}}_A(C, x_0, \dots, x_p, x_\infty)$ the standard compactification. On the pseudo-holomorphic map side, this involves the stable map compactification, and on the Morse-theoretic side, one allows the flow lines to break. Details are in [16, Section 5] (for illustration, see Fig. 3 there). To make the exposition more self-contained, we recall here that a point of the compactification consists of:

- A genus zero nodal Riemann surface $\hat{C}$ with $(p+2)$ smooth marked points $z_{\hat{C},0}, \dots, z_{\hat{C},\infty}$. One of the irreducible components of that surface is distinguished, and identified with C in a preferred way. Moreover, if one collapses all the other components (usually called bubble components), and transfers the marked points along with the collapse, then those marked points will end up in the same positions as in (3.2). In other words, if $z_{\hat{C},k}$ does not lie on the distinguished component, then it must lie on a bubble tree attached to that component at $z_{C,k}$.
- A map $\hat{u} : \hat{C} \to M$ which, on the distinguished component, is a solution of (3.3), and on the other components, is a J-holomorphic map. Moreover, those J-holomorphic maps must be stable (if they are constant on some non-distinguished component, then that component must have at least three special points). Finally, the map $\hat{u}$ still represents the homology class A.
- For each $k \in \{0, \dots, p\}$, a finite sequence of gradient flow lines $\hat{y}_{k,0} : \mathbb{R} \to M, \dots, \hat{y}_{k,m_k-1} : \mathbb{R} \to M$, $\hat{y}_{k,m_k} : (-\infty, 0] \to M$ (all but the last should be non-constant). These should satisfy

$$\begin{aligned}
&\lim_{s\to-\infty} \hat{y}_{k,0}(s) = x_k,\\
&\lim_{s\to+\infty} \hat{y}_{k,j}(s) = \lim_{s\to-\infty} \hat{y}_{k,j+1}(s),\\
&\hat{y}_{k,m_k}(0) = \hat{u}(z_{\hat{C},k}).
\end{aligned}$$

- Similarly, gradient flow lines $\hat{y}_{\infty,0} : [0,\infty) \to M$, $\hat{y}_{\infty,1} : \mathbb{R} \to M$, $\dots$, $\hat{y}_{\infty,m_\infty} : \mathbb{R} \to M$. Here, the conditions are that

$$\begin{aligned}
&\hat{y}_{\infty,0}(0) = \hat{u}(z_{\hat{C},\infty}),\\
&\lim_{s\to+\infty} \hat{y}_{\infty,j}(s) = \lim_{\to-\infty} \hat{y}_{\infty,j+1}(s),\\
&\lim_{s\to+\infty} \hat{y}_{\infty,m_\infty}(s) = x_\infty.
\end{aligned}$$

Lemma 3.4. *(i) If the dimension* (3.8) *is* 0*, we have*

$$\mathcal{M}_A(C, x_0, \ldots, x_p, x_\infty) = \bar{\mathcal{M}}_A(C, x_0, \ldots, x_p, x_\infty), \tag{3.9}$$

which means that the moduli space is a finite set.

(ii) If the dimension is 1*, the compactification is a manifold with boundary, with the interior being the space* $\mathcal{M}_A(\cdots)$*; the boundary points involve no bubbling, and only once-broken gradient flow lines.*

Sketch of proof. The proof is in [16, Theorem 3.4] for the 0-dimensional case, and [16, Section 3.3] for the 1-dimensional case. We will summarize it here. Recall that when compactifying the moduli space, what can occur is a mixture of Gromov compactification and breaking of Morse flowlines. Take a limit point in the form discussed above, assuming for simplicity that there is no breaking of Morse flow lines ($m_0 = \cdots = m_k = m_\infty = 0$). Collapse all the bubble components which carry constant J-holomorphic maps (called ghost components). Then, carry out the following further simplifications:

- Suppose that after that initial collapse of constant components, all marked points come to lie on the distinguished component. In that case, we forget all bubbles except for one, which carries a non-constant J-holomorphic map that intersects the image of the distinguished component at some point (these must be such a bubble). Finally, we also replace the map on that bubble component by its underlying simple map. That puts us in the situation of Assumption 3.3(ii), where (u, u_0) represents some class whose Chern number is less than equal that of A.
- Take the other case (after the initial collapse, at least one marked point does not lie on the distinguished component). In that case, we forget any bubble tree that carries no marked points. This leaves only the distinguished component and at most one holomorphic chain attached at each of its $(p+2)$ marked points; the component of that chain which is furthest from the distinguished component will carry the marked point. As before, we replace all multiply covered bubbles with the underlying simple maps. Moreover, if two holomorphic maps u_i, u_j with $i < j$ in one chain are reparametrisations of each other, we remove the bubbles carrying $u_i, \ldots, u_{j-1}$. After this, all attached holomorphic chains are simple, and we are in the situation of Assumption 3.3(iii), with at least one nontrivial bubble chain, and where again the Chern number has not increased from that of the original A.

This process is shown in Fig. 3. All these simplified limits have codimension ≥ 2, and hence cannot occur in the moduli spaces under consideration. The case that includes Morse-theoretic breaking is similar, and we will not discuss it further. □

Given some coefficient field $\mathbb{F}$, we denote by $\mathbb{F}_x$ the one-dimensional vector space generated by orientations of $W^s(x)$, where the sum of the two orientations is zero. The Morse complex is

$$\mathrm{CM}^k(f) = \bigoplus_{|x|=k} \mathbb{F}_x. \tag{3.10}$$

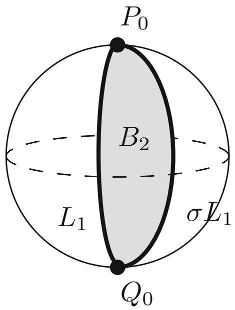


FIGURE 2. The cells from (2.16)–(2.18)

A choice of orientations of $W^s(x_0), \ldots, W^s(x_p), W^s(x_\infty)$ determines an orientation of the moduli space $\mathcal{M}_A(C, x_0, \ldots, x_p, x_\infty)$. In particular, every point in a zero-dimensional moduli space gives rise to a preferred isomorphism (an abstract version of a ± 1 contribution) $\mathbb{F}_{x_0} \otimes \cdots \otimes \mathbb{F}_{x_p} \cong \mathbb{F}_{x_\infty}$. One adds up those contributions to get a map

$$m_A(C, x_0, \ldots, x_p, x_\infty) : \mathbb{F}_{x_0} \otimes \cdots \otimes \mathbb{F}_{x_p} \longrightarrow \mathbb{F}_{x_\infty}, \tag{3.11}$$

and those maps are the coefficients of a chain map

$$S_A : \mathrm{CM}^*(f)^{\otimes p+1} \longrightarrow \mathrm{CM}^{*-2c_1(A)}(f). \tag{3.12}$$

Up to chain homotopy, this map is independent of the choice of almost complex structure and inhomogeneous term, by a parametrized version of our previous argument. Of course, the outcome is not in any sense surprising:

Lemma 3.5. *Up to chain homotopy, $S_A(x_0, x_1, \ldots, x_p)$ is the A-contribution to the $(p+1)$-fold quantum product $x_0 * x_1 * \cdots * x_p$.*

Proof. This is a familiar argument, which involves degenerating C to a nodal curve each of whose components has three marked points, one option being that drawn in Fig. 4(i); each component will again carry a Cauchy–Riemann equation with an inhomogeneous term. In our Morse-theoretic context, there is an additional step, familiar from the proof that the PSS map is an isomorphism, such as in [12, Theorem 6] (see Fig. 2), [1, Section 4] (see Fig. 6), [13, Section 4], or for more details [11]. Namely, one adds a length parameter, and inserts a finite length flow line of our Morse function at each node. As the length goes to infinity, each of the flow lines we have inserted breaks, see Fig. 4(ii); and that limit gives rise to the Morse homology version of the iterated quantum product. The parametrized moduli space (consisting of, first, the parameter used to degenerate C; and then in the second step, using the finite edge-length as a parameter) then yields a chain homotopy between those two operations. $\square$

Remark 3.6. Our use of inhomogeneous terms means that the moduli space could be nonzero for classes $A \in H_2(M; \mathbb{Z})$ which do not give rise to monomials in Λ (because $\int_A \omega_M$ is either negative, or it is zero but $A \neq 0$). However,

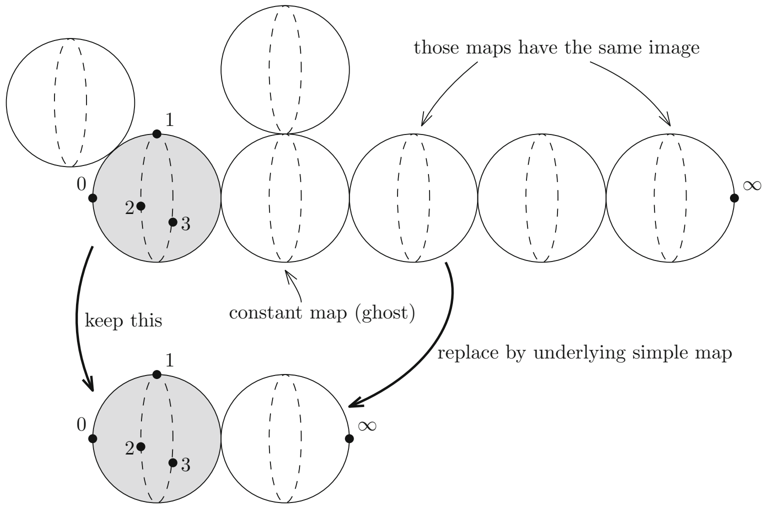


FIGURE 3. Simplification process from the proof of Lemma 3.4, for $p = 3$. The stable map at the top (with 7 components, and where the principal component is shaded) yields a solution of (3.3) with a length 1 simple chain attached

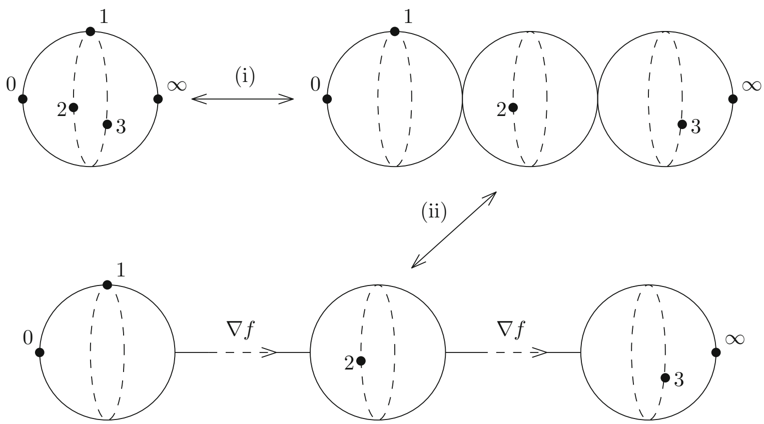


FIGURE 4. A schematic picture of the proof of Lemma 3.5, with $p = 3$

by choosing the inhomogeneous term small and using a compactness argument, one can rule out that undesired behaviour for any specific A. Since the outcome is independent of the choice up to chain homotopy, the resulting cohomology level structure is indeed defined over Λ.

3b. Fix an oriented codimension 2 submanifold $\Omega \subset M$. When choosing an almost complex structure, there are additional restrictions:

Assumption 3.7. In the situation of Assumption 3.2, we additionally require that the evaluation map on the space of simple J-holomorphic chains should be transverse to Ω.

We equip the Riemann surface (3.2) with "an additional marked point which can move freely" (and which will carry an Ω-incidence constraint). Formally, this means that we consider a family of genus zero nodal curve with sections

$$\begin{aligned} &\mathcal{C} \longrightarrow S, \\ &z_{\mathcal{C},0}, \dots, z_{\mathcal{C},p}, z_{\mathcal{C},\infty}, z_{\mathcal{C},*} : S \longrightarrow \mathcal{C}, \end{aligned} \tag{3.13}$$

where the parameter space S is again a copy of $\mathbb{C}P^1$, such that the following holds:

- The critical values of (3.13) are precisely the marked points from (3.2). If v is a regular value, the fibre $\mathcal{C}_v$ is canonically identified with C; that identification takes the points $z_{\mathcal{C}_r,0}, \dots, z_{\mathcal{C}_r,p}, z_{\mathcal{C}_r,\infty}$ arising from (3.13) to their counterparts in (3.2), and the remaining point $z_{\mathcal{C}_v,*}$ to v.
- If v is a singular value, $\mathcal{C}_v = \mathcal{C}_{v,+} \cup \mathcal{C}_{v,-}$ is a nodal surface with two components. The first component $\mathcal{C}_{v,+}$ is again identified with C, and the second component $\mathcal{C}_{v,-}$ is a rational curve attached to the first one at v. The first component carries all the marked points that C does, with the exception of the one which is equal to v; and the second component carries the two remaining marked points, considered to be distinct and also different from the node (so, the second component has three special points, which identifies it up to unique isomorphism).

Explicitly, (3.13) is constructed by starting with the trivial family $C \times S \to S$, and then blowing up the points (v, v), where v is one of the marked points in (3.2). One takes the proper transforms of the constant sections and of the diagonal section, which yield the $z_{\mathcal{C}}$'s from (3.13).

Denote by $\mathcal{C}^{\mathrm{sing}} \subset \mathcal{C}$ the set of $(p+2)$ nodes, and by $\mathcal{C}^{\mathrm{reg}}$ its complement. We write $T(\mathcal{C}^{\mathrm{reg}}/S)$ for the fibrewise tangent bundle, which is a complex line bundle on $\mathcal{C}^{\mathrm{reg}}$. A fibrewise inhomogeneous term on $\mathcal{C}$ is a complex anti-linear map $\nu_{\mathcal{C}} : T(\mathcal{C}^{\mathrm{reg}}/S) \to TM$, where both bundles involved have been pulled back to $\mathcal{C}^{\mathrm{reg}} \times M$, and with the property that $\nu_{\mathcal{C}}$ is zero outside a compact subset (meaning, in a neighbourhood of $\mathcal{C}^{\mathrm{sing}} \times M \subset \mathcal{C} \times M$). Suppose that we have chosen such a term. One can then consider the moduli space of pairs (v, u), where

$$\begin{aligned} &v \in S, \; u : \mathcal{C}_v \longrightarrow M, \\ &(\bar{\partial}_J u)_z = \nu_{\mathcal{C}_v, z, u(z)}. \end{aligned} \tag{3.14}$$

In the case where $\mathcal{C}_v$ has a node, the second equation is imposed separately on each of its components (with the assumption that both preimages of the node must be mapped to the same point, so as to constitute an actual map on $\mathcal{C}_v$). This makes sense since, near each of the preimages of the node, the equation reduces to the ordinary J-holomorphic curve equation. The incidence conditions are

$$u(z_{\mathcal{C}_v,0}) \in W^u(x_0), \; \dots, \; u(z_{\mathcal{C}_v,p}) \in W^u(x_p),$$

$$u(z_{\mathcal{C}_v,\infty}) \in W^s(x_\infty),\ u(z_{\mathcal{C}_v,*}) \in \Omega. \tag{3.15}$$

Assumption 3.8. We impose the following requirements:

(i) The space of all solutions of (3.14), (3.15) should be regular. This should be understood as two distinct conditions: on the open set of regular values v, regularity holds in the parametrized sense; and for each singular value v, it holds in the ordinary unparametrized sense.

(ii) Take an element (v, u) in the same space, with a simple J-holomorphic bubble attached at an arbitrary point, in the same sense as in Assumption 3.3(ii) (the attaching point can be a marked point, or even the node if v is singular). Then, that moduli space should be regular, as well. As in (i), this should be interpreted as two different conditions, depending whether v is regular or not.

(iii) Consider solutions for regular v, which have a simple holomorphic chain attached at a subset of the $(p+3)$ marked points, and where the incidence constraint has been transferred to the end of that chain, as in Assumption 3.3(iii). Then, the resulting moduli space should again be regular.

(iv) Take a singular v, We look at a situation similar to (iii), but where additionally, there may be a simple holomorphic chain separating the two components of $\mathcal{C}_v$. Let us spell out what that means (ignoring the possible existence of chains at the marked points). Write $z_\pm \in \mathcal{C}_{v,\pm}$ for the preimages of the node. In the definition of the moduli space (3.14), the $\mathcal{C}_{v,\pm}$ carry maps $u_\pm$ which necessarily satisfy $u_-(z_-) = u_+(z_+)$. However, in our limiting situation, we instead have a simple chain $(u_1, \ldots, u_l)$, such that

$$u_-(z_-) = u_1(0),\ u_+(z_+) = u_l(\infty). \tag{3.16}$$

Again, we require that the resulting space should be regular.

As before, given $A \in H_2(M;\mathbb{Z})$, we write $\mathcal{M}_A(\mathcal{C}, x_0, \ldots, x_\infty, \Omega)$ for the space of solutions of (3.14), (3.15) representing A. The added parameter $v \in S$ compensates exactly for the evaluation constraint at $z_{\mathcal{C},*}$, so that we get the same expected dimension as before

$$\dim \mathcal{M}_A(\mathcal{C}, x_0, \ldots, x_p, x_\infty, \Omega) = 2c_1(A) + |x_\infty| - |x_0| - \cdots - |x_p|. \tag{3.17}$$

Concerning the analogue of the stable map compactification, we have a version of Lemma 3.4 (with essentially the same proof):

Lemma 3.9. *(i) If the dimension* (3.17) *is* 0*, we have a finite set*

$$\mathcal{M}_A(\mathcal{C}, x_0, \ldots, x_p, x_\infty, \Omega) = \bar{\mathcal{M}}_A(\mathcal{C}, x_0, \ldots, x_p, x_\infty, \Omega). \tag{3.18}$$

(ii) If the dimension is 1*, the compactification is a manifold with boundary, with the boundary points only involving once-broken gradient flow lines.*

In both cases (i) and (ii), the moduli space and its compactification contain only points where v is a regular value.

We define $m_A(\mathcal{C}, x_0, \dots, x_p, x_\infty, \Omega)$ to be the signed count of points in the zero-dimensional moduli spaces. As before, one can assemble these into a chain map

$$P_{A,\Omega} : \mathrm{CM}^*(f)^{\otimes p+1} \longrightarrow \mathrm{CM}^{*-2c_1(A)}(f). \tag{3.19}$$

Up to chain homotopy, this is independent of the choices of J and $\nu_{\mathcal{C}}$, and also depends only on $[\Omega] \in H^2(M;\mathbb{Z})$.

3c. The remaining topic in this section is the analogue of the divisor axiom. As one would expect, this is not particularly difficult, but requires a bit of technical discussion around forgetting a marked point. For the submanifold Ω, we want to assume that it is transverse to the stable and unstable manifolds of the Morse function.

Lemma 3.10. *In the situation of Lemma 3.4, the following holds generically: any map u in a zero-dimensional space $\mathcal{M}_A(C, x_0, \dots, x_p, x_\infty)$ intersects Ω transversally, and moreover, all those intersections happen away from the marked points. The same is true within the smaller space of those ν_C which vanish close to the marked points.*

This is standard (transversality of evaluation maps). The only wrinkle specific to our case is that the intersections avoid the marked points: but if they did not, we would have an incidence constraint with $\Omega \cap W^s(x)$ or $\Omega \cap W^u(x)$, and those can be ruled out for dimension reasons.

Proposition 3.11. *Fix some A. For suitable choices made in the definitions, the maps (3.12) and (3.19) are related by $P_{A,\Omega} = (A \cdot \Omega)\, S_A$. (For arbitrary choices, the same relation will therefore hold up to chain homotopy.)*

Proof. Even more explicitly, our statement says that one can arrange that

$$m_A(\mathcal{C}, x_0, \dots, x_p, x_\infty, \Omega) = (A \cdot \Omega)\, m_A(C, x_0, \dots, x_p, x_\infty). \tag{3.20}$$

We start with J as in Assumption 3.7, and a ν_C as in Lemma 3.10. Because the inhomogeneous term is zero near the marked points, it can be pulled back to give a fibrewise inhomogeneous term $\nu_{\mathcal{C}}$. To clarify, if $\mathcal{C}_v$ is a singular fibre, then $\nu_{\mathcal{C}_v}$ is supported on $\mathcal{C}_{v,+} \cong C$, and zero on the other component $\mathcal{C}_{v,-}$. Let's consider the structure of the resulting moduli spaces. Given a point in the compactification $\bar{\mathcal{M}}_A(\mathcal{C}, x_0, \dots, x_p, x_\infty, \Omega)$, one can forget the position of the $*$ marked point, and then collapse unstable components (components which are not C, and which carry a constant J-holomorphic map and less than three special points). The outcome is a (continuous) map

$$\bar{\mathcal{M}}_A(\mathcal{C}, x_0, \dots, x_p, x_\infty, \Omega) \longrightarrow \bar{\mathcal{M}}_A(C, x_0, \dots, x_p, x_\infty). \tag{3.21}$$

Now, suppose that the dimension is zero. Then, the target in (3.21) is $\mathcal{M}_A(C, x_0, \dots, x_p, x_\infty)$, and consists only of maps $u : C \to M$ whose intersection points with Ω are not marked points. The preimage of u under (3.21) is necessarily an element of $\mathcal{M}_A(\mathcal{C}, x_0, \dots, x_p, x_\infty, \Omega)$, with v a regular value; such preimages correspond bijectively to points in $u^{-1}(\Omega)$, hence form a finite set, and (because of the transversality condition in Lemma 3.10) are regular points in the parametrized moduli space. Finally, the sign of their contribution to $m_A(\mathcal{C}, x_0, \dots, x_p, x_\infty, \Omega)$ is given by multiplying the contribution of u

to $m_A(C, x_0, \dots, x_p, x_\infty)$ with the local intersection number (sign) of u and Ω at the relevant point.

We have now shown that $\mathcal{M}_A(\mathcal{C}, x_0, \dots, x_p, x_\infty, \Omega) = \bar{\mathcal{M}}_A(\mathcal{C}, x_0, \dots, x_p, x_\infty, \Omega)$ is regular, and that counting points in it exactly yields the right-hand side of (3.20). The $\nu_{\mathcal{C}}$ used for this purpose may not satisfy Assumption 3.8, so this setting is not strictly speaking part of our general definition of $m_A(\mathcal{C}, x_0, \dots, x_p, x_\infty, \Omega)$. However, we can find a small perturbation of $\nu_{\mathcal{C}}$ which does satisfy Assumption 3.8, and points in the associated zero-dimensional moduli spaces will correspond bijectively to those for the original $\nu_{\mathcal{C}}$, because of the compactness and regularity of the original space. □

4. Quantum steenrod operations

This section concerns the operations (1.15) and (1.30). We first set up the various equivariant moduli spaces, then define $Q\Sigma_b$, and discuss its properties. Then we proceed to do the same for $Q\Pi_{a,b}$, and go as far as establishing (1.32).

4a. We equip $C = \mathbb{C}P^1$ with the $(\mathbb{Z}/p)$-action generated by the same rotation as in Sect. 2, but here denoted by σ_C. Fix a compatible almost complex structure J. An equivariant inhomogeneous term ν_C^{eq} is a smooth complex-anti-linear map $TC \to TM$, where both bundles have been pulled back to $S^\infty \times_{\mathbb{Z}/p} C \times M$, and with the same condition of vanishing near the marked points as before. More concretely, one can think of it as a family $\nu_{C,w}^{\mathrm{eq}}$ of inhomogeneous terms (in the standard sense) parametrized by $w \in S^\infty$, with the property that

$$\nu^{\mathrm{eq}}_{C,\tau(w),z,x} = \nu^{\mathrm{eq}}_{C,w,\sigma_C(z),x} \circ D\sigma_z : \mathrm{TC}_z \to \mathrm{TM}_x \quad \text{for } (w,z,x) \in S^\infty \times C \times M. \tag{4.1}$$

Such equivariant data always exist, because the $\mathbb{Z}/p$-action on the space $S^\infty \times C \times M$ is free. Consider the following parametrized moduli problem:

$$\begin{aligned} &w \in S^\infty, \quad u : C \longrightarrow M, \\ &(\bar\partial_J u)_z = \nu^{\mathrm{eq}}_{C,w,z,u(z))}. \end{aligned} \tag{4.2}$$

Note that this inherits a $(\mathbb{Z}/p)$-action, generated by

$$(w,u) \longmapsto (\tau(w), u \circ \sigma_C). \tag{4.3}$$

Fix critical points $x_0, \dots, x_p, x_\infty$, and impose the same incidence constraints as in (3.4) or equivalently (3.5). Moreover, we fix an integer $i \geq 0$ and use that to restrict the parameter w to one of the cells from (2.3), (2.4). More precisely, the condition is that

$$w \in \Delta_i \setminus \partial\Delta_i \subset S^\infty. \tag{4.4}$$

Take solutions of (4.2), (3.4), (4.4) that represent some class $A \in H_2(M;\mathbb{Z})$, and denote the resulting moduli space by $\mathcal{M}_A(\Delta_i \times C, x_0, \dots, x_p, x_\infty)$. The expected dimension increases by the number of parameters

$$\dim \mathcal{M}_A(\Delta_i \times C, x_0, \dots, x_p, x_\infty) = i + 2c_1(A) + |x_\infty| - |x_0| - \cdots - |x_p|. \tag{4.5}$$

Note that while one could define such moduli spaces for more general cells $\tau^j(\Delta_i)$, that is redundant because of (4.3). To express that more precisely, write $(x_1^{(j)}, \ldots, x_p^{(j)})$ for the p-tuple obtained by cyclically permuting $(x_1, \ldots, x_p)$ j times (to the right, so $x_1^{(1)} = x_p$). Then

$$\begin{aligned} &\mathcal{M}_A(\tau^j(\Delta_i) \times C, x_0, \ldots, x_p, x_\infty) \xrightarrow{\cong} \mathcal{M}_A(\Delta_i \times C, x_0, x_1^{(j)}, \ldots, x_p^{(j)}, x_\infty), \\ &(w, u) \longmapsto (\tau^{-j}(w), u \circ \sigma_C^{-j}). \end{aligned} \tag{4.6}$$

There is also a natural compactification, denoted by $\bar{\mathcal{M}}_A(\cdots)$ as usual. This combines the (parametrized) stable map compactification, breaking of Morse flow lines, and instances where the parameter w reaches the boundary of Δ_i.

Lemma 4.1. *For generic J and ν_C^{eq}, the following properties are satisfied.*

(i) If the dimension (4.5) is zero, we get a finite set

$$\mathcal{M}_A(\Delta_i \times C, x_0, \ldots, x_p, x_\infty) = \bar{\mathcal{M}}_A(\Delta_i \times C, x_0, \ldots, x_p, x_\infty). \tag{4.7}$$

(ii) If the dimension is 1, the moduli space is regular, and its compactification is a manifold with boundary. Besides the usual boundary points arising from broken Morse flow lines, one has solutions (w, u) where $w \in \partial\Delta_i$. Using (4.6), the set of such boundary points can be identified with a disjoint union

$$\bigcup_j \mathcal{M}_A(\Delta_{i-1} \times C, x_0, x_1^{(j)}, \ldots, x_p^{(j)}, x_\infty) \;\; \textit{over} \;\; \begin{cases} j = 0, \ldots, p-1 & i \textit{ even}, \\ j = 0, 1 & i \textit{ odd}. \end{cases} \tag{4.8}$$

In (ii), note that the only points $w \in \partial\Delta_i$ that occur lie in the interior of the cells of dimension $(i-1)$. In particular, the fact that the even-dimensional Δ_i have corners can be disregarded. The proof of Lemma 4.1 is simply a parametrized version of that of Lemma 3.4: one imposes Assumption 3.2 on J, and the parametrized analogue of Assumption 3.3 on ν_C^{eq}, where the parameter space is taken to be each $\Delta_i \backslash \partial\Delta_i$. We will not discuss the argument further, and move ahead to its implications.

As usual, we count points in zero-dimensional moduli spaces, and collect those coefficients into

$$\Sigma_A(\Delta_i, \ldots) : \mathrm{CM}^*(f) \otimes \mathrm{CM}^*(f)^{\otimes p} \longrightarrow \mathrm{CM}^{*-i-2c_1(A)}(f). \tag{4.9}$$

Lemma 4.1(ii), with the orientations of the Δ_i taken into account as in (2.7), (2.8), shows that, d being the Morse differential

$$\begin{aligned} &d\Sigma_A(\Delta_i, x_0, \ldots, x_p) - (-1)^i \sum_{j=0}^{p} (-1)^{|x_0|+\cdots+|x_{j-1}|} \Sigma_A(\Delta_i, x_0, \ldots, dx_j, \ldots, x_p) \\ &= \begin{cases} \sum_j (-1)^* \Sigma_A(\Delta_{i-1}, x_0, x_1^{(j)}, \ldots, x_p^{(j)}) & i \text{ even}, \\ (-1)^* \Sigma_A(\Delta_{i-1}, x_0, x_1^{(1)}, \ldots, x_p^{(1)}) - \Sigma_A(\Delta_{i-1}, x_0, x_1, \ldots, x_p) & i \text{ odd}. \end{cases} \end{aligned} \tag{4.10}$$

Here, $(-1)^*$ is the Koszul sign associated with permuting $(x_1, \ldots, x_p)$.

Remark 4.2. Our sign conventions for parametrized pseudo-holomorphic map equations are as follows. Consider, just for the simplicity of notation, operations induced by a Cauchy–Riemann equation on the sphere, with one input and one output. If we have a family of such equations depending on a parameter space Δ which is a manifold with boundary, then the resulting endomorphism of $\mathrm{CM}^*(f)$ satisfies

$$d\phi_\Delta - (-1)^{|\Delta|}\phi_\Delta d = \phi_{\partial\Delta}. \tag{4.11}$$

Note that this differs from the convention in [17, Section 4c]; one can translate between the two by multiplying ϕ_Δ with $(-1)^{|\Delta|(|\Delta|-1)}$.

From now on, we will exclusively work with coefficients in $\mathbb{F} = \mathbb{F}_p$.

Lemma 4.3. *Suppose that b is a Morse cocycle. Then, for each i and A*

$$x \longmapsto (-1)^{|b|\,|x|}\Sigma_A(\Delta_i, x, b, \dots, b) \tag{4.12}$$

is a chain map (an endomorphism of the Morse complex) of degree $p|b| - i - 2c_1(A)$.

This is immediate, by specializing (4.10) to $x_1 = \cdots = x_p = b$. In particular, in this case, the Koszul signs in (4.10) are 1: so for odd i, the expression on the right-hand side vanishes, whereas for even i that expression is $p\Sigma_A(\Delta_{i-1}, x_0, b, \dots, b)$, which vanishes modulo p.

We combine these operations into a series, which is a chain map

$$\begin{aligned}
&\Sigma_{A,b} : \mathrm{CM}^*(f) \longrightarrow (\mathrm{CM}(f)[[t,\theta]])^{*+p|b|-2c_1(A)},\\
&x \longmapsto (-1)^{|b|\,|x|}\sum_k \Big(\Sigma_A(\Delta_{2k}, x, b, \dots, b) + (-1)^{|b|+|x|}\Sigma_A(\Delta_{2k+1}, x, b, \dots, b)\theta\Big)t^k.
\end{aligned} \tag{4.13}$$

One can also sum formally over all A and extend the outcome Λ-linearly

$$\Sigma_b = \sum_A q^A \Sigma_{A,b} : \mathrm{CM}^*(f;\Lambda) \longrightarrow \mathrm{CM}^{*+p|b|}(f;\Lambda). \tag{4.14}$$

Lemma 4.4. *Up to homotopy, (4.13) depends only on cohomology class $[b]$, and moreover, that dependence is linear.*

Proof. Take $\mathrm{CM}^*(f)^{\otimes p}$, with the $\mathbb{Z}/p$-action given by cyclic permutation, and form the associated equivariant complex as in (2.35). Consider the t-linear map

$$\begin{aligned}
&\Sigma_A^{\mathrm{eq}} : \mathrm{CM}^*(f) \otimes (\mathrm{CM}^*(f)^{\otimes p})_{\mathrm{eq}} \longrightarrow (\mathrm{CM}(f)[[t,\theta]])^{*-2c_1(A)},\\
&x_0 \otimes (x_1 \otimes \cdots \otimes x_p) \longmapsto\\
&\qquad \sum_k \Big(\Sigma_A(\Delta_{2k}, x_0, \dots, x_p) + (-1)^{|x_0|+\cdots+|x_p|}\Sigma_A(\Delta_{2k+1}, x_0, \dots, x_p)\theta\Big)t^k,\\
&x_0 \otimes (x_1 \otimes \cdots \otimes x_p)\,\theta \longmapsto \sum_k \Big(\Sigma_A(\Delta_{2k}, x_0, \dots, x_p)\theta\\
&\qquad\qquad - (-1)^{|x_0|+\cdots+|x_p|}\sum_j j(-1)^*\Sigma_A(\Delta_{2k+1}, x_0, x_1^{(j)}, \dots, x_p^{(j)})t\Big)t^k,
\end{aligned} \tag{4.15}$$

where $(-1)^*$ is again the Koszul sign. Equation (4.10), along with (2.7) and (2.8), amounts to saying that (4.15) is a chain map with respect to d_{eq}. As an elementary algebraic consequence, one has the following: if c is any cocycle in $(\mathrm{CM}^*(f)^{\otimes p})_{\mathrm{eq}}$, then

$$x \longmapsto (-1)^{|c|\,|x|}\Sigma_A^{\mathrm{eq}}(x \otimes c) \tag{4.16}$$

is an endomorphism of the chain complex $\mathrm{CF}^*(f)$ of degree $|c| - 2c_1(A)$. The homotopy class of that endomorphism depends only on the cohomology class of c. Moreover, they are additive in c. Applying that construction to $c = b \otimes \cdots \otimes b$ yields precisely (4.13).

From Lemma 2.5, we know that the cohomology class $[b \otimes \cdots \otimes b] \in H^*_{\mathrm{eq}}(\mathrm{CM}^*(f)^{\otimes p})$ only depends on that of $[b]$, which proves our first claim. By the same Lemma, if we use $t(b \otimes \cdots \otimes b)$ instead, the associated operation becomes linear in $[b]$. But that operation is just t times (4.13), so it follows that (4.13) itself must be linear in $[b]$. □

Definition 4.5. For $b \in H^*(M;\mathbb{F}_p)$ and $A \in H_2(M;\mathbb{Z})$, we define the operation $Q\Sigma_{A,b}$ from (1.10) to be the cohomology level map induced by (4.13). Correspondingly, (4.14) is the chain map underlying $Q\Sigma_b$.

Here, we are implicitly using the fact that the chain-level operations are independent of all choices up to chain homotopy. The proof is standard, using moduli spaces with one extra parameter, and will be omitted. Among the previously stated properties of $Q\Sigma$, (1.16) concerns the contribution of the cell Δ_0, which is the operation from Sect. 3, hence is exactly Lemma 3.5. The next two Lemmas correspond to (1.11) and (1.18).

Lemma 4.6. *For $A = 0$, $Q\Sigma_{A,b}$ is the cup product with* $\mathrm{St}(b)$.

Sketch of proof. It will be convenient for this purpose to allow a slightly larger set of choices in the construction. Namely, we choose s-dependent vector fields for s in the relevant half-line given below, parametrising either an "incoming" or "outgoing" flowline, respectively

$$\begin{aligned} &Z_{0,w,s},\dots,Z_{p,w,s} \in C^\infty(TM) \text{ for } w \in S^\infty, s \le 0, \text{ with } Z_{k,w,s} = \nabla f \text{ if } s \ll 0,\\ &Z_{\infty,w,s} \in C^\infty(TM) \text{ for } w \in S^\infty, s \ge 0, \text{ with } Z_{\infty,w,s} = \nabla f \text{ if } s \gg 0. \end{aligned} \tag{4.17}$$

These are used to replace the gradient flow equations in (3.5) by $\mathrm{d}y_k/\mathrm{d}s = Z_{w,k,s}$. The effect is that in the incidence conditions (3.4), the (un)stable manifolds are replaced by perturbed versions. In particular, the transversality of those incidence conditions imposed on pseudo-holomorphic curves can then be achieved by choosing (4.17) generically. This strategy (with minor technical differences) goes back to the Morse-theoretic definition of ordinary Steenrod operations in [2, Section 2]. In [22, Appendix B.1], the iterative procedure to choose such $Z_{k,w,s}$ in a way that one obtains a moduli space cut-out transversely is given in detail, in addition to the fact that such a choice is generic.

We impose an additional symmetry condition, which ensures that (4.6) still holds

$$Z_{k+1,w,s} = Z_{k,\tau(w),s} \quad \text{for } k = 1,\dots,p-1. \tag{4.18}$$

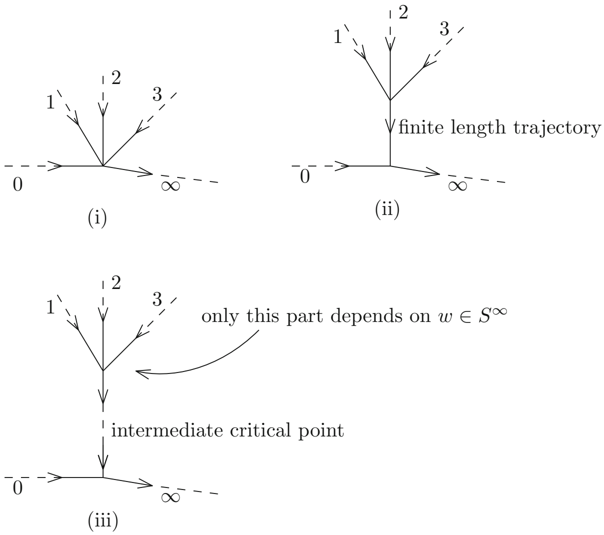


FIGURE 5. A schematic picture of the proof of Lemma 4.6

Considering just $A = 0$, this means that we can take the inhomogeneous term to be zero throughout, so that all maps u are constant (and of course regular). The resulting moduli spaces are purely Morse-theoretical; see Fig. 5(i) for a schematic representation. Without violating the symmetry property (4.6), we can deform our moduli spaces as indicated in Fig. 5(ii). This separates the coincidence condition at the endpoints of the half-flow lines into two parts, joined by a finite length flow line of some other auxiliary s-dependent vector field. More precisely, we use the length as an additional parameter, and all vector fields involved may depend on that. One can arrange that as the length goes to ∞, the limit consists of split solutions as in Fig. 5(iii), where the vector fields on the bottom part are independent on $w \in S^\infty$. It is now straightforward to see that this limit is the combination of the Morse-theoretic cup product and the Morse-theoretic version of the Steenrod operation [2, 4]. □

Lemma 4.7. $Q\Sigma_{A,b}(1)$ *agrees with the* A*-contribution to the quantum Steenrod operation* $Q\mathrm{St}(b)$*, as defined in [8, 22] (for* $p = 2$*) or [18] (all* p*).*

Sketch of proof. Morse-theoretically, 1 is represented by the sum of local minima of the Morse function. Hence, the associated incidence condition (3.4) requires $u(0)$ to lie in an open dense set, and is generically satisfied on every zero-dimensional moduli space. In other words, $\Sigma_{A,b}(1)$ can be computed by forgetting the zero-th marked point and its incidence condition. The outcome is exactly the definition of the quantum Steenrod operation, generalizing the $p = 2$ case from [22] in a straightforward way; compared to the slightly more abstract formulation in [18, Section 9], the only difference is that we stick to a specific cell decomposition of $B\mathbb{Z}/p = S^\infty/(\mathbb{Z}/p)$. □

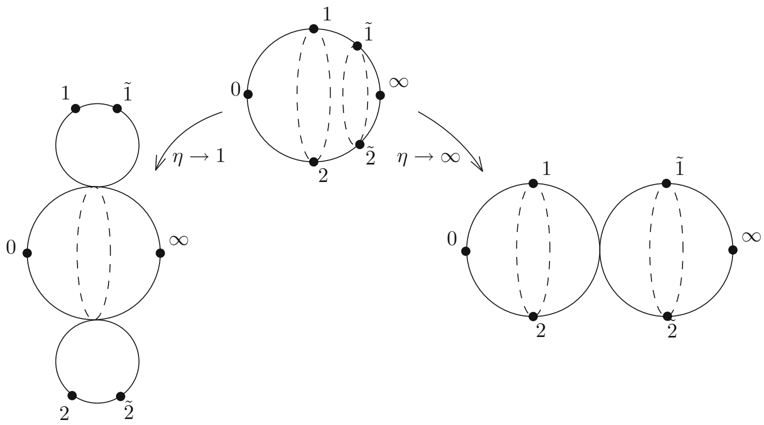


FIGURE 6. The family underlying the proof of Proposition 4.8, for $p = 2$

4b. The final piece of our discussion of $Q\Sigma$ operations concerns (1.20). We assume that the underlying cochain level map Σ_b has been extended to $b \in CM^*(f) \otimes \Lambda$, as in (1.19).

Proposition 4.8. *Fix Morse cocycles b and $\tilde{b}$, and write $\tilde{b} * b \in CM^*(f) \otimes \Lambda$ for a cochain representative of their quantum product. Then, there is a chain homotopy*

$$\Sigma_{\tilde{b}} \circ \Sigma_b \simeq (-1)^{|b|\,|\tilde{b}|\frac{p(p-1)}{2}} \Sigma_{\tilde{b}*b}. \tag{4.19}$$

Sketch of proof. We introduce a family of Riemann surfaces with $(2p + 2)$ marked points, which depends on an additional parameter $\eta \in (1, \infty)$. Each of those surfaces C_η is a copy of C, and the marked points are $z_{C_\eta,k} = z_{C,k}$, $k \in \{0, \dots, p, \infty\}$, from (3.2) together with

$$\tilde{z}_{C_\eta,1} = \eta z_{C,1}, \; \dots, \; \tilde{z}_{C_\eta,p} = \eta z_{C,p}. \tag{4.20}$$

There are natural degenerations at the end of our parameter space: as $\eta \to 1$, each point $\tilde{z}_{C_\eta,k}$ collides with its counterpart $z_{C_\eta,k}$, and one can see this as each pair bubbling off into an extra component of a nodal curve C_1. As $\eta \to \infty$, all the $\tilde{z}_{C_\eta,k}$ collide with $z_{C_\eta,\infty}$, and one can see as degeneration of C_η into a nodal curve C_∞ with two components, each of which is modelled on the original (3.2) (see Fig. 6).

We choose an equivariant inhomogeneous term $\nu^{\mathrm{eq}}_{C_\eta}$ on each of our curves, which is well behaved under the two degenerations (and is zero in a neighbourhood of the nodes and marked points; the details are similar to our previous definition of fibrewise inhomogeneous terms). Given critical points $x_0, x_1, \tilde{x}_1, \dots, x_p, \tilde{x}_p, x_\infty$ of the Morse function f, and a cell Δ_i, we define a moduli space of triples (η, w, u), where: $\eta \in (1, \infty)$, w is as in (4.4), and $u : C_\eta \to M$ is a map, representing the given homology class A, which satisfies the η-parametrized version of (4.2), and the incidence conditions

(3.4) as well as

$$u(\tilde{z}_{C_\eta,1}) \in W^u(\tilde{x}_1),\ \dots,\ u(\tilde{z}_{C_\eta,p}) \in W^u(\tilde{x}_p). \tag{4.21}$$

To understand the algebraic relations which this parametrized moduli space provides, we have to look at the contributions from limits with $\eta = 1$ or $\eta = \infty$. The $\eta = 1$ contribution is given by a suitable moduli space of maps on C_1, and is fairly easy to interpret. Namely, one follows the proof of Lemma 3.5 and separates the components of C_1 by finite length gradient trajectories (to preserve the $\mathbb{Z}/p$-symmetry, all the lengths must be the same, so there is only one length parameter). As the length goes to infinity, the Morse flow lines split, and we end up with a composition of quantum product (of x_k and x_k') and a remaining component where we have the previously defined operation (4.9). We can apply the same strategy to the $\eta = \infty$ limit, inserting a finite length gradient flow line between the two pieces. As the length goes to infinity, we end up with two separate components carrying equations of the kind which underlies (4.9). However, the two equations are coupled, because they carry the same parameter $w \in S^\infty$. In other words, the resulting moduli spaces end up being

$$\bigcup \mathcal{M}_{A_1}(\Delta_i \times C, x_0, \dots, x_p, x) \times_{S^\infty} \mathcal{M}_{A_2}(\Delta_i \times C, x, \tilde{x}_1, \dots, \tilde{x}_p, x_\infty), \tag{4.22}$$

where the (disjoint) union is over $A_1 + A_2 = A$ and all critical points x.

In the same spirit as in (4.9), we denote the operations obtained from (4.22) by

$$\Xi_A(\delta(\Delta_i), \dots) : CM^*(f) \otimes \mathrm{CM}^*(f)^{\otimes 2p} \longrightarrow \mathrm{CM}^{*-i-2c_1(A)}(f). \tag{4.23}$$

We also find it convenient to add up over all A, with the usual q^A coefficients. Fix cocycles b and $\tilde{b}$ and insert them into (4.23) at the marked points labeled $(1, \dots, p)$ and $(\tilde{1}, \dots, \tilde{p})$, respectively, with signs as in (4.13). This yields a chain map

$$\Xi_{\tilde{b},b}(\delta(\Delta_i), \cdot) : \mathrm{CM}^*(f) \longrightarrow (\mathrm{CM}(f) \otimes \Lambda)^{*+p|b|+p|\tilde{b}|}. \tag{4.24}$$

The outcome of the parametrized moduli space argument outlined above is a chain homotopy

$$\Xi_{\tilde{b},b}(\delta(\Delta_i), \cdot) \simeq \Sigma_{\tilde{b}*b}(\Delta_i, \cdot). \tag{4.25}$$

We will be somewhat brief about the final step, since that is a general issue involving equivariant cohomology, and not really specific to our situation. One can construct chain maps like (4.24) not just for $\delta(\Delta_i)$, but for other $\mathbb{F}_p$-coefficient cycles in $S^\infty/(\mathbb{Z}/p) \times S^\infty/(\mathbb{Z}/p)$, such as $\Delta_{i_1} \times \Delta_{i_2}$. In that case, there is a simple decomposition formula

$$\Xi_{\tilde{b},b}(\Delta_{i_1} \times \Delta_{i_2}, \cdot) = (-1)^{|b|\,|\tilde{b}|\,\frac{p(p-1)}{2}} \Xi_{\tilde{b}}(\Delta_{i_1}, \Xi_b(\Delta_{i_2}, \cdot)), \tag{4.26}$$

where the Koszul sign arises from reordering $(\tilde{b}, b, \tilde{b}, b, \dots)$ into $(\tilde{b}, \dots, \tilde{b}, b, \dots, b)$. Finally, homologous cycles give homotopic maps. One can use that, and the decomposition of $\delta(\Delta_i)$ into product cycles from Sect. 2, to obtain a further homotopy

$$\Xi_{\tilde{b},b}(\delta(\Delta_i),\cdot) \simeq \begin{cases} \displaystyle\sum_{i_1+i_2=i} \Xi_{\tilde{b},b}(\Delta_{i_1}\times\Delta_{i_2},\cdot) & \text{if } i \text{ is odd or } p=2, \\ \displaystyle\sum_{\substack{i_1+i_2=i\\ i_k \text{ even}}} \Sigma_{\tilde{b},b}(\Delta_{i_1}\times\Delta_{i_2},\cdot) & \text{if } i \text{ is even and } p>2. \end{cases} \tag{4.27}$$

The combination of (4.25), (4.26), and (4.27) then completes the argument. □

4c. We now merge ideas from Sects. 3 and 4, by which we mean that we take moduli spaces parametrized by cells in $S^\infty/(\mathbb{Z}/p)$, and add an additional freely moving marked point to the domain. The starting point is, once more, the family (3.13). From its construction as a blowup of $C \times S \to S$, this inherits a (diagonal) $(\mathbb{Z}/p)$-action, which we denote by $\sigma_{\mathcal{C}}$.

Fix an almost complex structure J. An equivariant fibrewise inhomogeneous term is a complex anti-linear map

$$\nu^{\mathrm{eq}}_{\mathcal{C}/S} : T(\mathcal{C}^{reg}/S) \longrightarrow TM, \tag{4.28}$$

where both bundles have been pulled back to $S^\infty \times_{\mathbb{Z}/p} \mathcal{C}^{\mathrm{reg}} \times M$. When restricted to any $S^{2k-1} \times_{\mathbb{Z}/p} \mathcal{C}^{\mathrm{reg}} \times M$, it should vanish outside a compact subset (meaning, it is zero in a neighbourhood of $S^{2k-1} \times_{\mathbb{Z}/p} \mathcal{C}^{\mathrm{sing}} \times M$; the restriction to S^{2k-1} follows our usual process of treating S^∞ as a direct limit of finite-dimensional manifolds). As before, one can think of it more explicitly as a family $\nu^{\mathrm{eq}}_{\mathcal{C}/S,w}$ of fibrewise inhomogeneous terms parametrized by $w \in S^\infty$, and satisfying a $(\mathbb{Z}/p)$-equivariance property as in (4.1)

$$\nu^{\mathrm{eq}}_{\mathcal{C}/S,\tau(w),z,x} = \nu^{\mathrm{eq}}_{\mathcal{C}/S,w,\sigma_{\mathcal{C}}(z),x} \circ D\sigma_{\mathcal{C}} : T(\mathcal{C}^{\mathrm{reg}}/S)_z \to \mathrm{TM}_x. \tag{4.29}$$

The associated moduli space consists of triples (w,v,u), where the parameters are $(w,v) \in S^\infty \times S$, v being a regular value of (3.13), and $u : \mathcal{C}_v \to M$ is a solution of the inhomogeneous Cauchy–Riemann equation given by $\nu^{\mathrm{eq}}_{\mathcal{C}_v,w}$. These inherit a $(\mathbb{Z}/p)$-action as in (4.3)

$$(w,v,u) \longmapsto (\tau(w), \sigma^{-1}(v), u \circ \sigma_{\mathcal{C}}). \tag{4.30}$$

We impose the usual incidence conditions, given by the (un)stable manifolds of critical points $x_0,\dots,x_p,x_\infty$, and by a codimension 2 submanifold Ω at the $*$ marked point. Finally, we restrict to the interior of cells (4.4). Denote the resulting moduli spaces by $\mathcal{M}_A(\Delta_i \times \mathcal{C}, x_0,\dots,x_p,x_\infty,\Omega)$. Their expected dimension remains as in (4.5).

We omit the discussion of transversality and of the compactifications, which is simply a combination of those in Sects. 3 and 4. The outcome of isolated-point-counting in our moduli space is maps

$$\Pi_A(\Delta_i,\dots) : \mathrm{CM}^*(f) \otimes \mathrm{CM}^*(f)^{\otimes p} \longrightarrow \mathrm{CM}^{*-i-2c_1(A)}(f) \tag{4.31}$$

which, due to the structure of the compactified one-dimensional moduli spaces, satisfy the same equation as the $\Sigma_A(\Delta_i,\dots)$, see (4.10). Specializing to coefficients in $\mathbb{F}_p$, and fixing a Morse cocycle b, one can therefore use (4.31) to define a chain

$$\Pi_{A,b} : \mathrm{CM}^*(f) \longrightarrow (\mathrm{CM}(f)[[t,\theta]])^{*+p|b|-2c_1(A)} \tag{4.32}$$

exactly as in (4.12). Moreover, up to homotopy that map depends linearly on $[b]$, as in Lemma 4.4. Again up to homotopy, it is also independent of all choices, including that of Ω within its cohomology class $a = [\Omega] \in H^2(M;\mathbb{Z})$.

Definition 4.9. For $a \in H^2(M;\mathbb{Z})$, $b \in H^*(M;\mathbb{F}_p)$ and $A \in H_2(M;\mathbb{Z})$, we define $Q\Pi_{A,a,b}$ to be the cohomology level map induced by (4.32). Adding up those maps with weights q^A yields (1.30).

Proposition 4.10. *Fix some A and integer i. For suitable choices made in the definition, we have $\Pi_A(\Delta_i, \dots) = (A \cdot \Omega)\Sigma_A(\Delta_i, \dots)$. As a consequence, we have $Q\Pi_{A,a,b} = (A \cdot \Omega)Q\Sigma_{A,b}$ for all i and A, which is equivalent to (1.32).*

Proof. The geometric part of this is exactly as in Proposition 3.11: for suitably correlated choices of inhomogeneous terms, the underlying moduli spaces bear the same relationship. Since that argument involves making a small perturbation, we can only apply it to finitely many moduli spaces at once, and that explains the bound on i in the statement. As a consequence, we get equality of the i-th coefficient in $Q\Pi_{A,a,b}$ and $(A \cdot \Omega)Q\Sigma_{A,b}$. □

Remark 4.11. Both in Sect. 3 and here, we have used an evaluation constraint at a codimension two submanifold $\Omega \subset M$, which limits $Q\Pi_{a,b}$ to $a \in H^2(M;\mathbb{Z})$. One can replace that by a pseudo-cycle of arbitrary dimension d (see, e.g., [23]) and, then, the definition goes through without any significant changes for $a \in H^d(M;\mathbb{Z})$. In fact, one could even take a mod p pseudo-cycle. This consists of an oriented manifold with boundary N^d, such that ∂N carries a free $(\mathbb{Z}/p)$-action, and a map $f : N \to M$, such that $f|\partial N$ is $(\mathbb{Z}/p)$-invariant, with the following properties: the limit points of f are contained in the image of a map from a manifold of dimension $(d-2)$, and the limit points of $f|\partial N$ are contained in the image of a map from a manifold of dimension $(d-3)$. While we do not intend to develop the theory of mod p pseudo-cycles here, this should allow one to define $Q\Pi_{a,b}$ for all $a \in H^d(M;\mathbb{F}_p)$. The proof of (1.31) given in the next section extends to such generalizations in a straightforward way, but of course, there is no analogue of (1.32) in codimensions $d > 2$.

5. Proof of Theorem 1.4

This section derives (1.31). Together with the previously established (1.32), that completes our proof of Theorem 1.4.

5a. We decompose the moduli spaces underlying $Q\Pi_{a,b}$ into pieces, where the position of the additional marked point is constrained to lie in one of the cells from Sect. 2. This means that instead of using $\Delta_i \times S \subset S^\infty \times S$ as parameter spaces, we look at the subspaces $\Delta_i \times W$, where

$$W \in \{P_0,\, Q_0,\, \sigma^j(L_1), \sigma^j(B_2)\}. \tag{5.1}$$

Within the framework of Sect. 4, it is unproblematic to ensure that all the resulting moduli spaces, denoted by $\mathcal{M}_A(\Delta_i \times \mathcal{C}|W, x_0, \dots, x_p, x_\infty, \Omega)$, satisfy the usual regularity and compactness properties. Point-counting in them gives rise to maps

$$\Pi_A(\Delta_i \times W, \dots) : \mathrm{CM}^*(f) \otimes \mathrm{CM}^*(f)^{\otimes p} \longrightarrow \mathrm{CM}^{*-i-2c_A(A)-|W|+2}(f). \tag{5.2}$$

As in (4.10), adjacencies between cells determine relations between the associated invariants. In our case, these are governed by (2.7)–(2.8) and (2.19)–(2.21). Explicitly, the relations are

$$\begin{aligned}
&d\Pi_A(\Delta_i \times W, x_0, \dots, x_p) - (-1)^{i+|W|} \\
&\sum_{k=0}^{p} (-1)^{|x_0|+\cdots+|x_{k-1}|} \Pi_A(\Delta_i \times W, x_0, \dots, dx_k, \dots, x_p) \\
&= \begin{cases} \sum_j (-1)^* \Pi_A(\Delta_{i-1} \times \sigma^j W, x_0, x_1^{(j)}, \dots, x_p^{(j)}) & i \text{ even}, \\ (-1)^* \Pi_A(\Delta_{i-1} \times \sigma W, x_0, x_1^{(1)}, \dots, x_p^{(1)}) & \\ -\Pi_A(\Delta_{i-1} \times W, x_0, x_1, \dots, x_p) & i \text{ odd} \end{cases} \\
&\quad + (\text{extra term depending on}\, W).
\end{aligned} \tag{5.3}$$

The last-mentioned term is zero if $W \in \{P_0, Q_0\}$, with the remaining cases being

$$\begin{aligned}
&(\text{extra term for}\, W = \sigma^j L_1) \\
&= (-1)^i \big(\Pi_A(\Delta_i \times Q_0, x_0, x_1^{(j)}, \dots, x_p^{(j)}) - \Pi_A(\Delta_i \times P_0, x_0, x_1^{(j)}, \dots, x_p^{(j)})\big),
\end{aligned} \tag{5.4}$$

$$\begin{aligned}
&(\text{extra term for}\, W = \sigma^j B_2) \\
&= (-1)^{i+1} \big(\Pi_A(\Delta_i \times \sigma^{j+1} L_1, x_0, x_1^{(1)}, \dots, x_p^{(1)}) \\
&\quad - \Pi_A(\Delta_i \times \sigma^j L_1, x_0, x_1, \dots, x_p)\big).
\end{aligned} \tag{5.5}$$

As usual, we now specialize to coefficients in $\mathbb{F} = \mathbb{F}_p$. The relations above immediately imply the following:

Lemma 5.1. *Fix a cocycle $b \in CM^*(f)$. Then, the t-linear map*

$$\begin{aligned}
&\Pi^{eq}_{A,b} : C_{-*}(S)_{\mathrm{eq}} \otimes \mathrm{CM}^*(f) \longrightarrow (\mathrm{CM}(f)[[t,\theta]])^{*+p|b|-2c_1(A)+2}, \\
&W \otimes x \longmapsto (-1)^{|b|\,(|W|+|x|)} \sum_k \Big(\Pi_A(\Delta_{2k} \times W, x, b, \dots, b) \\
&\qquad +(-1)^{|x|+|b|+|W|} \Pi_A(\Delta_{2k+1} \times W, x, b, \dots, b)\theta\Big) t^k, \\
&W\,\theta \otimes x \longmapsto (-1)^{|b|\,(|W|+|x|)} \sum_k \Big((-1)^{|x|} \Pi_A(\Delta_{2k} \times W, x, b, \dots, b)\theta \\
&\qquad -(-1)^{|b|+|W|} \sum_j j \Pi_A(\Delta_{2k+1} \times \sigma^j W, x, b, \dots, b)t\Big) t^k
\end{aligned} \tag{5.6}$$

is a chain map.

Following (4.15), one can think of (5.6) as a special case of a more general structure, which would be a t-linear chain map

$$(C_{-*}(S) \otimes \mathrm{CM}^*(f) \otimes \mathrm{CM}^*(f)^{\otimes p})_{\mathrm{eq}} \longrightarrow (\mathrm{CM}(f)[[t,\theta]])^{*-2c_1(A)+2}. \tag{5.7}$$

Here, the group $\mathbb{Z}/p$ acts on $C_{-*}(S)$, as well as on $\mathrm{CM}^*(f)^{\otimes p}$ by cyclic permutations. As in the previous situation, (5.7) would be useful to prove that

(5.6) only depends on the cohomology class of b, and is additive. For our purposes, however, we can work around that, since all necessary computations can be done using a fixed cocycle b.

5b. At this point, everything we need can be extracted from an analysis of the chain map (5.6).

Lemma 5.2. *Suppose that we specialize (5.6) to using* $W = B_2 + \sigma B_2 + \cdots + \sigma^{p-1} B_2 \in C_2(S)_{eq}$. *Then, the resulting chain map* $\mathrm{CM}^*(f) \to (\mathrm{CM}(f)[[t,\theta]])^{*+p|b|-2c_1(A)}$ *is equal to* $\Pi_{A,b}$.

Proof. This is essentially by definition. We are considering the map

$$x \longmapsto (-1)^{|b|\,|x|} \sum_{j,k} \Big(\Pi_A(\Delta_{2k} \times \sigma^j(B_2), x, b, \dots, b) \\ +(-1)^{|b|+|x|} \Pi_A(\Delta_{2k+1} \times \sigma^j(B_2), x, b, \dots, b)\theta \Big) t^k. \tag{5.8}$$

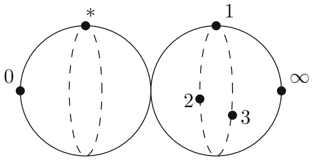
The regularity of the spaces $\mathcal{M}_A(\Delta \times \mathcal{C}|W, x_0, \dots, x_p, x_\infty, \Omega)$ for cells W of dimension < 2 implies that in a zero-dimensional space $\mathcal{M}_A(\Delta \times \mathcal{C}, x_0, \dots, x_p, x_\infty, \Omega)$, none of the points arises from a parameter value $v \in S$ which belongs to one of those cells. In other words, that space $\mathcal{M}_A(\Delta \times \mathcal{C}, x_0, \dots, x_p, x_\infty, \Omega)$ is the disjoint union of $\mathcal{M}_A(\Delta \times \mathcal{C}|\sigma^j B, x_0, \dots, x_p, x_\infty, \Omega)$. □

Lemma 5.3. *Suppose that we specialize (5.6) to using* $W = P_0 \in C_0(S)_{eq}$, *and pass to cohomology. Then, the resulting map is equal to the following: take all possible decompositions* $A = A_1 + A_2$, *and add up*

$$H^*(M;\mathbb{F}_p) \xrightarrow{*_{A_1} a} H^{*+2-2c_1(A_1)}(M;\mathbb{F}_p) \xrightarrow{Q\Sigma_{b,A_2}} (H(M;\mathbb{F}_p)[[t,\theta]])^{*+p|b|+2-2c_1(A)}, \tag{5.9}$$

where $a = [\Omega] \in H^2(M;\mathbb{Z})$.

Proof. This time, the reason is geometric. Using P_0 means that we are restricting to a particular fibre of (3.13), which is the nodal surface from Fig.

(i)


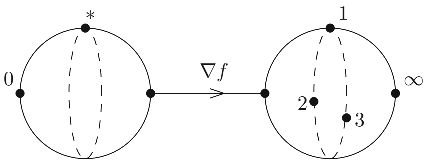
(ii)


FIGURE 7. A schematic picture of the proof of Lemma 5.3, with $p = 3$

7(i). Recall that each component of that surface carries an inhomogeneous term, which additionally depends on parameters in S^∞. However, without violating regularity or other restrictions, one can arrange that the inhomogeneous term on the component which is a three-pointed sphere ($\mathcal{C}_{0,-}$ in the notation from Sect. 3) is independent of those parameters.

After that, one inserts a finite length Morse flow line between the two components, as in Fig. 7(ii). In the same way as in Lemma 3.5, the resulting (varying length) moduli space gives a chain homotopy between our operation and the chain map underlying the composition (5.9), in its Morse-theoretic incarnation. □

Lemma 5.4. *Suppose that we specialize* (5.6) *to using* $W = Q_0 \in C_0(S)_{eq}$, *and pass to cohomology Then, the resulting map is equal to the following: take all possible decompositions* $A = A_1 + A_2$, *and add up*

$$H^*(M;\mathbb{F}_p) \xrightarrow{Q\Sigma_{b,A_1}} H(M;\mathbb{F}_p)[[t,\theta]])^{*+p|b|-2c_1(A_1)} \xrightarrow{*_{A_2} a} (H(M;\mathbb{F}_p)[[t,\theta]])^{*+p|b|+2-2c_1(A)}, \tag{5.10}$$

where $a = [\Omega]$ *as before.*

The proof is the same as for Lemma 5.3. Note that the operations in (5.10) appear in the opposite order from (5.9). The reason is that over $v = 0$, the component $\mathcal{C}_{y,-}$ is attached to $\mathcal{C}_{v,+}$ at the point $0 \in C$, which serves as input of the Σ operation; whereas for $v = \infty$, it is attached at the output point $\infty \in C$. Finally, we have the following, which establishes (1.31):

Proposition 5.5. $tQ\Pi_{a,b}$ *equals the difference between* (5.9) *and* (5.10).

Proof. By Lemma 5.2, $\Pi_{A,b}t$ is obtained by specializing (5.6) to $(B_2 + \cdots + \sigma^{p-1}B_2)t$. From (2.52) and (2.53), we see that this is chain homotopic to specializing the same map to $(P_0 - Q_0)$. Using Lemma 5.3 and 5.4 then yields the desired result. □

6. Computations

In this section, we explore the power of Theorem 1.4 as a computational tool.
6a. Our first task is to work out the details of Example 1.6, where M is the two-sphere. We use the standard generator of $H_2(M;\mathbb{Z})$, and correspondingly write Λ as a power series ring in one variable q. The quantum connection is

$$\nabla = tq\partial_q + \begin{pmatrix} 0 & q \\ 1 & 0 \end{pmatrix}. \tag{6.1}$$

Let us temporarily use $\mathbb{Q}$-coefficients, and allow inverses of t. If ξ satisfies

$$(tq\partial_q)^2\xi = q\xi, \tag{6.2}$$

then the following endomorphism is covariantly constant with respect to (6.1):

$$\Xi = \begin{pmatrix} -\xi(tq\partial_q\xi) & -(tq\partial_q\xi)^2 \\ \xi^2 & \xi(tq\partial_q\xi) \end{pmatrix}. \tag{6.3}$$

It is straightforward to write down an explicit solution of (6.2):

$$\xi = \sum_{k=0}^{\infty} \frac{1}{(k!)^2} q^k t^{-2k}. \tag{6.4}$$

Pick a prime $p > 2$. Take (6.3) with (6.4), and truncate it by dropping all powers q^p or higher. The remaining denominators are coprime to p, so we can reduce coefficients to $\mathbb{F}_p$. The outcome, using some elementary combinatorics to simplify the formulae, is the matrix Σ from (1.26). For example, the $q^k t^{-2k}$ term of the σ_{21} coordinate of (6.3) is

$$\sum_{k=k_1+k_2} \tfrac{1}{(k_1!)^2} \tfrac{1}{(k_2!)^2} = \frac{1}{(k!)^2} \sum_{k=k_1+k_2} \binom{k}{k_1}\binom{k}{k_2} = \frac{1}{(k!)^2}\binom{2k}{k},$$

the second equality being the Chu–Vandermonde identity. We notice that this is the σ_{21} component $\frac{(2k)!}{(k!)^4}$ of (1.26). Similarly, the coefficient of $q^k t^{2-2k}$ in the σ_{12} component of (6.3) is

$$\sum_{k=k_1+k_2} \tfrac{k_1}{(k_1!)^2} \tfrac{k_2}{(k_2!)^2},$$

and using the Chu–Vandermonde identity on

$$\tfrac{1}{(k-1)!^2} \sum_{k=k_1+k_2} \binom{k-1}{k_1}\binom{k-1}{k_2},$$

one obtains the coefficient of $q^k t^{2-2k}$ in the σ_{12} component of (1.26). A similar application of this identity can be used for the σ_{11} component.

By construction, this endomorphism is covariantly constant modulo q^p; and the constant term (in q) of $-t^{p-1}\Sigma$ matches the cup product with $\mathrm{St}(h) = -t^{p-1}h$ (see (1.13) for the sign convention). Therefore, $-t^{p-1}\Sigma$ and $Q\Sigma_h$ must agree modulo q^p. However, for degree reasons, $Q\Sigma_h$ cannot have terms of order q^p or higher. The consequence is that $Q\Sigma_h = -t^{p-1}\Sigma$, as previously stated.

Remark 6.1. It is worthwhile spelling out the comparison with the fundamental solution of the quantum differential equation, mentioned in Remark 1.5. For S^2, the fundamental solution is [10, Section 28.2] (note the differences in notation and conventions: our t is their $-\hbar$; our q is their e^t; our t is their H)

$$\Psi = \begin{pmatrix} -tq\partial_q \eta & -tq\partial_q \xi \\ \eta & \xi \end{pmatrix}, \tag{6.5}$$

where ξ is as in (6.4), and

$$\eta = \sum_{k=0}^{\infty} \frac{1}{(k!)^2} q^k t^{-2k-1}\big(-\log(q) + 2\sum_{j=1}^{k}\frac{1}{j}\big) \tag{6.6}$$

is a multivalued solution of the same Eq. (6.2) as ξ. By forming (1.23) with $\beta = h$, one gets exactly the matrix from (6.3)

$$\Xi = \Psi \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \Psi^{-1}. \tag{6.7}$$

6b. Following ideas from [22], let us look at the following situation:

Assumption 6.2. The second cohomology group $H^2(M;\mathbb{F}_p)$ generates $H^*(M;\Lambda)$ as a ring, with the quantum product.

This implies that $H^*(M;\mathbb{F}_p)$ is zero in odd degrees. It also implies that each class in $H^2(M;\mathbb{F}_p)$ can be lifted to $H^2(M;\mathbb{Z})$, as one sees by looking at

$$\begin{aligned}\cdots \to H^2(M;\mathbb{Z}) &\to H^2(M;\mathbb{F}_p) \to H^3(M;\mathbb{Z}) \xrightarrow{p} H^3(M;\mathbb{Z})\\ &\to H^3(M;\mathbb{F}_p) \to \cdots \end{aligned} \tag{6.8}$$

Lemma 6.3. *Suppose that Assumption 6.2 holds. Then, the quantum product and* $\mathrm{QSt}(b)$*, for* $b \in H^2(M;\mathbb{F}_p)$*, determine all the quantum Steenrod operations.*

Proof. Write the covariant constancy property as

$$Q\Sigma_b(a * c) = t\partial_a Q\Sigma_b(c) + a * Q\Sigma_b(c), \quad a \in H^2(M;\mathbb{Z}), \quad b, c \in H^*(M;\mathbb{F}_p). \tag{6.9}$$

This shows that $Q\Sigma_b(c)$ and the quantum product determine $Q\Sigma_b(a*c)$. Therefore, if one knows $\mathrm{QSt}(b) = Q\Sigma_b(1)$ and Assumption 6.3 holds, the entire operation $Q\Sigma_b$ can be computed from that. By (1.20)

$$\mathrm{QSt}(b * c) = Q\Sigma_b(\mathrm{QSt}(c)). \tag{6.10}$$

If we know $\mathrm{QSt}(b)$ and $\mathrm{QSt}(c)$, for some $b \in H^2(M;\mathbb{F}_p)$ and $c \in H^*(M;\mathbb{F}_p)$, then our previous argument determines $Q\Sigma_b$, and we can get $\mathrm{QSt}(b*c)$ from that by (6.10). In view of Assumption 6.3, this implies the desired result. □

Here is a concrete class of examples to which this strategy applies.

Proposition 6.4. *Suppose that* M *is a monotone symplectic manifold, satisfying Assumption 6.2. Then, the quantum Steenrod operations can be computed in terms of the quantum product and classical Steenrod operations.*

Proof. Take $b \in H^2(M;\mathbb{F}_p)$. Then, $\mathrm{QSt}(b)$ has degree $2p$. The monomials in it that can have nonzero coefficients are $t^j q^A$, where $j + c_1(A) \leq p$. The terms with $j = 0$ and $c_1(A) = p$ are part of (1.16). The remaining terms are determined by covariant constancy, since any monomial q^A that lies in I_{diff} must necessarily have $c_1(A) \geq p$. Having determined $\mathrm{QSt}(b)$, Lemma 6.3 does the rest. □

As a concrete illustration, let us consider a cubic surface $M \subset \mathbb{C}P^3$, which is a del Pezzo surface, and hence a monotone symplectic manifold. For simplicity, instead of the whole Novikov ring, we will work with a single Novikov variable q, which counts the Chern number of holomorphic curves. Let us first take coefficients in $\mathbb{Z}$. Take h_2 to be the first Chern class of M, and h_4 to be the Poincaré dual of a point. Computations in [5,6] show that

$$\begin{aligned} h_2 * h_2 &= 3h_4 + 9q\, h_2 + 108q^2, \\ h_2 * h_4 &= 36q^2\, h_2 + 252q^3. \end{aligned} \tag{6.11}$$

At one point, we will use another class in $H^2(M)$, the Poincaré dual of a Lagrangian sphere, denoted by l_2. This satisfies

$$l_2 * l_2 = -2h_4 + 4q\, h_2 + 12q^2. \tag{6.12}$$

Example 6.5. Take the cubic surface with $p = 2$ (this computation is of the same kind as those in [22], only expressed in slightly different language). First of all

$$\mathrm{QSt}(c) = c * c + tc \;\text{ for all } c \in H^2(M;\mathbb{F}_2). \tag{6.13}$$

A priori, $\mathrm{QSt}(c)$ could also have a tq term, which would lie in $H^0(M;\mathbb{F}_2)$. This would come from classes with $c_1(A) = 1$. To get a nonzero output in $H^0(M;\mathbb{F}_2)$, one would need to have a stable A-curve going through every point of M. However, each A is represented by a unique embedded (-1)-sphere; hence, the term must vanish, leaving (6.13).

By combining (6.11), (6.13), and (6.9), one gets

$$\begin{aligned}\mathrm{QSt}(h_2) &= h_4 + (q+t)h_2,\\ Q\Sigma_{h_2}(h_2) &= tq\partial_q \mathrm{QSt}(h_2) + h_2 * \mathrm{QSt}(h_2) = (q+t)h_4 + q^2 h_2,\\ Q\Sigma_{h_2}(h_2 * h_2) &= tq\partial_q Q\Sigma_{h_2}(h_2) + h_2 * Q\Sigma_{h_2}(h_2) = (tq+q^2)h_4 + q^3 h_2.\end{aligned} \tag{6.14}$$

Using (6.10), we get the result announced in Example 1.7

$$\begin{aligned}\mathrm{QSt}(h_4) &= \mathrm{QSt}(h_2 * h_2 + qh_2) = Q\Sigma_{h_2}(\mathrm{QSt}(h_2)) + q^2\mathrm{QSt}(h_2)\\ &= Q\Sigma_{h_2}(h_2 * h_2 + th_2) + q^2 \mathrm{QSt}(h_2) = t^2 h_4.\end{aligned} \tag{6.15}$$

Example 6.6. Let us again look at the cubic surface, but now with $p = 3$. Here, the fact that we work with a single Novikov variable q will limit the effectiveness of our computation, leading to an incomplete result. As explained in Proposition 6.4, we can use covariant constancy to determine the quantum Steenrod operations on $H^2(M;\mathbb{F}_3)$. In the same way, one can compute $Q\Sigma_b(c)$ for $b, c \in H^2(M;\mathbb{F}_3)$ except for the $q^3 t$ term, which lies in $H^0(M;\mathbb{F}_3)$. We will only describe the outcome (code that carries out this computation is available at [19])

$$\begin{aligned}\mathrm{QSt}(h_2) &= -t^2 h_2,\\ \mathrm{QSt}(l_2) &= -t^2 l_2,\\ Q\Sigma_{l_2}(l_2) &= -t^2 h_4 + (\text{term lying in } H^0(M;\mathbb{F}_3) q^3 t).\end{aligned} \tag{6.16}$$

From that one gets, using (6.12)

$$\begin{aligned}\mathrm{QSt}(h_4) &= \mathrm{QSt}(l_2 * l_2 - qh_2) = Q\Sigma_{l_2}(\mathrm{QSt}(l_2)) - q^3 \mathrm{QSt}(h_2)\\ &= t^4 h_4 + q^3 t^2 h_2 + (\text{term lying in } H^0(M;\mathbb{F}_3) q^3 t^3).\end{aligned} \tag{6.17}$$

Note that, unlike the $p = 2$ case, $\mathrm{QSt}(h_4)$ contains a non-classical (quantum) term.

6c. We conclude our discussion with a higher dimensional case: the intersection of two quadrics in $\mathbb{C}P^5$, which is a monotone symplectic 6-manifold. Let us first work with $\mathbb{Z}$-coefficients. The even degree cohomology has a basis $\{1, h_2, h_4, h_6\}$, where the subscript denotes the dimension. There is also odd degree cohomology, $H^3(M;\mathbb{Z}) = \mathbb{Z}^4$, but that will play no role in our argument. We can identify the Novikov ring with $\mathbb{Z}[[q]]$, but since $c_1(M)$ is twice the positive area generator of $H^2(M;\mathbb{Z})$, the formal variable q has degree 4. The quantum product, as computed in [7], satisfies

$$\begin{aligned} h_2 * h_2 &= 4(h_4 + q), \\ h_2 * h_4 &= h_6 + 2qh_2, \\ h_2 * h_6 &= 4qh_4 + 4q^2, \\ h_4 * h_4 &= 2qh_4 + 3q^2. \end{aligned} \tag{6.18}$$

Example 6.7. Taking our intersection of quadrics, let us set $p = 2$. The classical Steenrod operations are

$$Sq(h_k) = t^{k/2} h_k. \tag{6.19}$$

For h_2, this is because $h_2^2 = 0 \in H^4(M;\mathbb{F}_2)$, which one can read off from the classical term in (6.18). For h_4, its Poincaré dual of is represented by a line $\mathbb{C}P^1 \subset M$. The normal bundle of that line has first Chern class 0; by the geometric description of Steenrod squares through Stiefel–Whitney classes, this implies vanishing of $\mathrm{Sq}^2(h_4)$.

Since the quantum product with h_2 agrees with its classical counterpart, the cup product with any element of $H^*(M;\mathbb{F}_2)$ is a covariantly constant endomorphism for the quantum connection. From that, Theorem 1.4, and (6.19), one gets

$$\begin{aligned} &Q\Sigma_{h_2}(c) = th_2 c + (\text{terms lying in}\, H^k(M;\mathbb{F}_2) \text{ with } k < |c| - 4), \\ &Q\Sigma_{h_4}(c) = t^2 h_4 c + q^2 c + (\text{terms lying in}\, H^k(M;\mathbb{F}_2) \text{ with } k < |c|). \end{aligned} \tag{6.20}$$

Therefore

$$\begin{aligned} &\mathrm{QSt}(h_2) = th_2, \\ &\mathrm{QSt}(h_4) = t^2 h_4 + q^2, \\ &\mathrm{QSt}(h_6) = Q\Sigma_{h_2 * h_4}(1) = Q\Sigma_{h_2}(\mathrm{QSt}(h_4)) = Q\Sigma_{h_2}(t^2 h_4 + q^2) = t^3 h_6 + q^2 t h_2. \end{aligned} \tag{6.21}$$

Example 6.8. Still for our intersection of quadrics, take $p = 3$. Then, the quantum product and covariant constancy completely determine $Q\Sigma_{h_2}$, for degree reasons (in fact, the same is true for any $p > 2$). Explicitly (see again [19] for code), the action on $H^{even}(M;\mathbb{F}_3)$ is

$$Q\Sigma_{h_2} = \begin{pmatrix} qt & q^2 & -q^2 t & q^3 \\ -t^2 & qt & 0 & q^2 t \\ 0 & -t^2 + q & -qt & q^2 \\ 1 & 0 & -t^2 & -qt \end{pmatrix}. \tag{6.22}$$

From that, we get

$$\begin{aligned} \mathrm{QSt}(h_2) &= Q\Sigma_{h_2}(1) = qt\,1 - t^2\,h_2 + h_6, \\ \mathrm{QSt}(h_4) &= Q\Sigma_{h_4}(1) = Q\Sigma_{h_2 * h_2 - q1}(1) = Q\Sigma_{h_2}^2(1) - q^3\,1 \\ &= qt(q + t^2)\,h_2 + (q + t^2)^2\,h_4, \\ \mathrm{QSt}(h_6) &= Q\Sigma_{h_6}(1) = Q\Sigma_{h_2 * h_2 * h_2}(1) = Q\Sigma_{h_2}^3(1) \\ &= q^2 t(q^2 - qt^2 - t^4)\,1 + q^2 t^4\,h_2 + qt^3(q + t^2)\,h_4 \end{aligned}$$

$$+(q^3 - q^2t^2 + qt^4 - t^6)h_6.$$

Acknowledgements

Both authors were partially supported by a Simons Investigator award from the Simons Foundation. Additional support for the first author was provided by the Simons Collaboration for Homological Mirror Symmetry, and by NSF grant DMS-1904997. The second author was additionally supported by a Heilbronn Research Fellowship.

Funding Information Open Access funding provided by the MIT Libraries

References

[1] Albers, P.: A Lagrangian Piunikhin–Salamon–Schwarz morphism and two comparison homomorphisms in Floer homology. Int. Math. Res. Notes (IMRN) (2008) (art. ID rnm134)

[2] Betz, M., Cohen, R.: Graph moduli spaces and cohomology operations. Turkish J. Math. **18**, 23–41 (1994)

[3] Cineli, E., Ginzburg, V., Gurel, B.: From pseudo-rotations to holomorphic curves via quantum Steenrod squares. Int. Math. Res. Notes (IMRN) 2274–2297 (2022)

[4] Cohen, R., Norbury, P.: Morse field theory. Asian J. Math. **16**, 661–711 (2012)

[5] Crauder, B., Miranda, R.: Quantum cohomology of rational surfaces. In: The moduli space of curves (Texel Island, 1994), pp. 33–80. (Birkhäuser, 1995)

[6] Di Francesco, P., Itzykson, C.: Quantum intersection rings. In: The moduli space of curves (Texel Island, 1994), pp. 81–148. (Birkhäuser, 1995)

[7] Donaldson, S.K.: Floer homology and algebraic geometry. In Vector bundles in algebraic geometry, pp. 119–138. (Cambridge University Press, Cambridge, 1993)

[8] Fukaya, K.: Morse homotopy and its quantization. In: Geometric topology (Athens, GA, 1993), pp. 409–440. (American Mathematics Society, New York, 1997)

[9] Hofer, H., Salamon, D.: Floer homology and Novikov rings. In: The Floer memorial, pp. 483–524. (Birkhauser, 1995)

[10] Hori, K., Katz, S., Klemm, A., Pandharipande, R., Thomas, R., Vafa, C., Vakil, R., Zaslow, E.: Mirror symmetry. (American Mathematics Society/Clay Mathematics Institute, 2003)

[11] Katić, J.: Compactification of mixed moduli spaces in Morse-Floer theory. Rocky Mt. J. Math. **38**(3), 923–939 (2008)

[12] Katić, J., Milinković, D.: Piunikhin–Salamon–Schwarz isomorphisms for Lagrangian intersections. Differ. Geom. Appl. **22**(2), 215–227 (2005)

[13] Lu, G.: An explicit isomorphism between Floer homology and quantum homology. Pac. J. Math. **213**(2), 319–363 (2004)

[14] McDuff, D., Salamon, D.: J-holomorphic curves and symplectic topology. (American Mathematics Society, New York, 2004)

[15] Pandharipande, R.: Rational curves on hypersurfaces (after A. Givental). In: Séminaire Bourbaki 1997/98, Astérisque 252, pp. 307–340 (1998)

[16] Schwarz, M.: A quantum cup-length estimate for symplectic fixed points. Invent. Math. **133**(2), 353–397 (1998)

[17] Seidel, P.: Connections on equivariant Hamiltonian Floer cohomology. Comment. Math. Helv. **93**, 587–644 (2018)

[18] Seidel, P.: Formal groups and quantum cohomology. Geom. Topol. (to appear)

[19] http://math.mit.edu/~seidel/code.html. Accessed 20 May 2022

[20] Shelukhin, E.: Pseudorotations and Steenrod squares. J. Mod. Dyn. **16**, 289–304 (2020)

[21] Shelukhin, E.: Pseudorotations and Steenrod squares revisited. Math. Res. Lett. **28**, 1255–1261 (2021)

[22] Wilkins, N.: A construction of the quantum Steenrod squares and their algebraic relations. Geom. Topol. **24**, 885–970 (2020)

[23] Zinger, A.: Pseudocycles and integral homology. Preprint arxiv:math/0605535

Paul Seidel and Nicholas Wilkins
Department of Mathematics
MIT
77 Massachusetts Ave
Cambridge MA02139
USA
e-mail: `pseidel@mit.edu`;
`nwilkins@mit.edu`

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:25
https://doi.org/10.1007/s11784-022-00958-5
Published online April 5, 2022

Journal of Fixed Point Theory and Applications

An algebraic approach to the algebraic Weinstein conjecture

Vivek Shende

Abstract. How does one measure the failure of Hochschild homology to commute with colimits? Here, I relate this question to a major open problem about dynamics in contact manifolds—the assertion that Reeb orbits exist and are detected by symplectic homology. More precisely, I show that for polarizably Weinstein fillable contact manifolds, said property is equivalent to the failure of Hochschild homology to commute with certain colimits of representation categories of tree quivers. So as to be intelligible to algebraists, I try to include or black-box as much of the geometric background as possible.

Mathematics Subject Classification:. 53D37, 53D40.

Existence of closed geodesics on a compact Riemannian manifold M is guaranteed for topological reasons [14, 25]. Let us recall the argument. Morse theory tells us that the homology of the free loop space $\mathcal{L}M = \mathrm{Maps}(S^1 \to M)$ can be computed by a complex generated by geodesics. The trivial loops contribute a subcomplex computing the homology of the original manifold, so there must be nontrivial loops unless $\mathrm{H}_\bullet(M) \to \mathrm{H}_\bullet(\mathcal{L}M)$ is an isomorphism. It is obviously not an isomorphism unless M is simply connected, and in this case, we study the based loop space $\Omega M := \mathrm{Maps}((S^1, 0) \to (M, m))$ and the fibration $\Omega M \to \mathcal{L}M \to M$. As this is split by constant loops $M \to \mathcal{L}M$, we find $\pi_k(\mathcal{L}M) = \pi_k(M) \oplus \pi_{k+1}(M)$, so by Hurewicz, the first nontrivial homotopy group $\pi_{k+1}(M)$ contributes nontrivially to $\mathrm{H}_k(\mathcal{L}M)$, while $\mathrm{H}_k(M)$ vanishes. In fact, in the simply connected case, one can obtain rather more refined information [41].

We would like to think of the map $\mathrm{H}_\bullet(M) \to \mathrm{H}_\bullet(\mathcal{L}M)$ as arising from the following local-to-global construction. We write Ω for the constant cosheaf of spaces over M, with costalk a point. By definition, this assigns a point

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

to any contractible open set, and sends covers to colimits.[1] By the (∞) van Kampen theorem, $\Omega(U)$ is the path groupoid of U, explaining the notation. Composition of loops gives $\mathrm{H}_\bullet(\Omega(M))$ the structure of a ring, and its Hochschild homology gives the homology of the free loop space [7,20]: $\mathrm{HH}_\bullet(\mathrm{H}_\bullet(\Omega M)) = \mathrm{H}_\bullet(\mathcal{L}M)$. The inclusion of constant loops is

$$\mathrm{H}_\bullet(M) = \mathrm{colim}_U \, \mathrm{HH}_\bullet(\mathrm{H}_\bullet(\Omega(U))) \to \mathrm{HH}_\bullet(\mathrm{colim}_U \, \mathrm{H}_\bullet(\Omega(U))) = \mathrm{H}_\bullet(\mathcal{L}M). \tag{1}$$

The following is the first example of the problem we are interested in:

Question 1. Can the failure of (1) to be an isomorphism be seen in terms of some general machinery measuring the failure of Hochschild homology to commute with homotopy colimit?

We turn to contact geometry. The contact-geometric formulation of the geodesic flow is the following. On a cotangent bundle T^*M, there is the tautological 1-form λ, which at a given covector ξ is the function on tangent vectors given by ξ. Fixing a metric on M, we may restrict λ to the unit cosphere bundle S^*M; here, it is *contact*, meaning that $\lambda \wedge (\mathrm{d}\lambda^{n-1})$ is nowhere vanishing. The *Reeb* vector field R on S^*M is characterized by lying in the kernel of $d\lambda$ and normalized by $\lambda(R) = 1$. The metric identifies S^*M with the unit sphere bundle in TM, and the Reeb flow with geodesic flow.

More generally, the same formulas define the Reeb flow for any contact form on any manifold V. The Weinstein conjecture asserts the existence of a closed trajectory [46]. It has long been a central problem in contact geometry [21–23,43], and known in general only in dimension 3, by an argument whose ingredients have no known analogue in higher dimension [40]. The result is also known for flexible contact structures in general [4]; we will be interested here in what is in some sense the opposite setting [31], of Weinstein fillable contact manifolds.

It is natural to try and generalize the Morse theoretic approach to geodesics to the study of Reeb orbits. That is, one wants a complex generated by orbits, so that nonvanishing of the homology groups implies the existence of orbits. The reason to impose a differential is to provide invariance under deformations: these homologies depend only on $\ker \lambda$ rather than λ itself, i.e., on the *contact structure* rather than the *contact form*.

Just as the cosphere bundle bounds the codisk bundle, we may ask that some general contact (V, λ) is the boundary of some W to which λ extends, and over which $(d\lambda)^n$ is everywhere nonvanishing. We also ask that the 'Liouville' vector field Z characterized by $\mathrm{d}\lambda(Z, \cdot) = \lambda$ points out at the boundary; in the case of the cotangent bundle, this vector field is radial in the fiber directions. Such W are called Liouville domains, and determine a symplectic cohomology $\mathrm{SH}^\bullet(W)$, which may be taken to be generated by the Reeb orbits of (V, λ) and the critical points of a Morse function on W [45].

[1] We abuse terminology to respect intuition and write 'homology' when we mean that we think of the complex as an object in the derived category. Similarly, we write '=' to mean, e.g., identified by a canonical quasi-isomorphism, etc. We always work with ∞-categories, etc., and (pre)sheaves or cosheaves should be understood accordingly.

(See [33] for a leisurely introduction to Liouville manifolds and symplectic cohomology.) The differential is such that these Morse critical points form a subcomplex on which the differential is the Morse differential, giving an exact triangle[2]

$$\mathrm{H}^\bullet(W) \to \mathrm{SH}^\bullet(W) \to \mathrm{SH}^\bullet_+(W) \xrightarrow{[1]} . \tag{2}$$

In particular, if $\mathrm{H}^\bullet(W) \to \mathrm{SH}^\bullet(W)$ fails to be an isomorphism, then the Weinstein conjecture holds for V, as $\mathrm{SH}^\bullet_+(W)$, which is generated by the Reeb orbits of V, must be nonzero.

Viterbo's *algebraic Weinstein conjecture* is the assertion that $\mathrm{SH}^\bullet_+(W)$ always detects an orbit. It is not typically easy to compute $\mathrm{SH}^\bullet(W)$. But when $W = T^*M$, one knows:

Theorem 2. [42,44] *There is an isomorphism* $\mathrm{SH}^{\bullet+n}(T^*M) \cong \mathrm{H}_{-\bullet}(\mathcal{L}M)$.[3]

This isomorphism has seen many further developments; see, e.g., [3,10].

The composition $\mathrm{H}^{\bullet+n}(T^*M) \to \mathrm{SH}^{\bullet+n}(T^*M) \cong \mathrm{H}_{-\bullet}(\mathcal{L}M)$ is identified with the inclusion of constant loops $\mathrm{H}_{-\bullet}(M) \to \mathrm{H}_{-\bullet}(\mathcal{L}M)$ under Poincaré duality [2, Lem. 3.6]. Therefore, the symplectic homology detects geodesics in essentially the same way as the Morse homology of the loop space did. However, already in this case it does more: it shows that a contact level of T^*M, not necessarily the unit cosphere bundle for any Riemannian metric, will also have Reeb orbits.

A class of Liouville domains including but rather more general than codisk bundles are the *Weinstein* domains.[4] By definition, these are those for which the Liouville vector field is gradient-like for a Morse–Bott function. In this case, the critical points of Z have index $\leq \dim W/2$, and union of descending level sets is a singular isotropic subset termed the *core* or *skeleton*. Stein domains from complex analysis are Weinstein when viewed as symplectic manifolds, and conversely, any Weinstein domain is deformation equivalent to a Stein domain [9].

Weinstein domains for which the indices of critical points are $< \dim W/2$ are said to be *subcritical*, and for these, it is known that $\mathrm{SH}^\bullet(W) = 0$; in particular, the Weinstein conjecture holds for their contact boundaries [42]. Simple examples: the ball; the cotangent bundle of a noncompact manifold. Beyond these, the Weinstein conjecture is not known for contact boundaries of Weinstein domains in any reasonable generality, and it would be a major advance to establish it.

One available tool for computing symplectic homology is the open-closed morphism from the Hochschild homology of the wrapped Fukaya category[5] [1,6,15,19,32,35]:

---

[2] There are many differing conventions for the grading of symplectic cohomology, and also for which item to call symplectic cohomology and which symplectic homology. We follow [1,19].

[3] When M is not spin, it is necessary to twist one side or the other by a local system [3,27].

[4] Weinstein manifolds and the Weinstein conjecture have the same eponym, but a priori no other relation.

[5] [19,34] contain the relevant definitions. We will soon cite some results which compute the Fukaya category in all relevant cases, so the gist of the article will not be lost to the reader with no idea what the Fukaya category is.

$$\mathrm{HH}_\bullet(\mathrm{Fuk}(W)) \to \mathrm{SH}^{\bullet+n}(W). \tag{3}$$

By either [1,16] or [6,12,13] plus a 'generation' result [8,18], this morphism is by now known to be an isomorphism for Weinstein domains.

For cotangent bundles, there is an object (the cotangent fiber) of $F \in \mathrm{Fuk}(T^*M)$, which generates the category and for which $\mathrm{Hom}(F,F) = \mathrm{H}_{-\bullet}(\Omega M)$. Thus, the open-closed map induces

$$\mathrm{HH}_\bullet(\mathrm{H}_{-\bullet}(\Omega M)) = \mathrm{HH}_\bullet(\mathrm{Hom}(F,F)) \to \mathrm{SH}^{\bullet+n}(T^*M) = \mathrm{H}_{-\bullet}(\mathcal{L}M).$$

This is the same as the corresponding such morphism mentioned above.[6]

How can the open-closed map help us? At first, it does not look promising. We have not said what the Fukaya category is, but its definition involves the same sort of geometrical structures as are involved in symplectic homology. On top of this, we have now added the nontrivial step of taking Hochschild homology. However, just as ΩM has better local-to-global behavior than $\mathcal{L}M$, we also have (as anticipated by [26]):

Theorem 3. [17–19,30,37] *The Fukaya category of a Weinstein manifold is the global sections of a constructible cosheaf of categories over the skeleton. Moreover, this cosheaf is isomorphic to the cosheaf of microlocal sheaves.*[7]

We have not said what microlocal sheaves are and it will not be relevant; but for a definition, see [30], which is built on the technology of [24]. What is relevant is that microlocal sheaves are in principle combinatorial-topological in nature, but in practice, the stalks of the above cosheaf may be complicated categories at complicated singularities of the skeleton. When the skeleton is smooth, the cosheaf is locally constant with stalk the category of chain complexes.

For $W = T^*M$, the cosheaf is simply the path groupoid Ω (twisted by a local system if M is not spin). More generally, Nadler found an explicit collection of the so-called 'arboreal' singularities [28,29], with the property that:

Theorem 4. [29] *When the skeleton is arboreal, the cosheaf of microlocal sheaves has stalks given by representation categories of tree quivers. The cogenerization morphisms are explicit.*

When $\dim W = 2$, the skeleton of W will be a (ribbon) graph, and 'arboreal' essentially amounts to asking that the graph is trivalent. The cosheaf $\mathcal{A}$ will assign the category of chain complexes at smooth points, and the category of exact triangles (aka $\mathrm{Perf}(\bullet \to \bullet)$) at the trivalent vertices, with the obvious cogenerization morphisms. This case was studied in [11]. For $\dim W = 4$, the skeleton is two dimensional; the topology of the typical new kind of singularity is depicted in Fig. 1. More geometric pictures can be found in [5,39].

[6] I am not certain whether this follows from [2], but in any case, it certainly does from [17–19].

[7] Strictly speaking, what is presently in the literature requires that the Weinstein manifold is 'stably polarizable'. It is known to experts how to remove this hypothesis; on the other hand, we will impose it later for different reasons.

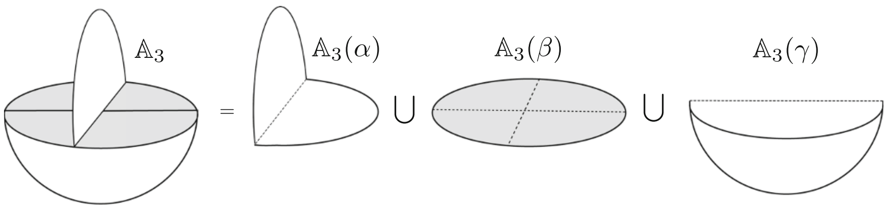


FIGURE 1. The arboreal A3 singularity, and how to glue it from smooth spaces. All arboreal singularities admit analogous gluing descriptions

By an arboreal space, we will mean a pair $(\mathbf{X}, \mathcal{A})$ of a space and a constructible cosheaf, which are locally given by Nadler's explicit models. In particular, the stalks are representation categories of tree quivers. An explicitly combinatorial exposition of this notion can be found in [38].

Some further restrictions on the local models lead to the notion of *positive arboreal space* [5]. (Positivity is encoded by a small amount of additional combinatorial data in addition to $\mathcal{A}$, which models Lagrangian polarizations.) These provide skeleta for a large class of Weinstein manifolds:

Theorem 5. [5] *There is an equivalence of categories between:*

- *Stably polarized Weinstein manifolds and their homotopies*
- *Positive arboreal spaces and their concordances.*

Here, a stable polarization is a choice of Lagrangian sub-bundle of $TW \oplus \mathbb{C}^n$ for some n.

Remark. It is not clearly understood what weaker hypothesis corresponds to taking all arboreal spaces; in dimension 4, none is needed [39], but this is apparently not the case in general [5]. It is expected that this can be repaired by adding some further explicit list of presently unknown singularities. There are also other tricks for reducing problems to the stably polarized case, like [30, Sec. 10]. In any case, I expect that any technique which works for stably polarized Weinstein manifolds should work for Weinstein manifolds in general.

In some sense, we have already arrived at a reduction to algebra: one could try and develop tools for computing the colimit of categories giving $\mathcal{A}(\mathbf{X})$ or its Hochschild homology $\mathrm{HH}_\bullet(\mathcal{A}(\mathbf{X}))$. Indeed, as far as anyone knows, $\mathrm{SH}^\bullet(W)$ is always either infinite dimensional, or zero. If one could show this zero or infinite property for $\mathrm{HH}_\bullet(\mathcal{A}(\mathbf{X}))$, the algebraic Weinstein conjecture would follow. Or if one could show that the natural circle action on any nonvanishing $\mathrm{HH}_\bullet(\mathcal{A}(\mathbf{X}))$ is nontrivial, the result would again follow.

Here, we want to point out that in fact it is possible to directly generalize Eq. (1) and make direct contact with Eq. (2). Consider a cosheaf of categories $\mathcal{A}$ over a space $\mathbf{X}$. The Hochschild homologies form a precosheaf $\mathrm{HH}_\bullet(\mathcal{A})$ which is *not* generally a cosheaf, since Hochschild homology does not commute with colimits. We may cosheafify it and obtain a cosheaf $\mathcal{HH}_\bullet(\mathcal{A})$. There is a natural map

$$\Gamma(\mathbf{X}, \mathcal{HH}_\bullet(\mathcal{A})) \to \mathrm{HH}_\bullet(\mathcal{A}(\mathbf{X})). \tag{4}$$

Because $\mathcal{A}$ is constructible, the LHS is the colimit of the Hochschild homologies of the stalks of $\mathcal{A}$, and the RHS is the Hochschild homology of the colimit of the stalks.

Moreover, $\mathcal{HH}_\bullet(\mathcal{A})$ can be computed explicitly. Indeed, for a tree quiver T, it is the case that $\mathrm{HH}_\bullet(\mathrm{Perf}(T))$ is concentrated in degree zero, and is a free module whose rank is the number of vertices of T. This gives the stalks of $\mathcal{HH}_\bullet(\mathcal{A})$, and in fact one can show:

Theorem 6. [38] *When* $(\mathbf{X}, \mathcal{A})$ *arises from the skeleton of a stably polarizable Weinstein manifold,* $\mathcal{HH}_{\bullet-n}(\mathcal{A})$ *is the cosheaf of compactly supported cohomologies. As* $\mathbf{X}$ *is compact, we have* $\Gamma(\mathbf{X}, \mathcal{HH}_{\bullet-n}(\mathcal{A})) \cong \mathrm{H}^\bullet(\mathbf{X})$.

In fact, it is possible to show that the resulting map

$$\mathrm{H}^\bullet(\mathbf{X}) \cong \Gamma(\mathbf{X}, \mathcal{HH}_{\bullet-n}(\mathcal{A})) \to \mathrm{HH}_{\bullet-n}(\mathcal{A}(\mathbf{X})) \to \mathrm{SH}^\bullet(W) \tag{5}$$

agrees with the original $\mathrm{H}^\bullet(\mathbf{X}) = \mathrm{H}^\bullet(W) \to \mathrm{SH}^\bullet(W)$. This follows from [19, Eq. 1.7] given that the local arboreal models are *Weinstein sectors*, as was shown in [36]. In some more detail: in [36], it is shown that the nondegenerate arboreal sectors are *stopped*; hence, the top row of [19, Eq. 1.7] consists of isomorphisms. Degenerate arboreal singularities are obtained by stop removal; the desired commutativity descends using the stop removal localization sequence and the fact that Hochschild homology sends localizations to exact triangles.

Putting all this together, we have:

Theorem 7. *The algebraic Weinstein conjecture for contact manifolds with stably polarizable Weinstein filling is equivalent to the assertion that* $\Gamma(\mathbf{X}, \mathcal{HH}_\bullet(\mathcal{A})) \to \mathrm{HH}_\bullet(\mathcal{A}(\mathbf{X}))$ *is never an isomorphism for positive arboreal spaces* $(\mathbf{X}, \mathcal{A})$.

We are left with the following:

Question 8. What measures the failure of Hochschild homology to commute with colimits?

Acknowledgements

Beyond the evident intellectual debt that this work owes to the ideas of Claude Viterbo, it happens that I was inspired to write it by the upcoming celebration of his mathematics on the occasion of his sixtieth birthday.

References

[1] Abouzaid, M.: A geometric criterion for generating the Fukaya category. Publ. Math. l'IHÉS **112**, 191–240 (2010)

[2] Abouzaid, M.: A cotangent fibre generates the Fukaya category. Adv. Math. **228**(2), 894–939 (2011)

[3] Abouzaid, M.: Symplectic cohomology and Viterbo's theorem. In: Latschev, J., Oancea, A. (eds.) Free Loop Spaces in Geometry and Topology, pp. 271–485. European Mathematical Society, Zurich (2015)

[4] Albers, P., Hofer, H.: On the Weinstein conjecture in higher dimensions. arXiv:0705.3953

[5] Alvarez-Gavela, D., Eliashberg, Y., Nadler, D.: Positive arborealization of polarized Weinstein manifolds. arXiv:2011.08962

[6] Bourgeois, F., Ekholm, T., Eliashberg, Y.: Effect of Legendrian surgery. Geometry Topol. **16**(1), 301–389 (2012)

[7] Burghelea, D., Fiedorowicz, Z.: Cyclic homology and algebraic k-theory of spaces-ii. Topology **25**(3), 303–317 (1986)

[8] Baptiste, C., Rizell, G.D., Ghiggini, P., Golovko, R.: Geometric generation of the wrapped Fukaya category of Weinstein manifolds and sectors. arXiv:1712.09126

[9] Cieliebak, K., Eliashberg, Y.: From Stein to Weinstein and Back: Symplectic Geometry of Affine Complex Manifolds. American Mathematical Society, Providence (2012)

[10] Cieliebak, K., Latschev, J.: The role of string topology in symplectic field theory. New perspectives and challenges in symplectic field theory **49**, 113–146 (2009)

[11] Dyckerhoff, T., Kapranov, M.: Triangulated surfaces in triangulated categories. J. Eur. Math. Soc. **20**(6), 1473–1524 (2018)

[12] Ekholm, T.: Holomorphic curves for Legendrian surgery. arXiv:1906.07228

[13] Ekholm, T., Lekili, Y.: Duality between Lagrangian and Legendrian invariants. arXiv:1701.01284

[14] Fet, A., Lyusternik, L.: Variational problems on closed manifolds. Dokl. Akad. Nauk SSSR **81**, 17–18 (1951)

[15] Fukaya, K., Oh, Y.-G., Ohta, H., Ono, K.: Lagrangian Intersection Floer Theory: Anomaly and Obstruction. American Mathematical Society, Providence (2010)

[16] Ganatra, G.: Symplectic cohomology and duality for the wrapped Fukaya category. arXiv:1304.7312

[17] Ganatra, G., Pardon, J., Shende, V.: Microlocal Morse theory of wrapped Fukaya categories. arXiv:1809.08807

[18] Ganatra, G., Pardon, J., Shende, V.: Sectorial descent for wrapped Fukaya categories. arXiv:1809.03427

[19] Ganatra, S., Pardon, J., Shende, V.: Covariantly functorial wrapped Floer theory on Liouville sectors. Publ. Math. l'IHÉS **131**(1), 73–200 (2020)

[20] Goodwillie, T.G.: Cyclic homology, derivations, and the free loopspace. Topology **24**(2), 187–215 (1985)

[21] Hofer, H.: Pseudoholomorphic curves in symplectizations with applications to the Weinstein conjecture in dimension three. Invent. Math. **114**(1), 515–563 (1993)

[22] Hofer, H., Viterbo, C.: The Weinstein conjecture in cotangent bundles and related results. Annal. Scuola Norm. Superiore Pisa-Classe Sci. **15**(3), 411–445 (1988)

[23] Hofer, H., Viterbo, C.: The weinstein conjecture in the presence of holomorphic spheres. Commun. Pure Appl. Math. **45**(5), 583–622 (1992)

[24] Kashiwara, M., Schapira, P.: Sheaves on Manifolds. Springer, Berlin (2013)

[25] Klingenberg, W.: Lectures on Closed Geodesics. Springer, Berlin (2012)

[26] Kontsevich, M.: Symplectic geometry of homological algebra. (available at the author's webpage) (2009)

[27] Kragh, T.: The Viterbo transfer as a map of spectra. arXiv:0712.2533

[28] Nadler, D.: Non-characteristic expansions of Legendrian singularities. arXiv:1507.01513

[29] Nadler, D.: Arboreal singularities. Geometry Topol. **21**(2), 1231–1274 (2017)

[30] Nadler, D., Shende, V.: Sheaf quantization in Weinstein symplectic manifolds. arXiv:2007.10154

[31] Niederkrüger, K.: The plastikstufe—a generalization of the overtwisted disk to higher dimensions. Algebr. Geometr. Topol. **6**(5), 2473–2508 (2006)

[32] Seidel, P.: Fukaya categories and deformations. arXiv:math/0206155

[33] Seidel, P.: A biased view of symplectic cohomology. Curr. Dev. Math. **2006**(1), 211–254 (2006)

[34] Seidel, P.: Fukaya Categories and Picard–Lefschetz Theory, vol. 10. European Mathematical Society, Zurich (2008)

[35] Seidel, P.: Symplectic homology as Hochschild homology. Algebr. Geometry Seattle **2005**, 415–434 (2009)

[36] Shende, V.: Arboreal singularities from Lefschetz fibrations. arXiv:1809.10359

[37] Shende, V.: Microlocal category for Weinstein manifolds via h-principle. arXiv:1707.07663

[38] Shende, V., Takeda, A.: Calabi-Yau structures on topological Fukaya categories. arXiv:1605.02721

[39] Starkston, L.: Arboreal singularities in Weinstein skeleta. Sel. Math. New Ser. **24**(5), 4105–4140 (2018)

[40] Clifford Henry Taubes: The Seiberg–Witten equations and the Weinstein conjecture. Geometry Topol. **11**(4), 2117–2202 (2007)

[41] Vigué-Poirrier, M., Sullivan, D.: The homology theory of the closed geodesic problem. J. Differ. Geometry **11**(4), 633–644 (1976)

[42] Viterbo, C.: Functors and Computations in Floer homology with applications, II. arXiv:1805.01316

[43] Viterbo, C.: A proof of Weinstein's conjecture in $\mathbb{R}^{2n}$. Ann. l'Inst. Henri Poincare (C) Non Linear Anal. **4**(4), 337–356 (1987)

[44] Viterbo, C.: Generating functions, symplectic geometry, and applications. In: Proceedings of the International Congress of Mathematicians, pp. 537–547. Springer (1995)

[45] Viterbo, C.: Functors and computations in Floer homology with applications. I. Geom. Funct. Anal. **9**(5), 985–1033 (1999)

[46] Weinstein, A.: On the hypotheses of Rabinowitz' periodic orbit theorems. J. Differ. Equ. **33**(3), 353–358 (1979)

Vivek Shende
Center for Quantum Mathematics
Syddansk Universitet
Odense
Denmark
e-mail: vivek.vijay.shende@gmail.com

and

Department of Mathematics
UC Berkeley
Berkeley
California
USA

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:48
https://doi.org/10.1007/s11784-022-00965-6
Published online June 7, 2022

Journal of Fixed Point Theory and Applications

Quantum cohomology as a deformation of symplectic cohomology

Matthew Strom Borman, Nick Sheridan and Umut Varolgunes

Abstract. We prove that under certain conditions, the quantum cohomology of a positively monotone compact symplectic manifold is a deformation of the symplectic cohomology of the complement of a simple crossings symplectic divisor. We also prove rigidity results for the skeleton of the divisor complement.

Mathematics Subject Classification. 53D37, 53D40.

1. Introduction

1.1. Geometric setup

Let (M, ω) be a compact symplectic manifold satisfying the monotonicity condition:

$$2\kappa c_1(TM) = [\omega], \text{ for some } \kappa > 0$$

in $H^2(M; \mathbb{R})$.

Let $D = \cup_{i=1}^N D_i \subset M$ be an SC symplectic divisor (in the sense of [36, Section 2]) and set $X = M \backslash D$.[1]

We assume that there exist positive rational numbers $\lambda_1, \ldots, \lambda_N$ called *weights* such that

$$2c_1(TM) = \sum_i \lambda_i \cdot \mathrm{PD}(D_i) \quad \text{in } H^2(M; \mathbb{R}).$$

Note that the number of weights in the setup depends on the divisor. If $\mathrm{PD}(D_i)$ are linearly independent classes in $H^2(M; \mathbb{R})$ (e.g., if D is smooth), the weights are canonically determined. Otherwise, the choice of weights is extra data.

[1] From now on, we systematically shorten SC symplectic divisor to SC divisor as we believe this will not cause confusion.

This article is part of the topical collection "Symplectic geometry—a Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder and Felix Schlenk.

The classes $\mathrm{PD}(D_i)$ have canonical lifts to the relative cohomology group $H^2(M, X; \mathbb{R})$ along the canonical map

$$H^2(M, X; \mathbb{R}) \to H^2(M; \mathbb{R}),$$

see Sect. 2.1 for more details. Let us denote these classes by $\mathrm{PD}^{rel}(D_i)$ and note that they form a basis of $H^2(M, X; \mathbb{R})$. We define the class

$$\boldsymbol{\lambda} := \sum_i \lambda_i \cdot \mathrm{PD}^{rel}(D_i) \in H^2(M, X; \mathbb{R}),$$

which is a lift of $2c_1(TM)$ by construction. Consequently, $\kappa\boldsymbol{\lambda}$ is a lift of $[\omega]$.

Let us denote by $\Omega^*(M, X)$ the relative de Rham complex, which is by definition the cone of the restriction map $\Omega^*(M) \to \Omega^*(X)$. Note that there is a relative de Rham isomorphism

$$H^*(\Omega^*(M, X)) \to H^*(M, X; \mathbb{R}),$$

which in particular tells us that there exists a one-form $\theta \in \Omega^1(X)$ satisfying

$$\omega|_X = \mathrm{d}\theta, \qquad \text{and} \qquad \int_u \omega - \int_{\partial u} \theta = \kappa\boldsymbol{\lambda} \cdot u \qquad \text{for all } u \in H_2(M, X).$$

Using that $\kappa\lambda_i > 0$ for all i, we may arrange that (X, θ) is a finite type convex symplectic manifold. Moreover $c_1(TX) = 0$, and a preferred homotopy class of trivializations of a power of the canonical bundle of X is determined by the choice of weights λ_i (see Sect. 3.3 for details).

Example 1.1. Suppose that M is a smooth complex projective variety, D a simple normal crossings divisor, and $\mathbf{D}^{\boldsymbol{\lambda}} = \sum_i \lambda_i D_i$ is an effective ample $\mathbb{Q}$-divisor whose class in $\mathrm{CH}_*(M)_{\mathbb{Q}}$ is twice the anticanonical class:

$$[\mathbf{D}^{\boldsymbol{\lambda}}] = -2K_M.$$

Let us also choose an arbitrary $\kappa > 0$.

Choose a positive integer k such that $k\mathbf{D}^{\boldsymbol{\lambda}}$ has integral coefficients, and let $\mathcal{L}$ be the corresponding complex line bundle with section s. Ampleness implies that we can choose a positive Hermitian metric $\|\cdot\|$ on $\mathcal{L}$ with curvature 2-form F. We define the symplectic form $\omega := \frac{-i\kappa}{k} F$ on M. We can also define the primitive $\theta := \frac{\kappa}{k} d^c \log \|s\|$ of ω on X. Using D as our SC divisor and λ_i as the weights, this puts us in the geometric set-up described above. Note that in this case X is an affine variety.

The setup that we described thus far is among the most studied in symplectic geometry. Now, we introduce an hypothesis which is less common, but which is very crucial for our results.

Hypothesis A. We have $\lambda_i \leq 2$ for all $i = 1, \ldots, N$.

Remark 1.2. Recalling that $[\mathbf{D}^{\boldsymbol{\lambda}}] = -2K_M$, we note that the extreme case of Hypothesis A, namely $\lambda_i = 2$ for all i, corresponds to (M, D) being log Calabi–Yau. If we in addition assume that each irreducible component of D is ample, then Hypothesis A implies that (M, D) is either log Calabi–Yau or log general type.

Example 1.3. Consider the setup of Example 1.1. Let us take $M = \mathbb{CP}^n$, D a simple normal crossings divisor of degree d. Then we may choose weights λ_i such that Hypothesis A holds if and only if $d \geq n+1$. Note that (M, D) is log Calabi–Yau if $d = n+1$, and log general type if $d > n+1$. To see one direction of the implication, assume that $D = \cup_{i=1}^N D_i$ with D_i smooth of degree d_i and Hypothesis A holds. Then we have

$$2(n+1) = \sum_i \lambda_i d_i \leq \sum_i 2d_i = 2d.$$

Note that in the log Calabi–Yau case $d = n+1$, all weights λ_i must be equal to 2.

1.2. Quantum cohomology is a deformation of symplectic cohomology

We fix, once and for all, a commutative ring $\mathbb{k}$. Let $A \subset \mathbb{Q}$ be the subgroup generated by the weights λ_i, and set $\Lambda = \mathbb{k}[A]$ to be the group algebra of A.[2] We define a $\mathbb{Q}$-grading on Λ by putting e^a in degree $i(e^a) = a$. Let $a_0 > 0$ be a generator of A, and define $q := e^{a_0}$. Hence, we have an isomorphism of algebras $\Lambda \cong \mathbb{k}[q, q^{-1}]$.

Throughout the paper, we will consider various filtrations associated with filtration maps (see Sect. A.1 for a review of this notion). We will abuse notation using the same symbol for the filtration map and the associated filtration. In the first instance of this abuse of notation, we introduce the filtration $\mathcal{Q}_{\geq \bullet}$ on Λ associated with the filtration map $\mathcal{Q} : \Lambda \to \mathbb{Z}$ induced by $\mathcal{Q}(q^a) = a$. Thus, $\mathcal{Q}_{\geq p}\Lambda$ consists of all linear combinations of monomials q^a with $a \geq p$.

We define the graded Λ-module $QH^*(M; \Lambda) := H^*(M; \mathbb{k}) \otimes_{\mathbb{k}} \Lambda$, equipped with the tensor product grading.[3] We are concerned with the following idealized and vague conjecture:

Conjecture 1.4. *Under certain hypotheses:*

(a) $QH^*(M; \Lambda)$ *is the cohomology of a natural deformation of the symplectic cochain complex* $SC^*(X; \mathbb{k})$ *over* Λ*;*
(b) *Furthermore, the associated spectral sequence converges:* $E_1 = SH^*(X; \mathbb{k}) \otimes_{\mathbb{k}} \Lambda \Rightarrow QH^*(M; \Lambda)$.

We will prove a modified version of Conjecture 1.4 in the setup described in Sect. 1.1. Notably, for the analogue of part (b) we will need Hypothesis A.

Remark 1.5. Conjecture 1.4 part (b) is not true in general. For example, if we take $M = \mathbb{CP}^n$ and D a hyperplane, then $X = M \setminus D = \mathbb{C}^n$ has vanishing symplectic cohomology. But we cannot have a spectral sequence with vanishing E_1 page, converging to the non-vanishing cohomology of $\mathbb{CP}^n$! Note that Hypothesis A is not satisfied in this case by Example 1.3. More generally, it is not satisfied for D a union of $N \leq n$ hyperplanes; and $X = \mathbb{C}^{n+1-N} \times (\mathbb{C}^*)^{N-1}$ still has vanishing symplectic cohomology in these cases.

[2] Explicitly, Λ is the $\mathbb{k}$-algebra of $\mathbb{k}$-linear combinations of the symbols e^a where $a \in A$, and $e^a \cdot e^b = e^{a+b}$.

[3] Our main results do not concern the quantum cup product on $QH^*(M; \Lambda)$, but it plays a role in some of the conjectures in Sect. 1.6.

Note that Conjecture 1.4 (a) is a statement about the chain complex $SC^*(X;\Bbbk)$, which depends on various auxiliary data which we have not included in the notation. Given such a choice, we consider the chain complex

$$SC^*_\Lambda := (SC^*(X;\Bbbk) \otimes_\Bbbk \Lambda, d \otimes id_\Lambda) \tag{1.1}$$

with the tensor product grading.[4] It admits a $\mathcal{Q}$-filtration induced by the filtration map $\mathcal{Q}(\gamma \otimes r) = \mathcal{Q}(r)$. In the modified version of Conjecture 1.4 (a) that we prove, we will need to replace SC_Λ with an 'equivalent' filtered complex $\widetilde{SC}_\Lambda$:

Theorem B (a modified version of Conjecture 1.4 (a)). *Assume that we are in the setup described in Sect. 1.1. Then there exists a choice of the auxiliary data needed to define $SC^*(X;\Bbbk)$, and a filtered cochain complex of $\mathcal{Q}_{\geq 0}\Lambda$-modules, $\widetilde{SC}_\Lambda := \left(\widetilde{SC}^*_\Lambda, \widetilde{d}, \widetilde{\mathcal{Q}}_{\geq \bullet}\right)$, with the following properties:*

(1) $\left(\widetilde{SC}^*_\Lambda, \widetilde{d}, \widetilde{\mathcal{Q}}_{\geq \bullet}\right)$ *is filtered quasi-isomorphic to* $(SC^*_\Lambda, d \otimes id_\Lambda, \mathcal{Q}_{\geq \bullet})$ (*see Sect. 5.2 for the precise meaning of this statement*).

(2) *There exists a second $\mathcal{Q}_{\geq 0}\Lambda$-linear differential ∂ on $\widetilde{SC}^*_\Lambda$, such that $\partial - \widetilde{d}$ strictly increases the $\widetilde{\mathcal{Q}}$-filtration. We call ∂ the deformed differential.*

(3) *We have* $H^*\left(\widetilde{SC}^*_\Lambda, \partial\right) \cong QH^*(M;\Lambda)$.

By considering the spectral sequence associated with the deformed filtered complex $\left(\widetilde{SC}^*_\Lambda, \partial, \widetilde{\mathcal{Q}}_{\geq \bullet}\right)$, we then obtain:

Theorem C (Conjecture 1.4 (b)). *Assume now that Hypothesis A holds. Then the spectral sequence associated with the filtered complex $(\widetilde{SC}_\Lambda, \partial, \widetilde{\mathcal{Q}}_{\geq \bullet})$ converges, and has*

$$E_1^{j,k} = SH^{k+j(1+a_0)}(X;\Bbbk) \otimes_\Bbbk \Bbbk \cdot q^{-j} \Rightarrow QH^{j+k}(M;\Lambda),$$

where $j \in \mathbb{Z}$ and $k \in \mathbb{Q}$.

Remark 1.6. Because our Floer complexes are $\mathbb{Q}$-graded, our spectral sequence $(E_i^{j,k}, d_i^{j,k})$ has $i, j \in \mathbb{Z}$ and $k \in \mathbb{Q}$, rather than the usual $k \in \mathbb{Z}$. All the standard theory of spectral sequences goes through in this slightly more general context. Indeed, one can think of such a spectral sequence as a collection of ordinary spectral sequences indexed by $\{c \in \mathbb{Q} : 0 \leq c < 1\}$, by setting $E(c)_i^{j,k} = E_i^{j,k-c}$.

Let us note an immediate corollary:

Corollary 1.7. *Under Hypothesis A, the affine variety X from Example 1.1 has non-vanishing symplectic cohomology. In particular, it is not flexible.*

Remark 1.8. We expect that analogues of Theorems B and C hold also under the assumption that M is Calabi–Yau, i.e., $c_1(TM) = 0$. Indeed, Yuhan Sun has recently proved very closely related results [35]. In this case, the key notion is 'index boundedness', as used by McLean in [25], together with certain lower bounds on the indices of the one-periodic orbits 'on the divisor'. We refer the reader to [35] for more details.

[4] In general this will be a $\mathbb{Q}$-grading.

1.3. Rigidity results

By applying the same techniques as those used to prove Theorems B and C, we will prove a rigidity result for certain subsets of M.

The main tool used to prove the result is a version of the relative symplectic cohomology developed by the third author in [40] (with which we assume some familiarity). Slightly modifying the construction there, for any compact $K \subset M$, we can define a $\mathbb{Q}$-graded Λ-module $SH^*_M(K;\Lambda)$. The definition of this invariant involves choosing acceleration data to define a Floer 1-ray, and then the chain-level invariant is defined to be not the telescope but a degreewise-completed telescope. More details are given in Sect. 3.2.[5]

Following [38], we say that K is *SH-invisible* if $SH^*_M(K;\Lambda) = 0$, and *SH-visible* otherwise. One can prove that SH-visible subsets are not stably displaceable (see Theorem 3.6).[6] For example, PSS isomorphisms imply that $QH^*(M;\Lambda) \cong SH_M(M;\Lambda)$, so M is SH-visible; and as a result M is not stably displaceable (as a subset of itself). Moreover, there are restriction maps $SH^*_M(K';\Lambda) \to SH^*_M(K;\Lambda)$, whenever K' contains K. By a unitality argument, it follows that any compact subset of an SH-invisible subset is SH-invisible (see Theorem 3.7).

We say that K is *nearly SH-visible* if any compact domain that contains K in its interior is SH-visible. As straightforward consequences of the previous paragraph, one can show that SH-visible subsets are nearly SH-visible,[7] and nearly SH-visible subsets are not stably displaceable.

We say that K is *SH-full* if for any compact K' contained in $M\backslash K$, $SH^*_M(K';\Lambda) = 0$. SH-full subsets are nearly SH-visible, as a consequence of the Mayer–Vietoris property of relative symplectic cohomology [40]. One can prove that an SH-full subset cannot be displaced from a nearly SH-visible subset by a symplectomorphism. It is also the case that SH-full subsets are not stably displaceable from themselves (see [38, Corollary 1.9]). By a closed–open map plus unitality argument (see [38, proof of Theorem 1.2 (6)]), it can

[5] The construction that we give here can be generalized to all symplectic manifolds with the property

$$2c_1(TM) = \eta[\omega] \quad \text{for some } \eta \in \mathbb{R},$$

and subgroup $B \subset \mathbb{R}$ which contains $\omega(\pi_2(M))$. Namely, we define the filtered graded algebra $\mathbb{k}[B]$ where $i(e^b) = \eta b$ and the filtration level of e^b is b. We then define the Novikov-type algebra

$$\Lambda_{B,\eta} := \widehat{\mathbb{k}[B]},$$

where the completion is degreewise. Our Λ in this paper is nothing but $\Lambda_{\kappa A,\kappa^{-1}}$, whereas the Novikov field used in e.g. [38] is $\Lambda_{\mathbb{R},0}$. The construction produces a $\mathbb{Z} + \eta B$-graded $\Lambda_{B,\eta}$-module $SH^*_M(K;\Lambda_{B,\eta})$. When $c_1(M) = 0$, and taking into account only the contractible orbits, the invariant that is denoted by $SH^*_M(K;\Lambda)$ in [38] is a special case of this construction as well. It would have been called $SH^*_M(K;\Lambda_{\mathbb{R},0})$ in our notation here, and a capped orbit (γ,u) here would be interpreted as $T^{\mathcal{A}(\gamma,u)}\gamma$ in that paper's notation. Let us also note that $\eta < 0$ requires using virtual techniques, which forces us to make certain assumptions on $\mathbb{k}$.

[6] Stably displaceable means $K \times Z \subset M \times T^*S^1$ is displaceable from itself by a Hamiltonian diffeomorphism, where Z is the zero section of T^*S^1.

[7] We do not have examples of nearly SH-visible subsets which are not SH-visible.

be shown that Floer-theoretically essential (over $\Bbbk$) monotone Lagrangians cannot be displaced from SH-full subsets by a symplectomorphism.

Now we state our result, which will need some extra hypotheses beyond those already mentioned in Sect. 1.1. First of all, we assume that D is an orthogonal SC divisor. Then there exist Hamiltonian circle actions rotating about the D_i, and commuting on the overlaps, by [23]; we assume that θ is 'adapted' to such a system of commuting Hamiltonians, in an appropriate sense. We make these notions precise in Sect. 2 below. We remark that the data we need is weaker than what McLean calls a 'standard tubular neighbourhood' in [25].

Let Z be the Liouville vector field of (X, θ). We define the continuous function $\rho^0 : X \to \mathbb{R}$, so that the Liouville flow starting at x is defined precisely for time $t < -\log(\rho^0(x))$. Note that $\mathbb{L} = \{\rho^0 = 0\}$ is the Lagrangian skeleton of (X, θ). We extend the function to $\rho^0 : M \to \mathbb{R}$ by setting $\rho^0|_D = 1$.

Definition 1.9. We define

$$\tilde{\sigma}_{\mathrm{crit}} := 1 - \frac{2}{\max_i \lambda_i}, \qquad \sigma_{\mathrm{crit}} := \max(0, \tilde{\sigma}_{\mathrm{crit}}),$$

and set

$$K_{\mathrm{crit}} := \{\rho^0 \leq \sigma_{\mathrm{crit}}\} \subset M.$$

Note that $\sigma_{\mathrm{crit}} = 0$, and hence $K_{\mathrm{crit}} = \mathbb{L}$, if and only if Hypothesis A is satisfied.

Equivalently, K_{crit} is the image of the Liouville flow for time $\log(\sigma_{\mathrm{crit}})$.

Theorem D. *The subset $K_{\mathrm{crit}} \subset M$ is SH-full. In particular, if Hypothesis A is satisfied, then $\mathbb{L}$ is SH-full.*

For example, this means that when Hypothesis A is satisfied, $\mathbb{L}$ cannot be displaced from any Floer-theoretically essential (over $\Bbbk$) monotone Lagrangian.

Remark 1.10. It is possible for a compact subset to be SH-full for one choice of $\Bbbk$ but not for another. We did not make a big deal about this as our result is uniform for all ground fields. We expect this to play a real role in the context of Conjecture 1.20. We also refer the reader to Remark 1.8 of [38] for another weakening of the notion of SH-fullness.

Remark 1.11. An analogue of Theorem D, in the case that M is Calabi–Yau, was proved in [38].

1.4. Floer theory conventions

We give a quick outline of our conventions for Hamiltonian Floer theory on M, for the purposes of giving an overview of the proofs of our main results in the following section (see Sect. 3 for more details). Let $A' \subset \mathbb{Z}$ be the image of $2c_1(TM) : \pi_2(M) \to \mathbb{Z}$, and set $\Lambda' := \Bbbk[A']$. Note $A' \subset A$, so $\Lambda' \subset \Lambda$.

A 'cap' for an orbit $\gamma : S^1 \to M$ of a Hamiltonian $H : S^1 \times M \to \mathbb{R}$ is an equivalence class of discs u bounding γ, where two discs are considered equivalent if they have the same symplectic area. One can associate an index $i(\gamma, u)$ and action $\mathcal{A}(\gamma, u)$ to a capped orbit (γ, u) of a non-degenerate Hamiltonian. The 'mixed index'

$$i_{\text{mix}}(\gamma) = i(\gamma, u) - \kappa^{-1}\mathcal{A}(\gamma, u)$$

is independent of the cap u.

We define $CF^i(M, H)$ to be the free $\mathbb{Z}$-graded $\Bbbk$-module generated by capped orbits (γ, u) of H satisfying $i(\gamma, u) = i$. This becomes a graded Λ'-module, where $e^a \cdot (\gamma, u) = (\gamma, u\#B)$ where $2c_1(TM)(B) = a$. It also admits an action filtration, associated with the filtration map induced by $\mathcal{A}(\gamma, u)$. We define $CF^*(M, H; \Lambda) := CF^*(M, H) \otimes_{\Lambda'} \Lambda$. It has a $\Bbbk$-basis of 'fractional caps': a fractional cap for γ is a formal expression $u + a$, where u is a cap for γ and $a \in A$, and we declare $u + a \sim u' + a'$ iff $a - a' \in A'$ and $(\gamma, u') = e^{a-a'} \cdot (\gamma, u)$.

The Floer differential increases degree by 1, and respects the action filtration (i.e., it does not decrease action). The PSS isomorphism identifies $HF^*(M, H; \Lambda) \cong QH^*(M; \Lambda)$. If $H_1 \leq H_2$ pointwise, then there exists a continuation map $CF^*(M, H_1; \Lambda) \to CF^*(M, H_2; \Lambda)$ which respects action filtrations.

We now explain our conventions for relative symplectic cohomology. Given $K \subset M$ compact, a choice of acceleration data (H_τ, J_τ) is the data required to define a Floer 1-ray

$$\mathcal{C}(H_\tau, J_\tau) := CF^*(M, H_1; \Lambda) \to CF^*(M, H_2; \Lambda) \to \cdots$$

consisting of Floer cohomology groups and continuation maps, where the monotone sequence of Hamiltonians $H_1 \leq H_2 \leq \cdots$ converges to 0 on K and $+\infty$ outside of K. We consider the telescope complex $tel(\mathcal{C})$, which is constructed so that

$$H^*(tel(\mathcal{C})) = \varinjlim_i HF^*(M, H_i; \Lambda) = QH^*(M; \Lambda).$$

We define $\widehat{tel}(\mathcal{C})$ to be the degreewise completion of $tel(\mathcal{C})$ with respect to the action filtration, and $SH^*_M(K; \Lambda) := H^*(\widehat{tel}(\mathcal{C}))$.

1.5. Outline of proofs

In this section, we give an extended overview of the proofs of our main results, trying to convey the main ideas while avoiding technicalities. We assume that we are in the geometric setup described in Sect. 1.1, with the additional properties and data explained in Sects. 1.2 and 1.3.

We will construct a function $\rho : M \to \mathbb{R}$ which is a smoothing of ρ^0 (really, a family of smoothings ρ^R for $R > 0$ sufficiently small) with the following properties:

- it will be continuous on M, and smooth on the complement of $\mathbb{L}$;
- $\rho|_{\mathbb{L}} = 0$ and $\rho|_D \approx 1$;[8]

[8]If D is smooth then we can arrange that $\rho|_D = 1$; if D is normal crossings then $\rho|_D$ will be equal to 1 away from a neighbourhood of the singularities of D, where an error is introduced by 'rounding corners'.

- it will satisfy $Z(\rho) = \rho$ on $X \backslash \mathbb{L}$.

It also has the property that $K_\sigma := \{\rho \le \sigma\}$ is a Liouville subdomain of X for any $\sigma \in (0,1)$. Because $Z(\rho) = \rho$, $K_\sigma \to \mathbb{L}$ as $\sigma \to 0$.

1.5.1. Theorem B. We choose $\sigma \in (\sigma_{\text{crit}}, 1)$, and construct acceleration data (H_τ, J_τ) for $K_\sigma \subset M$ as follows. Fix $0 < \ell_1 < \ell_2 < \cdots$ such that the Reeb flow on $Y = \partial K_\sigma$ has no ℓ_n-periodic orbits for all n, and $\ell_n \to \infty$ as $n \to \infty$. We choose an increasing family of smooth functions $h_n : \mathbb{R} \to \mathbb{R}$, approximating the piecewise-linear functions $\max(0, \ell_n(\rho - \sigma))$ with increasing accuracy as $n \to \infty$, and being linear with slope ℓ_n for $\rho \ge \sigma$ (see Fig. 2). We consider acceleration data (H_τ, J_τ) for $K_\sigma \subset M$ such that J_τ is of contact type near ∂K_σ and H_n is equal to a carefully chosen perturbation of $h_n \circ \rho$. The 1-periodic orbits of Hamiltonians H_n then fall into two groups (1) SH-type: contained in K_σ and (2) D-type: outside of K_σ. We also make sure that the SH-type orbits that are not "Reeb type" are constant.

We now consider the Floer 1-ray

$$\mathcal{C}(H_\tau, J_\tau) := CF^*(M, H_1; \Lambda) \to CF^*(M, H_2; \Lambda) \to \cdots$$

associated with our choice of acceleration data. We decompose the associated telescope complex as a direct sum of the SH-type generators and the D-type generators:

$$tel(\mathcal{C}) = tel(\mathcal{C})_{SH} \oplus tel(\mathcal{C})_D.$$

This is a direct sum as Λ-modules, not as cochain complexes: the differential, which we denote by ∂, mixes up the factors.

By restricting the acceleration data with K_σ, we also obtain a Floer 1-ray of $\mathbb{k}$-cochain complexes

$$\mathcal{C}_{SH}(H_\tau, J_\tau) := CF^*(K_\sigma, H_1|_{K_\sigma}; \mathbb{k}) \to CF^*(K_\sigma, H_2|_{K_\sigma}; \mathbb{k}) \to \cdots$$

and we set

$$SC^*(X; \mathbb{k}) := tel(\mathcal{C}_{SH}).$$

We denote the differential by d. Strictly speaking, this is the cochain complex defining the symplectic cohomology of the Liouville domain K_σ à la Viterbo [45]. Our notation is justified by the fact that in [23, Section 4], McLean shows that $H^*(SC^*(X; \mathbb{k}))$ only depends on X.

We associate a canonical fractional cap u_{in} to each SH-type orbit γ, by setting $u_{\text{in}} := u - u \cdot \boldsymbol{\lambda}$ for an arbitrary cap u (one easily checks that u_{in} is independent of u). There is then an isomorphism of Λ-modules (recall Equation (1.1))

$$\begin{aligned} SC^*_\Lambda &\xrightarrow{\sim} tel(\mathcal{C})_{SH} \\ \gamma \otimes q^a &\mapsto q^a \cdot (\gamma, u_{\text{in}}). \end{aligned} \tag{1.2}$$

However, this is not a chain map: indeed, the matrix component $\partial_{SH,SH}$ need not even be a differential.

Proposition 1.12 (= Proposition 5.10). *For any Floer solution u that contributes to $\mathcal{C}(H_\tau, J_\tau)$ with both ends asymptotic to SH-type orbits, we have $u \cdot \boldsymbol{\lambda} \ge 0$. In case of equality, u is contained in K_σ.*

One could think of Proposition 1.12 as a manifestation of positivity of intersection of Floer trajectories with the components of the divisor D (c.f. [37, Lemma 4.2]), although we actually prove it using an argument related to Abouzaid–Seidel's 'integrated maximum principle' [2, Lemma 7.2].

The consequence of Proposition 1.12 is that $d \otimes id_\Lambda - \partial_{SH,SH}$ strictly increases the $\mathcal{Q}$-filtration. Using PSS isomorphisms, we also see that the homology of $tel(\mathcal{C})$ is isomorphic to $QH^*(M;\Lambda)$. Thus, we are some way towards proving Theorem B, but we are troubled by the existence of D-type orbits. The following proposition is the most important ingredient in the proof of Theorem B, as it allows us to 'throw out' the D-type orbits.

Proposition 1.13. *There exists $\delta > 0$ such that*

$$i_{\text{mix}}(\gamma) \geq \kappa^{-1}\delta\ell_n$$

for any D-type orbit γ of H_n.

Sketch of proof when D is smooth. The Hamiltonian H_n is approximately equal to $\ell_n(\rho - \sigma)$ near D. When D is smooth we have $\rho = r/\kappa\lambda$, where r is the moment map for a Hamiltonian circle action rotating a neighbourhood of D about D with unit speed. In particular, the Hamiltonian flow of H_n approximately rotates around D at speed $\ell_n/\kappa\lambda$, and the D-type orbits are approximately constant. (This is in contrast to the Hamiltonians used, for example, in [37], which are approximately constant near D, and which have non-constant D-type orbits linking D.)

We compute the mixed index with respect to the approximately constant cap, which is called u_{out} in the body of the paper. As the Hamiltonian flow of H_n rotates around D at speed $\ell_n/\kappa\lambda$, we have $i(\gamma, u_{\text{out}}) \approx 2\ell_n/\kappa\lambda$. On the other hand, we have $H_n \approx h_n(1) \approx \ell_n(1-\sigma)$ along D, and $\omega(u_{\text{out}}) \approx 0$, so $\mathcal{A}(\gamma, u_{\text{out}}) \approx \ell_n(1-\sigma)$. Combining we have

$$\begin{aligned} i_{\text{mix}}(\gamma) &= i(\gamma, u_{\text{out}}) - \kappa^{-1}\mathcal{A}(\gamma, u_{\text{out}}) \\ &\approx \frac{2\ell_n}{\kappa\lambda} - \kappa^{-1}\ell_n(1-\sigma) \\ &\geq \kappa^{-1}\ell_n(\sigma - \sigma_{\text{crit}}), \end{aligned}$$

which gives the desired result, as we chose $\sigma > \sigma_{\text{crit}}$. □

Our first thought, in trying to 'throw out' the D-type orbits, might be to consider the submodule of $tel(\mathcal{C})$ spanned by orbits satisfying $i_{\text{mix}}(\gamma) < \kappa^{-1}\delta\ell_n$, as that is contained in $tel(\mathcal{C})_{SH}$ by Proposition 1.13. However this does not behave well with respect to the differential: it is neither subcomplex, quotient complex, nor subquotient. Instead, we consider a family of subquotient complexes $(SC_\Lambda^{(p)}, \partial_p)$ of $tel(\mathcal{C})$, indexed by $p \in \mathbb{R}$, spanned by generators (γ, u) satisfying

$$i(\gamma, u) < p \leq \frac{\mathcal{A}(\gamma, u) + \delta\ell_n}{\kappa}.$$

(Note that these are contained in $tel(\mathcal{C})_{SH}$ by Proposition 1.13, which is identified with SC_Λ by (1.2).)

To see that this is a subquotient of $tel(\mathcal{C})$, we first observe that the differential clearly increases the quantity $\mathcal{F}(\gamma,u) = \frac{\mathcal{A}(\gamma,u)+\delta\ell_n}{\kappa}$: it increases action, and increases n and hence ℓ_n by the definition of the telescope complex. Therefore, it defines a filtration map, so $\mathcal{F}_{\geq p}tel(\mathcal{C})$ is a subcomplex. On the other hand, the degree truncation $\sigma_{<p}C^\bullet := \bigoplus_{i<p} C^i$ is always a quotient complex of any cochain complex. Thus $(SC_\Lambda^{(p)}, d\otimes id_\Lambda) = \sigma_{<p}\mathcal{F}_{\geq p}SC_\Lambda$ is a subquotient of SC_Λ, whose generators are all of SH-type by Proposition 1.13.

Proposition 1.14. *For any $p\in\mathbb{R}$, both $\mathcal{F}_{\geq p}tel(\mathcal{C})\subset tel(\mathcal{C})$ and $\mathcal{F}_{\geq p}tel(\mathcal{C}_{SH})\subset tel(\mathcal{C}_{SH})$ are quasi-isomorphic subcomplexes.*

Sketch of proof. We may identify $\mathcal{F}_{\geq p}tel(\mathcal{C})$ as the telescope complex of the 1-ray of Floer groups $\mathcal{A}_{\geq\kappa p-\delta\ell_n}CF^*(M,H_n,\Lambda)$. The key point is that $\kappa p - \delta\ell_n\to-\infty$ as $n\to\infty$, and the action filtration is exhaustive, so the direct limit 'eventually catches everything' (see Appendix A.2). The argument for $\mathcal{F}_{\geq p}tel(\mathcal{C}_{SH})\subset tel(\mathcal{C}_{SH})$ is identical. $\square$

Because $H^j(\sigma_{<p}C^\bullet) = H^j(C^\bullet)$ for $j<p-1$, we have

$$H^j(\sigma_{<p}\mathcal{F}_{\geq p}tel(\mathcal{C}),\partial) = H^j(M;\Lambda) \qquad \text{for } j<p-1.$$

If we were willing to weaken the statement in Theorem B, and only achieve the isomorphism of item (3) up to degree $p-1$, we would now be done: we could simply take $\widetilde{SC}_\Lambda = SC_\Lambda^{(p)}$, with $\widetilde{\mathcal{Q}}$ equal to the filtration induced by $\mathcal{Q}$. However, to get the corresponding statement in all degrees, we observe that there are natural maps $SC_\Lambda^{(p)}\to SC_\Lambda^{(q)}$ for all $p\geq q$, induced by the inclusion $\mathcal{F}_{\geq p}\subset\mathcal{F}_{\geq q}$ and the projection $\sigma_{<p}\twoheadrightarrow\sigma_{<q}$. We define $(\widetilde{SC}_\Lambda,\partial)$ to be the homotopy inverse limit of the inverse system of chain complexes $(SC_\Lambda^{(p)},\partial_p)$, and $\widetilde{\mathcal{Q}}$ the filtration induced by the $\mathcal{Q}$-filtration on SC_Λ. The result is that

$$H^*(\widetilde{SC}_\Lambda,\partial) = \varprojlim_p H^*(SC_\Lambda^{(p)},\partial_p) = QH^*(M;\Lambda)$$

as desired. (We remark that this step requires us to check that $\varprojlim^1 H^*(SC_\Lambda^{(p)},\partial_p)=0$; indeed the inverse system is easily seen to satisfy the Mittag–Leffler property.) This completes the sketch proof of Theorem B.

1.5.2. Theorem C. To prove Theorem C, it suffices to prove that the $\widetilde{\mathcal{Q}}$-filtration is bounded below and exhaustive, by the 'Classical Convergence Theorem' [46, Theorem 5.5.1]. The $\mathcal{Q}$-filtration on each $SC_\Lambda^{(p)}$ is exhaustive by definition, but the $\widetilde{\mathcal{Q}}$-filtration on $\widetilde{SC}_\Lambda$ is not exhaustive, due to the direct product taken in the construction. Nevertheless one can show that the inclusion $\cup_q\widetilde{\mathcal{Q}}_{\geq q}\widetilde{SC}_\Lambda\subset\widetilde{SC}_\Lambda$ is a quasi-isomorphism, and the $\widetilde{\mathcal{Q}}$-filtration on this quasi-isomorphic subcomplex is exhaustive by construction.

Thus the main thing to prove, to apply the Classical Convergence Theorem, is that the $\widetilde{\mathcal{Q}}$-filtration is bounded below. The key ingredient is the following:

Proposition 1.15. *Suppose that Hypothesis A is satisfied. Then for any SH-type orbit γ, we have $i(\gamma, u_{\text{in}}) \geq 0$.*

Sketch of proof when D is smooth. Note that the result is trivial for constant SH-type orbits, as $i(\gamma, u_{\text{in}})$ is equal to a Morse index which is non-negative. For a Reeb-type orbit γ, we define u_{out} be the small cap passing through D. Then the orbit γ winds $\nu = u_{\text{out}} \cdot D$ times around D, so $i(\gamma, u_{\text{out}}) \approx 2\nu$. Thus we have

$$i(\gamma, u_{\text{in}}) = i(\gamma, u_{\text{out}}) - \lambda u_{\text{out}} \cdot D = (2 - \lambda)\nu \geq 0,$$

as required. $\square$

We now show that the $\widetilde{\mathcal{Q}}$-filtration is bounded below. To be precise, we need to show that for any i there exists $q(i)$ such that $\widetilde{\mathcal{Q}}_{\geq q(i)} \widetilde{SC}^{i}_{\Lambda} = 0$.[9] Indeed, we observe that for $i(\gamma \otimes e^a) = i$ fixed, we have

$$a_0 \widetilde{\mathcal{Q}}(\gamma \otimes e^a) = a = i(\gamma \otimes e^a) - i(\gamma, u_{\text{in}}) \leq i$$

by Proposition 1.15; thus we may take $q(i) = i/a_0$.

The following result is an immediate consequence of Theorem D and the Mayer–Vietoris property of relative symplectic cohomology [40]. However it also admits a simple direct proof using Proposition 1.15, which we feel is illuminating, so we give it here.

Proposition 1.16. *Suppose Hypothesis A is satisfied. Then the restriction map*

$$SH_M(M; \Lambda) \to SH_M(K_\sigma; \Lambda)$$

is an isomorphism for all $\sigma \in (0,1)$. In particular, $K_\sigma \subset M$ is SH-visible for all $\sigma \in (0,1)$ and $\mathbb{L}$ is weakly SH-visible, hence not stably displaceable from itself.

Proof. Note that we have $i(\gamma, u_{\text{in}}) \geq 0$ for any SH-type orbit, by Proposition 1.15. We also have $\mathcal{A}(\gamma, u_{\text{in}}) = h(\rho) - \rho \cdot h'(\rho) \leq 0$, where $\rho = \rho(\gamma)$, by the well-known formula [45, Section 1.2].[10] It follows that $i_{\text{mix}}(\gamma) \geq 0$. This inequality is satisfied for D-type orbits as well (recall Proposition 1.13), and therefore it is satisfied for all relevant one periodic orbits.

Now if we fix the index $i(\gamma, u) = i$, then the inequality $i_{\text{mix}}(\gamma) \geq 0$ yields an upper bound on the action: $\mathcal{A}(\gamma, u) \leq \kappa \cdot i$. Therefore the degreewise completion of the telescope complex has no effect:

$$\widehat{tel}(\mathcal{C}(H_\tau, J_\tau)) = tel(\mathcal{C}(H_\tau, J_\tau)).$$

It follows that $SH^*_M(M; \Lambda) \to SH^*_M(K_\sigma; \Lambda)$ is an isomorphism as required. $\square$

[9]The terminology is counterintuitive as our filtrations are decreasing, whereas the standard conventions for spectral sequences are for the filtrations to be increasing.
[10]Note that our conventions are different from Viterbo's.

1.5.3. Theorem D. To prove Theorem D, we need to consider the dependence of our constructions on the 'smoothing parameter' $R > 0$, so we include it in the notation. The proof starts with the same strategy that was used in the proof of [38, Theorem 1.24]. For R sufficiently small and σ sufficiently close to 1, $M \backslash K_\sigma^R$ is stably displaceable (this follows from an h-principle as popularized by McLean in [25]). Therefore, $SH_M\left(\overline{M \backslash K_\sigma^R}; \Lambda\right) = 0$ for such R, σ. We then prove that there exists a continuous function $\sigma_{\mathrm{crit}}^D(R)$, with $\sigma_{\mathrm{crit}}^D(0) = \sigma_{\mathrm{crit}}$, such that the following holds:

Proposition 1.17 (Proposition 5.14). *Let $\sigma_{\mathrm{crit}}(R) < \sigma_1 < \sigma_2 < 1$. Then, there exists an isomorphism*

$$SH_M^*\left(\overline{M \backslash K_{\sigma_1}^R}; \Lambda\right) \cong SH_M^*\left(\overline{M \backslash K_{\sigma_2}^R}; \Lambda\right).$$

In particular, $SH_M\left(\overline{M \backslash K_\sigma^R}\right) = 0$ for all $\sigma \in (\sigma_{\mathrm{crit}}^D(R), 1)$; as the compact sets $\left\{\overline{M \backslash K_\sigma^R}\right\}_{R>0, \sigma>\sigma_{\mathrm{crit}}^D(R)}$ exhaust $M \backslash K_{\mathrm{crit}}$, this implies that K_{crit} is SH-full.

The proof of Proposition 1.17 uses the 'contact Fukaya trick' of [38]. This allows us to set up acceleration data (H_τ, J_τ) for $\overline{M \backslash K_{\sigma_2}}$ and $(\tilde{H}_\tau, \tilde{J}_\tau)$ for $\overline{M \backslash K_{\sigma_1}}$, so that there is an isomorphism of Floer 1-rays $\mathcal{C}(H_\tau, J_\tau) \cong \mathcal{C}(\tilde{H}_\tau, \tilde{J}_\tau)$, which however need not respect action filtrations. The key to proving the Proposition, then, is to show that the action filtrations on the corresponding telescope complexes are topologically equivalent. The reason why this last step worked in [38] was the index-boundedness property (also popularized in [25]). In our setting we need estimates on the mixed index, which have a different nature.

1.6. Conjectures

1.6.1. Filtration on $QH^*(M; \Lambda)$. Note that, as an immediate corollary of Theorem B (3), there exists a filtration $\widetilde{\mathcal{Q}}_{\geq \bullet}$ on $QH^*(M; \Lambda)$ induced by the $\widetilde{\mathcal{Q}}$-filtration on $\left(\widetilde{SC}_\Lambda, \partial\right)$. (In general this is different from the 'obvious' filtration on $QH^*(M; \Lambda)$, i.e., the one with filtration map $\alpha \otimes r \mapsto \mathcal{Q}(r)$ for $\alpha \in H^*(M; \Bbbk)$, $r \in \Lambda$.) We give a conjectural description of this $\widetilde{\mathcal{Q}}$-filtration. Consider the function $f : M \to \mathbb{R}$ defined by

$$f(x) = \sum_{k : x \in D_k} \lambda_k - 2,$$

and set $M^j := \{f < j\}$.

Conjecture 1.18. *We have*

$$\widetilde{\mathcal{Q}}_{\geq j} H^i(M; \Lambda) \supset \ker(H^i(M; \Lambda) \to H^i(M^{j a_0 - i}; \Lambda)).$$

When Hypothesis A holds, this inclusion is an equality.

We first observe that the Conjecture is consistent with the fact that

$$q \cdot \widetilde{\mathcal{Q}}_{\geq j} H^i(M; \Lambda) = \widetilde{\mathcal{Q}}_{\geq j+1} H^{i+a_0}(M; \Lambda).$$

It is motivated by this together with the natural expectation that the isomorphism of Theorem B (3) sends

$$\mathrm{PD}(C) \mapsto \left[e^{\sum_{i\in I}\lambda_i} \cdot \mathrm{PSS}_{log}(C) + (\text{higher-order terms})\right],$$

where C is a cycle contained in D_I, and PSS_{log} is the log PSS map of [14]. Thus we expect $\widetilde{\mathcal{Q}}(\mathrm{PD}(C)) \geq \sum_{i\in I}\lambda_i/a_0$.

Remark 1.19. The filtration in Conjecture 1.18 exhibits intriguing parallels with the weight filtration in Hodge theory, c.f. [9,19].

1.6.2. Analogue of Theorem C in the absence of Hypothesis A. Let us consider the spectral sequence associated with the filtered complex $(\widetilde{SC}_\Lambda, \partial, \widetilde{\mathcal{Q}}_{\geq\bullet})$ of Theorem B. If Hypothesis A holds, then it converges to $QH^*(M;\Lambda)$ by Theorem C; but it is also interesting to study the spectral sequence when this Hypothesis does not hold.

As we saw in Sect. 1.5.2, the reason Hypothesis A is necessary for Theorem C to hold is that it guarantees the $\widetilde{\mathcal{Q}}$-filtration on $\widetilde{SC}_\Lambda$ is bounded below, and in particular complete. Let us denote by $(\overline{SC}_\Lambda, \partial)$ the completion of $(\widetilde{SC}_\Lambda, \partial)$ with respect to the $\widetilde{\mathcal{Q}}$-filtration. Note that taking the completion does not change the spectral sequence.

We give a conjectural description of $H^*(\overline{SC}_\Lambda, \partial)$, based on suggestions made to us independently by Pomerleano and Seidel. For each $i \in I$, define $QH^*(M;\Lambda)_i$ to be the 0-generalized eigenspace of the operator $\mathrm{PD}(D_i)\star(-)$ on $QH^*(M;\Lambda)$, where $\star$ denotes the quantum cup product. I.e., it is the subspace of $\alpha \in QH^*(M;\Lambda)$ such that $\mathrm{PD}(D_i)^{\star k}\star\alpha = 0$ for some k. We then define

$$QH^*(M;\Lambda)_{\mathrm{crit}} := \bigcap_{i:\lambda_i>2} QH^*(M;\Lambda)_i.$$

Conjecture 1.20. *We have $H^*(\overline{SC}_\Lambda, \partial) \cong QH^*(M;\Lambda)_{\mathrm{crit}}$. Furthermore, the resulting spectral sequence converges to $QH^*(M;\Lambda)_{\mathrm{crit}}$.*

As evidence for the conjecture, we use Conjecture 1.18 to argue that whenever $\lambda_i > 2$, the degree-0 class $c - e^{-2}\mathrm{PD}(D_i)$ is invertible in the $\widetilde{\mathcal{Q}}$-completed quantum cohomology, for any $c \neq 0$. Indeed its inverse is

$$\left(c - e^{-2}\mathrm{PD}(D_i)\right)^{-1} = c^{-1}\cdot\sum_{j=0}^{\infty}\left(c^{-1}e^{-2}\mathrm{PD}(D_i)\right)^{\star j},$$

which converges because $\widetilde{\mathcal{Q}}(e^{-2}\mathrm{PD}(D_i)) \geq (\lambda_i - 2)/a_0 > 0$. Therefore, any c-generalized eigenvector of $e^{-2}\mathrm{PD}(D_i)\star(-)$ dies in the $\widetilde{\mathcal{Q}}$-completion:

$$\left(c - e^{-2}\mathrm{PD}(D_i)\right)^{\star k}\star\alpha = 0 \qquad\Rightarrow\qquad \alpha = 0,$$

by multiplying on the left by the inverse.

Assuming that the $\Bbbk$-linear endomorphisms $e^{-2}\mathrm{PD}(D_i)\star(-)$ admit Jordan normal forms, the above argument suggests that only the 0-generalized eigenspaces can 'survive'. This gives some evidence for Conjecture 1.20 in the case that $\Bbbk$ is an algebraically closed field. It is reasonable to believe that one can bootstrap from there to the case of a general commutative ring $\Bbbk$. For the rest of this section we will assume that $\Bbbk$ is an algebraically closed field.

Remark 1.21. We strongly expect that $H^*(\overline{SC}_\Lambda, \partial)$ is nothing but the relative symplectic cohomology of the skeleton of X. There is an intriguing contrast between Conjecture 1.20 and Ritter's work [29]: precisely, let us consider the case that D is smooth and $\lambda > 2$, and let $\mathcal{N}$ be the total space of the inverse of the normal bundle to D. Then Conjecture 1.20 (together with the above expectation) says that $QH^*(M)_{\mathrm{crit}}$, which is the 0-generalized eigenspace of $QH^*(M)$, 'lives on the skeleton of X'; whereas Ritter shows that $SH^*(\mathcal{N})$ is the quotient of $QH^*(\mathcal{N})$ by its 0-generalized eigenspace. Note that we can obtain $\mathcal{N}$ from the Liouville completion $\widehat{X}$ of X by replacing a neighbourhood of the skeleton with a copy of D (more precisely, the symplectic cut of $\widehat{X}$ along the hypersurface $\{\rho = 1\}$ is $M \coprod \mathcal{N}$).

Remark 1.22. In light of Venkatesh's quantitative generalization of Ritter's results [41], we expect that considering Liouville domain neighborhoods V of the skeleton of varying sizes (vaguely speaking, 'in the directions of the components of the divisors'), one might observe that additional simultaneous generalized eigenspaces start contributing to $SH^*_M(V;\Lambda)$. It might be possible to interpret Theorem D as the other end of this size dependence: if the size of V is large enough in all directions (e.g., if it contains K_{crit}), then all simultaneous generalized eigenspaces contribute to $SH^*_M(V;\Lambda)$.

Further evidence for Conjecture 1.20 is provided in [8], in the case $M = \mathbb{CP}^1 \times \mathbb{CP}^1$, where D is a $(1,1)$ hypersurface: indeed the conjecture is confirmed in this case. We discuss further examples in Sects. 1.7.4 and 1.7.5 below.

We now recall a variation on the definition of relative symplectic cohomology from [38, Remark 1.8]. The relative symplectic cohomology $SH^*_M(K;\Lambda)$ is a module over $SH^*_M(M;\Lambda) = QH^*(M;\Lambda)$, via the restriction map. For any idempotent $a \in QH^0(M;\Lambda)$, we define the 'a-relative symplectic cohomology of K' to be $a \cdot SH^*_M(K;\Lambda)$. We define corresponding properties of subsets of M: a-SH-visible, a-SH-full, etc.

Lemma 1.23. *The subspace $QH^*(M;\Lambda)_{\mathrm{crit}} \subset QH^*(M;\Lambda)$ is an ideal which is generated by an idempotent a.*

Proof. We first observe that for any even element α in a supercommutative Frobenius algebra, the decomposition into generalized eigenspaces of $\alpha \star (-)$ is orthogonal (with respect to the pairing and the algebra structure), and hence the generalized eigenspaces are ideals generated by idempotents. It follows for each i, the subspace $QH^*(M;\Lambda)_i$ is an ideal generated by an idempotent; so the intersection is an ideal generated by the product of these idempotents. □

Conjecture 1.24. *Under the same hypotheses as for Theorem D (without assuming Hypothesis A), the skeleton $\mathbb{L}$ is a-SH-full, where a is the idempotent from Lemma 1.23.*

Conjecture 1.24 implies, for example, that $\mathbb{L}$ must intersect every a-Floer-theoretically essential (over $\mathbb{k}$) monotone Lagrangian, where the latter condition means that $CO(a \otimes_\Lambda \mathbb{k}) \in HF^0(L; \mathbb{k})$ is non-zero. (Here we have used the algebra homomorphism $\Lambda \to \mathbb{k}$, which sends $q \mapsto 1$, to define an idempotent $a \otimes_\Lambda \mathbb{k} \in QH^0(M; \mathbb{k})$).

1.6.3. Maurer–Cartan element. For the purpose of this section, we assume that $\mathbb{k}$ is a field of characteristic zero, and we assume that Hypothesis A holds.

Recall that the symplectic cochain complex $SC^*(X; \mathbb{k})$ carries an L_∞ structure [13]. This consists of a sequence of operations $\ell^k : SC^*(X; \mathbb{k})^{\otimes k} \to SC^*(X; \mathbb{k})$ of degree $3 - 2k$, satisfying the L_∞ relations; and $\ell^1 = d$ is the standard differential. We extend these linearly to make SC^*_Λ into an L_∞ algebra. We recall that a *Maurer–Cartan element* for the L_∞ algebra (SC^*_Λ, ℓ^k) is an element $\beta \in \mathcal{Q}_{\geq 1} SC^2_\Lambda$, satisfying the Maurer–Cartan equation:

$$\sum_k \frac{\ell^k(\beta, \ldots, \beta)}{k!} = 0.$$

We remark that this is in fact a finite sum, because the terms live in successively higher levels of the $\mathcal{Q}$-filtration, which Hypothesis A ensures is bounded below (see Sect. 1.5.3).

A Maurer–Cartan element β can be used to deform the L_∞ structure to get a new one ℓ^k_β on SC_Λ (see, e.g. [16, Section 4]). In particular, the resulting operation ℓ^1_β defines a new differential on SC_Λ.

Conjecture 1.25. *There exists a Maurer–Cartan element $\beta \in SC^2_\Lambda$ such that in the statement of Theorem B, we may take $\widetilde{SC}_\Lambda = SC_\Lambda$ and $\partial = \ell^1_\beta$.*

Remark 1.26. Cieliebak and Latschev have outlined ideas closely related to Conjecture 1.25 (but in a more general context) in talks as far back as 2014.

Remark 1.27. Moreover, one expects that Floer-theoretic operations on quantum cohomology of M (such as the quantum cup product) are deformations of the corresponding operations on symplectic cohomology of X by β, c.f. [12].

Remark 1.28. In the proof of Theorem B presented in this paper, we need to replace SC_Λ with $\widetilde{SC}_\Lambda$. Conjecture 1.25 suggests an alternative proof, in which no such replacement is necessary. The cost is that the construction is significantly more elaborate, relying on the L_∞ structure and a version of the homotopy transfer theorem, which makes it harder to see the key geometric ideas, which are the same in both proofs.

Remark 1.29. It is natural to envision generalizations of our results, as well as of Conjecture 1.25, where M is allowed to be only a partial compactification of X; and furthermore, where some of the weights λ_i are allowed to be equal to 0. We present several examples in Sect. 1.7 below which illustrate such a generalization. For example, Remark 1.41 gives evidence for this generalized conjecture in the case $M = T^*\mathbb{RP}^2$, with $D \subset M$ a smooth divisor equipped with weight $\lambda = 0$; the generalized conjecture in this case says that $SC^*(M;\Bbbk)$ is a 'deformation' of $SC^*(X;\Bbbk)$ (note that there is no need for a Novikov ring in the definition of symplectic cohomology of M, as it is exact). We put scare quotes around 'deformation' because when the weights are 0, the extra terms in the deformed differential may simply preserve the $\mathcal{Q}$-filtration, rather than strictly increasing it; so there is no sense in which they are 'small'. To make a useful version of the conjecture one would need an alternative to the $\mathcal{Q}$-filtration, which *is* strictly increased by the extra terms; it would probably be defined in terms of the grading.

Note that the projection of β to $\mathrm{Gr}_1 SC^2_\Lambda$ is d-closed, and hence defines a class $[\beta_1] \in \mathrm{Gr}_1 SH^2(X;\Lambda)$. It is immediate from Conjecture 1.25 that the differential on the E_1 page of the spectral sequence is given by $[[\beta_1], -]$, where $[-,-]$ denotes the Lie bracket on $SH^*(X;\Bbbk)$.

We now explain how our conjectures connect with work of Tonkonog [37]. Tonkonog considers the following setup: $\bar{M}$ is a compact Fano variety equipped with its monotone Kähler form, $\bar{D} \subset \bar{M}$ a simple normal crossings anticanonical divisor, $X = \bar{M} \backslash \bar{D}$, and $M = \bar{M} \backslash \cup_{i=1}^J \bar{D}_i$ is a partial compactification of X, with compactifying divisor $D = M \cap \bar{D}$. Tonkonog defines a class $\mathcal{BS} \in SH^0(X;\Bbbk)$ by counting pseudoholomorphic 'caps' in M, such that the following holds:

Theorem 1.30 (Theorem 6.5 in [37]). *For any exact closed Lagrangian $L \subset X$ equipped with a $\Bbbk^*$-local system ξ, we have $\mathcal{CO}(\mathcal{BS}) = \mathfrak{m}^0_{(L,\xi)}$, where $\mathcal{CO} : SH^*(X;\Lambda) \to H^*(L;\Lambda)$ is the closed–open map, and $\mathfrak{m}^0_{(L,\xi)} \in H^2(L;\Lambda)$ is the disc potential.*

This fits into the generalized geometric setup alluded to in Remark 1.29 (we are in the log Calabi–Yau setting, and we equip each component of D with its canonical weight 2). It connects with our conjectures as follows:

Conjecture 1.31. *We have $\mathcal{BS} = [\beta_1]$.*

In many settings, we can tightly constrain the class β using grading considerations. For each i we can define a cocycle $B_i \in SC^{2-\lambda_j}(X;\Bbbk)$ by 'counting caps passing through D_i', following [37] or [15]. We define

$$B := \sum_i e^{\lambda_i} \cdot B_i \in SC^2(X;\Lambda).$$

Conjecture 1.32. *Suppose we are in the log Calabi–Yau case: i.e., $\lambda_i = 2$ for all i, and furthermore that the minimal Chern number of M is ≥ 2. Then we have $\beta = B$.*

Remark 1.33. If the minimal Chern number of M is 1, then we conjecture that $\beta = B + e^2 \cdot B_0$, where $B_0 \in SC^0(X;\Bbbk)$ is a multiple of the unit, and counts certain holomorphic spheres in M of Chern number 1. Note that the additional term B_0 is irrelevant for the purposes of Conjecture 1.25, as $\ell^1_B = \ell^1_{B+B_0}$ using the fact that B_0 is a multiple of the unit.

As evidence for the Conjecture, we first observe that $\mathrm{Gr}_1^{\mathcal{Q}} SC^2_\Lambda$ is generated by the classes qB_i, together with the unit $q \cdot 1$; and argue that the coefficient of the unit in β must count certain Chern-number-1 spheres. We further observe that $\mathcal{Q}_{\geq 2} SC^2_\Lambda = 0$. This follows as we have $a_0 = 2$, so any generator $\gamma \otimes q^j$ of SC^2_Λ with $j \geq 2$ must have $i(\gamma) \leq -2$; however, $i(\gamma) \geq 0$ by Proposition 1.15.

Remark 1.34. Based on [34, Lemma 6.4], we also expect Conjecture 1.32 to hold under either of the following hypotheses:

- D is smooth and Hypothesis A is satisfied.
- M is a projective variety, D a complex divisor, and $c_1(TM)$ lies in the interior of the cone $Amp'(M, D) \subset H^2(M; \mathbb{R})$ defined in [34, Definition 3.26].

In settings where Conjecture 1.32 holds, the Maurer–Cartan element β is determined up to gauge equivalence by the cohomology classes $[B_i]$. Furthermore, the components of β get 'turned on' one by one as the corresponding divisors get added compactifying X.

1.6.4. Mirror symmetry in the log Calabi–Yau case. Let us consider the log Calabi–Yau case, where $X = M \backslash D$ and X is equipped with its preferred Liouville structure and trivialization of canonical bundle. In this case we have $a_0 = 2$, so $\Lambda = \Bbbk[q, q^{-1}]$, where $i(q) = 2$.

Assume that Y is a mirror scheme to X over $\Bbbk$, which is smooth. Even though we choose to leave what this means vague, we will assume that it implies

$$SH^i(X; \Bbbk) \simeq \bigoplus_{p+q=i} H^q(Y, \Lambda^p TY), \tag{1.3}$$

and in particular

$$SH^0(X; \Bbbk) \simeq H^0(Y, \mathcal{O}_Y).$$

Therefore, the classes $B_i \in SH^0(X; \Bbbk)$ are mirror to functions $w_i \in H^0(Y, \mathcal{O}_Y)$. We set $W := \sum_i w_i$. This sum includes the constant term w_0, which may be non-zero in the case that the minimal Chern number of M is 1.

Now let Y_Λ denote the base change of Y to Λ, and $W_\Lambda = qW$ be a function on Y_Λ.

Conjecture 1.35. *The Landau–Ginzburg model (Y_Λ, W_Λ) is mirror to M.*

Remark 1.36. In fact, Conjecture 1.35 should extend beyond the log Calabi–Yau case we consider here. However, it becomes difficult (and confusing) to interpret the mirror in terms of the language of classical algebraic geometry: the polyvector fields on Y_Λ are given a non-standard grading, and in general W_Λ may be a polyvector field rather than a function. In contrast, in the log Calabi–Yau case one can give a transparent interpretation of Conjecture 1.35 in terms of the classical algebraic geometry of the Landau–Ginzburg model (Y, W) defined over $\Bbbk$, which we now do. (We discuss the non-log-Calabi–Yau case in Remark 1.39 at the end of this section.)

We consider the Koszul complex associated with the section dW of T^*Y:

$$K(dW) := \left\{ \cdots \to \Lambda^{p+1}(TY) \xrightarrow{dW} \Lambda^p(TY) \to \cdots \to TY \xrightarrow{dW} \mathcal{O}_Y \right\}.$$

This is a complex of vector bundles over Y. When the critical locus $Z := \mathrm{Crit}(W)$ is isolated, $K(dW)$ is a resolution of $\mathcal{O}_Z$, and therefore, its hypercohomology gives the algebra of functions on the critical locus: $\mathbb{H}^*(K(dW)) \cong \mathcal{O}(Z)$ (the hypercohomology is concentrated in degree $* = 0$). In general, we define $\mathcal{O}(Z^h) := \mathbb{H}^*(K(dW))$, because this hypercohomology is, essentially by definition, the graded algebra of functions on the 'derived critical locus of W' (see e.g. [42]).

Conjecture 1.35 implies, among other things, that we have an isomorphism of graded Λ-algebras

$$\mathcal{O}(Z^h) \otimes_{\Bbbk} \Lambda \cong QH^*(M; \Lambda). \tag{1.4}$$

We expect that the mirror to the spectral sequence of Theorem C on the RHS, is the hypercohomology spectral sequence on the LHS, in a sense we now make clear.

We recall the construction of the hypercohomology spectral sequence

$${}^I E_1^{p,q} = H^q(\Lambda^{-p} TY) \Rightarrow \mathcal{O}(Z^h),$$

following [46, Section 5.7]. We take a Cartan–Eilenberg resolution $\mathcal{C}^{p,q}$ of $K(dW)$, and consider the resulting bicomplex $C^{p,q} = \Gamma(\mathcal{C}^{p,q})$. We define a filtration map on this complex by $\mathcal{Q}(c) = p$ for $c \in C^{p,q}$ (i.e., we have $\mathcal{Q}(c) = -p$ for c a section of $\Lambda^p TY$). The resulting $\mathcal{Q}$-filtration induces the spectral sequence with E_1 page as above. The differential on the E_1 page is given by contracting with dW.

We now consider the bicomplex $C^{p,q} \otimes_{\Bbbk} \Lambda$, and equip it with the filtration map $\mathcal{Q}(c \otimes r) = \mathcal{Q}(c) + \mathcal{Q}(r)$. We conjecture that the resulting filtered complex is filtered quasi-isomorphic to $(\widetilde{SC}_{\Lambda}, \partial, \widetilde{\mathcal{Q}}_{\geq \bullet})$, and in particular the corresponding spectral sequence is isomorphic to the one from Theorem C. As evidence, we compute that the spectral sequence has

$$\begin{aligned} E_1^{j,k} &= \bigoplus_{p+q=3j+k} H^q(\Lambda^p TY) \otimes_{\Bbbk} \Bbbk \cdot q^{-j+p} \\ &\cong SH^{3j+k}(X; \Bbbk) \otimes_{\Bbbk} \Bbbk \cdot q^{-j+p}, \end{aligned}$$

which is clearly isomorphic to the E_1 page of the spectral sequence from Theorem C.

Remark 1.37. The attentive reader may notice the presence of an extra 'p' in the exponent of q, compared with the E_1 page from Theorem C. This is because the isomorphism of E_1 pages

$$\begin{aligned} SH \otimes_{\Bbbk} \Lambda &= \bigoplus_{q,p} H^q(\Lambda^p TY) \otimes_{\Bbbk} \Lambda \qquad \text{sends} \\ SH \otimes_{\Bbbk} \Bbbk &\mapsto \bigoplus_{q,p} H^q(\Lambda^p TY) \otimes_{\Bbbk} \Bbbk \cdot q^{-p}. \end{aligned}$$

This reflects the fact that $\mathcal{Q}(c) = -p$ for $c \in \Lambda^p TY$.

We now explain how this fits with the picture from the previous section. The isomorphism (1.3) is expected to respect the natural graded Lie algebra structures on both sides (among other things), where the Lie bracket on the polyvector field cohomology is given by the Schouten–Nijenhuis bracket. The differential on the E_1 page of the symplectic spectral sequence is given by $[B, -]$. The differential on the E_1 page of the hypercohomology spectral sequence is given by contraction with qdW, which coincides with $[qW, -]$ (as one can see from the definition of the Schouten–Nijenhuis bracket); thus the two differentials match.

More precisely, we expect that the isomorphism of Lie algebras (1.3) can be refined to a quasi-isomorphism of L_∞ algebras, and the Maurer–Cartan element β matches with the Maurer–Cartan element qW up to gauge equivalence. This would yield a chain-level quasi-isomorphism underlying (1.4), which would imply the isomorphism of spectral sequences discussed above.

Note that when Y is affine, there is no need to take a Cartan–Eilenberg resolution: we may take $C^{p,0} = \Gamma(\Lambda^{-p}TY)$ and $C^{p,q} = 0$ for $q \neq 0$, with differential given by contracting with dW, and the bicomplex is simply a complex. In particular, the hypercohomology spectral sequence degenerates at E_2. This leads us to make the following:

Conjecture 1.38. *If X in addition (to the conditions from the first paragraph of this section) admits a homological Lagrangian section and $SH^0(X; \Bbbk)$ is a smooth algebra, then the spectral sequence of Theorem C degenerates at E_2 page.*

Under these assumptions on X one can take Y to be the smooth affine scheme $Spec(SH^0(X; \Bbbk))$ (see [27]), which would satisfy (1.3), which is our reason to make this conjecture.

For example, the conjecture holds in the toric Fano examples (see Sect. 1.7.1), essentially by the argument given above. This degeneration also follows from the fact that one can construct $SC^*(X; \Bbbk)$ with zero differential in this case!

Remark 1.39. We now discuss the non-log-Calabi–Yau case of Conjecture 1.35, which will appear in several examples in Sect. 1.7 below. There are three complicating factors:

(1) The mirror to X will in general be a Landau–Ginzburg model (Y, w), rather than simply a variety Y;
(2) The algebra of polyvector fields on Y must be equipped with a non-standard grading;
(3) *a priori*, β will be mirror to a gauge equivalence class of Maurer–Cartan elements for the differential graded Lie algebra of polyvector fields on (Y, w), rather than simply a function W on Y.

Issue (2) is already present if one wants to talk about the mirror of T^*S^1 with a non-standard trivialization of its canonical bundle and then consider the correspondence between compactifications and deformations. In this case one cannot use a traditional SYZ approach as the zero section of T^*S^1 does not even have vanishing Maslov class with respect to such a trivialization.

It seems that to develop some general geometric intuition in the non-log Calabi–Yau cases, it would be helpful to use the language of derived algebraic geometry but we do not feel comfortable enough to do this at this point.

Concerning issue (3), we actually expect that β should be mirror to a function in broad generality, although it is not clear how to prove this. In some cases, it follows from grading considerations, as in Conjecture 1.32 and the ensuing remarks.

Remark 1.40. Even though we avoid a general discussion, we do use our expectations in the log Fano case in some examples in Sect. 1.7 below. Here is our starting ansatz in these examples: start with a log Calabi–Yau pair (M, D'), where

$$D' = \bigcup_{i=1}^{N+J} D'_i, \qquad \text{and set} \qquad D = \bigcup_{i=1}^{N} D'_i.$$

Suppose that $X' = M \backslash D'$ is mirror to Y as at the start of this section. This means that we could choose all weights $\lambda'_i = 2$; we assume, however, that there exists a valid choice of weights with $\lambda'_i > 0$ for all $1 \leq i \leq N$, and $\lambda'_i = 0$ for $N+1 \leq i \leq N+J$. We equip X' with the trivialization of its canonical bundle corresponding to these weights, and equip the algebra $SH^0(X'; \Bbbk)$ with its induced grading. We posit that this is the graded algebra of functions on the mirror of X' (with the alternative trivialization), which we regard as a 'graded scheme'. We set $X = M \backslash D$, and posit that the mirror to X is (Y, w) where $w = \sum_{i=N+1}^{N+J} w_i$. We furthermore posit that the Maurer–Cartan element β corresponding to $X \subset M$ is mirror to $W_\Lambda = \sum_{i=1}^{N} e^{\lambda_i} w_i$, and therefore that the mirror to M is $(Y_\Lambda, w + W_\Lambda)$.

1.7. Examples

1.7.1. Fano toric varieties. Let $\Delta \subset \mathbb{R}^n$ be a Fano Delzant polytope. This means that it is a Delzant polytope equal to the intersection of half-spaces (with no redundancy)

$$\nu_i(x) + 2\kappa \geq 0, \qquad i = 1, \ldots, m$$

for $\kappa > 0$ and $\nu_i \in (\mathbb{Z}^n)^\vee$ primitive. Using the symplectic boundary reduction construction (one of the many options), we construct a symplectic manifold (M_Δ, ω) with a Hamiltonian T^n action and moment map

$$\pi : M_\Delta \to \mathbb{R}^n.$$

The image of the moment map is by construction Δ. Finally, note that M_Δ satisfies the monotonicity condition $2\kappa c_1(TM_\Delta) = [\omega]$.

We define the toric SC divisor D_Δ as the preimage of the boundary of Δ under the moment map. Note that $D_\Delta = \bigcup_{i=1}^{m} D_i$ is automatically an orthogonal SC divisor. We define $X_\Delta = M_\Delta \backslash D_\Delta$. Again by construction X_Δ is a product $int(\Delta) \times (\mathbb{R}^n)^\vee / (\mathbb{Z}^n)^\vee$. Denoting the coordinates on $\mathbb{R}^n$ by $x_1, \ldots, x_n$ and the circle valued coordinates on $(\mathbb{R}^n)^\vee / (\mathbb{Z}^n)^\vee$ by $\phi_1, \ldots, \phi_n$, we have

$$\omega|_X = \sum \mathrm{d}x_i \mathrm{d}\phi_i.$$

We note the short exact sequence

$$0 \longrightarrow H^1(X_\Delta;\mathbb{R}) \xrightarrow{f} H^2(M_\Delta, X_\Delta;\mathbb{R}) \xrightarrow{g} H^2(M_\Delta;\mathbb{R}) \longrightarrow 0.$$

A choice of weights is (as always) equivalent to the choice of a rational class

$$\boldsymbol{\lambda} \in H^2(M_\Delta, X_\Delta;\mathbb{R}) \cong \mathbb{R}^m,$$

which is sent to $2c_1(TM_\Delta)$ by g and which has positive coordinates. We have a preferred lift given by

$$\boldsymbol{\lambda}^{\mathrm{can}} = (2, \ldots, 2).$$

Let us also use the natural isomorphism $H^1(X_\Delta;\mathbb{R}) \cong \mathbb{R}^n$. The map f is easily computed to be

$$x \mapsto \nu_i(x).$$

Hence, the set of all possible positive weights is the image of the rational points in the interior of $\frac{1}{\kappa}\Delta$ under the map $\mathbb{R}^n \to \mathbb{R}^m$ given by

$$(x_1, \ldots, x_n) \mapsto (\nu_1(x) + 2, \ldots, \nu_f(x) + 2).$$

We see that the only weight that satisfies Hypothesis A is the canonical weight, which corresponds to $0 \in \frac{1}{\kappa}\Delta$.

Now let us outline how Theorems B and C work in this context, assuming the conjectural results of Sect. 1.6.3. We can arrange that

$$SC^*(X_\Delta;\Bbbk) \cong \Bbbk[z_1^{\pm 1}, \ldots, z_n^{\pm 1}, \partial/\partial z_1, \ldots, \partial/\partial z_n]$$

where the variables z_i are commuting and have degree 0, and the variables $\partial/\partial z_i$ are anticommuting and have degree 1 (where the degrees are induced by $\boldsymbol{\lambda}^{\mathrm{can}}$). We can also arrange that the L_∞ structure is trivial, with the exception of the Lie bracket ℓ^2, which coincides with the Schouten–Nijenhuis bracket. We can compute, for instance via Theorem 1.30 and Cho–Oh's computation of the disc potential of toric Fano varieties [5], that $\beta = qW$, where

$$W = \sum_i z^{\nu_i}.$$

Now Conjecture 1.25 says that in the statement of Theorem B, we can take

$$\widetilde{SC}_\Lambda = SC_\Lambda = \Lambda[z_1^{\pm 1}, \ldots, z_n^{\pm 1}, \partial/\partial z_1, \ldots, \partial/\partial z_n], \qquad \text{with} \quad \partial = [qW, -].$$

As explained in Sect. 1.6.4, this is the Koszul complex for dW, tensored with Λ. One can show that W has isolated singularities, so the cohomology of the Koszul complex is

$$\mathcal{O}(Z) = \frac{\Bbbk[z_1^{\pm 1}, \ldots, z_n^{\pm 1}]}{\left\langle \frac{\partial W}{\partial z_1}, \ldots, \frac{\partial W}{\partial z_n} \right\rangle} = Jac(W).$$

Thus, assuming Conjecture 1.25, Theorem B gives

$$QH^*(M_\Delta;\Lambda) \cong H^*(SC_\Lambda, [qW, -]) \cong Jac(W) \otimes_\Bbbk \Lambda,$$

which is the familiar statement of closed-string mirror symmetry for toric Fano varieties, c.f. [4]. Note that the spectral sequence of Theorem C has $E_0 = E_1 = SC_\Lambda$, $E_2 = Jac(W)\otimes_{\Bbbk}\Lambda$, and degenerates at E_2 because the differential on SC_Λ vanishes (or alternatively, because $Jac(W)$ is concentrated in even degree).

Now let us outline how Theorem D works in this context. For each $i = 1,\dots,m$, we have a Hamiltonian circle action with moment map $\nu_i \circ \pi$, which rotates around D_i, and these actions commute on the overlaps. It follows that they define a system of commuting Hamiltonians for D_Δ, in the sense of Sect. 2. For any $p \in int(\Delta)$ we define the corresponding weights $\boldsymbol{\lambda}_p := f(\frac{p}{\kappa}) + \boldsymbol{\lambda}^{\mathrm{can}}$ and primitive (of $\omega|_{X_\Delta}$)

$$\theta_p = \sum(x_i - p_i)\mathrm{d}\phi_i.$$

The relative de Rham class of (ω, θ_p) is easily seen to be $f(p) + \kappa\boldsymbol{\lambda}^{\mathrm{can}} = \kappa\boldsymbol{\lambda}_p$. The Liouville vector field corresponding to θ_p is $Z_p = \sum(x_i - p_i)\partial/\partial x_i$. It follows that θ_p is adapted to the system of commuting Hamiltonians in the sense of Sect. 2. The skeleton $\mathbb{L}_p$ for θ_p is nothing but the Lagrangian torus above p. The corresponding subset $K_{\mathrm{crit},p}$ is easily computed to be $\pi^{-1}(\tilde{K}_{\mathrm{crit},p})$, where $\tilde{K}_{\mathrm{crit},p} \subset \Delta$ is the smallest rescaling of Δ, centred at p, which contains the origin. In particular, $K_{\mathrm{crit},p}$ coincides with $\mathbb{L}_p$ if and only if $\boldsymbol{\lambda}_p = \boldsymbol{\lambda}^{\mathrm{can}}$, if and only if Hypothesis A is satisfied.

Our Theorem D says that the monotone torus fiber $\mathbb{L}_0$ is SH-full. It follows that it is not stably displaceable. This result can also be obtained using Lagrangian Floer theory, using the fact that the disc potential always has a critical point in this case. Our result says nothing about the skeleta $\mathbb{L}_p$ for $p \neq 0$. Indeed it is known that for $n \leq 3$ all of these non-monotone fibers are displaceable by probes [22, Corollary 3.9 and Proposition 4.7].

The fact that $\mathbb{L}_0$ is SH-full also implies that it intersects every Floer theoretically essential (over some commutative ring) monotone Lagrangian. This result also follows from the fact that $\mathbb{L}_0$, equipped with appropriate local systems, split-generates each component of the monotone Fukaya category over an arbitrary field [11, Corollary 1.3.1].

1.7.2. Skeleta in S^2. Let us move on to a non-toric example. Consider S^2 with a symplectic structure ω such that $[\omega] = 4\kappa\mathrm{PD}(\mathrm{pt})$. Let D be the union of N distinct points $p_1,\dots,p_N \in S^2$. Consider weights $\lambda_1,\dots,\lambda_N > 0$, which needs to satisfy

$$\lambda_1 + \cdots + \lambda_N = 4.$$

Let θ be a primitive of ω on $S^2\backslash D$ compatible with the weights and with some choice of local moment maps for the circle actions rotating about the p_i. Let $\mathbb{L}$ be the induced skeleton. The complement $S^2 - \mathbb{L}$ is a disjoint union of open disks U_i, $i = 1,\dots N$, one for each point $p_1,\dots p_N$. $\mathbb{L}$ itself is the union of all critical points, homoclinical and heteroclinical orbits, and periodic orbits of the Liouville vector field by the Poincaré–Bendixson theorem. It is elementary to compute (using the compatibility with weights) that the symplectic area of U_i is equal to $\kappa\lambda_i$. If we restrict the function $\rho : M \to \mathbb{R}$

to the disc U_i, then it extends continuously to 0 along the boundary of the closed disk, it is equal to 1 at p_i, and it generates a Hamiltonian circle action rotating U_i about p_i.

Hypothesis A is satisfied if and only if no weight is bigger than 2, which means no disc U_i has area more than half the area of S^2. In this case the subset K_{crit} coincides with the skeleton $\mathbb{L}$. Otherwise, we have $\lambda_i > 2$ for some i, and K_{crit} is the union of $\mathbb{L}$ with a collar around the boundary of U_i, so that the rest of U_i has area equal to half the area of S^2. Theorem D says that K_{crit} is SH-full. This implies that it is not stably displaceable, and furthermore that no two such subsets can be disjoint from each other. It is easy to see explicitly that it is necessary to add the collar to K_{crit} in order for these results to hold.

1.7.3. The case $M = S^2$, $D =$ a point. Let $M = S^2$, and D be a single point. We start by sketching how Theorem B works in this case. It is possible to take simpler models for $SC^*(X;\Bbbk)$ and $\widetilde{SC}_\Lambda$ than those which appear in the actual proof of the Theorem.

We take a model for $SC^*(X;\Bbbk)$ which is isomorphic to $\Bbbk[z, z\theta]$ where z is a commutative variable of degree -2, and θ is anticommutative of degree 1. The generator 1 corresponds to the unique constant orbit, z^j to the fundamental cycle of the Reeb orbit going j times around D, and $z^j\theta$ to the point class of the same Reeb orbit. The differential d sends $z^j \mapsto 0$ and $z^j\theta \mapsto z^{j-1}$. In particular the cohomology vanishes: symplectic cohomology of the disc is zero.

We have $\Lambda = \mathbb{C}[q]$ with $i(q) = 4$. We take $\widetilde{SC}_\Lambda = SC_\Lambda$, and consider the deformed differential ∂, where $\partial - d$ sends $z^j \mapsto 0$ and $z^j\theta \mapsto qz^{j+1}$. The cohomology of this differential is free of rank 2 over Λ, with a basis given by 1 and qz. In particular, it is isomorphic to $QH^*(M;\Lambda)$, in accordance with Theorem B: the class 1 corresponds to $1 \in QH^0(M;\Lambda)$, and the class qz corresponds to $\mathrm{PD}(\mathrm{pt}) \in QH^2(M;\Lambda)$.

Theorem C does not apply in this case, because Hypothesis A is not satisfied: we have $\lambda = 4 > 2$. And indeed the conclusion of the Theorem fails, because we cannot have a spectral sequence with E_1 page vanishing, converging to $QH^*(M;\Lambda) \neq 0$. The reason the proof of Theorem C does not run is that the $\mathcal{Q}$-filtration on SC_Λ is not degreewise complete. For example, the classes $q^k z^{2k}$ all have degree 0, but their $\mathcal{Q}$-values go to $+\infty$. The convergence theorems for spectral sequences all require completeness, and indeed it could not be otherwise: taking the completion does not change the spectral sequence associated with a filtered complex, by inspection of the construction. It is easy to verify that the degreewise completion of (SC_Λ, ∂) is acyclic: for example,

$$1 = \partial\left(\sum_{j=0}^{\infty}(-qz^2)^j \cdot z\theta\right).$$

This confirms Conjecture 1.20 in this case, as $QH^*(S^2;\Lambda)_{\text{crit}} = 0$.

Theorem D simply says in this case that a disc occupying half the area of the sphere is SH-full, c.f. [39, Section 1.2.2].

We now offer another perspective on this computation following Remark 1.40, which will be useful in the next two sections. First, we take $M = S^2$ and $D' = D'_1 \cup D'_2$ to be an anticanonical divisor on S^2, where D'_1 and D'_2 are distinct points. If we equip each point with weight $\lambda'_i = 2$, then this is a special case of Sect. 1.7.1: we see $QH^*(M;\Lambda')$ as a deformation of $SC^*(X';\Bbbk)$ where $X' = M\backslash D'$. In this case $(SC_{\Lambda'}, \partial')$ is quasi-isomorphic to the complex

$$\Lambda'[x, x^{-1}, \partial_x],$$

where $\partial_x^2 = 0$, the generator q' of the Novikov ring is in degree 2, x is in degree 0, and ∂_x is in degree 1; the differential ∂ is $\Lambda'[x, x^{-1}]$-linear and sends

$$1 \mapsto 0, \qquad \partial_x \mapsto q'(1 - x^{-2}).$$

As expected, this chain complex is degree-wise complete with respect to the $\mathcal{Q}$-filtration and we obtain

$$QH^*(S^2;\Lambda') \cong \Lambda[x, x^{-1}]/\left\langle x^2 - 1\right\rangle.$$

Now we consider the case that $\lambda_1 = 0$, $\lambda_2 = 4$. Following the recipe of Remark 1.40, if $X = M\backslash D'_2$ then $SC^*(X;\Bbbk)$ should be quasi-isomorphic to $\Bbbk[x, x^{-1}, \partial_x]$, where $\partial_x^2 = 0$, x is in degree 2, and ∂_x is in degree -1; the differential d is $\Bbbk[x, x^{-1}]$-linear and sends

$$1 \mapsto 0, \qquad \partial_x \mapsto 1.$$

As expected, this chain complex is acyclic. The chain complex (SC_Λ, ∂) is quasi-isomorphic to $\Lambda[x, x^{-1}, \partial_x]$, with x and ∂_x graded as before, and the generator q of Λ in degree 4; the differential ∂ sends

$$1 \mapsto 0, \qquad \partial_x \mapsto 1 - qx^{-2}.$$

Note that as expected, we have an isomorphism of chain complexes

$$\begin{aligned} (SC_\Lambda, \partial) \otimes_\Lambda \Lambda' \cong (SC_{\Lambda'}, \partial') \qquad &\text{via the algebra map sending} \\ x \mapsto q'x, \\ x\partial_x \mapsto x\partial_x. \end{aligned}$$

We learned nothing new so far but we believe that this exercise might help unraveling the more complicated examples in the next two sections below.

1.7.4. The quadric in $\mathbb{CP}^2$. Consider $\mathbb{CP}^2$ with its Fubini-Study symplectic form, and D a smooth quadric with its canonical weight 3, which does not satisfy Hypothesis A. $\mathbb{L}$ in this case is the monotone Lagrangian $\mathbb{RP}^2$, which is known to be stably non-displaceable. On the other hand $\mathbb{RP}^2$ can be displaced from the Chekanov torus (see [47]), hence it is not SH-full for a general $\Bbbk$. As was pointed out to us by Leonid Polterovich, it is also known that $\mathbb{RP}^2$ is $[\mathbb{CP}^2]$-superheavy over $\mathbb{Z}/2$, see [10, Example 4.12].

Let us now test Theorem B and Conjecture 1.20 in this case, using the mirror picture outlined in Remark 1.40. The expectation, following [3, Section 5.2], is as follows.

Consider the graded ring

$$R := \Bbbk[x,y,z]/(z(xy-1)-1), \qquad \text{where } |x|=-1,\ |y|=1,\ |z|=0,$$

and consider elements $w_1 = y^2 z$ and $w_2 = x$ of R. We set $Y = Spec(R)$. Then X should be mirror to the Landau–Ginzburg model (Y, w_1) while M should be mirror to $(Y_\Lambda, w_1 + qw_2)$, where $|q| = 3$.

We expect $(SC^*(X;\Bbbk), d)$ to be quasi-isomorphic to

$$\left(\bigoplus_p \Lambda^p TY, [w_1, -]\right), \tag{1.5}$$

while (SC_Λ, ∂) should be filtered quasi-isomorphic to

$$\left(\bigoplus_p \Lambda^p TY \otimes_\Bbbk \Lambda, [w_1 + qw_2, -]\right),$$

with the filtration map given by $\mathcal{Q}(c \otimes q^a) = -p + a$ for $c \in \Lambda^p TY$. We can compute the cohomology of this complex: it comes out as the Jacobian ring of $w_1 + qw_2$, which is

$$\begin{aligned}
&\Lambda[x,y,z]/(q - y^3 z^2, 2yz - xy^2 z^2, z(xy-1)-1)\\
&\quad\cong \Lambda[x,y,z]/(z-1, x - 2q^{-1}y^2, y^3 - q)\\
&\quad\cong \Lambda[y](y^3 - q)\\
&\quad\cong QH^*(\mathbb{CP}^2;\Lambda).
\end{aligned}$$

This agrees with Theorem B in this case.

Now we turn to Conjecture 1.20. We consider two cases:

Case 1: 2 is invertible in $\Bbbk$. We easily deduce that $1 - q(x/2)^3$ is nullhomologous; it is also clearly invertible in the $\mathcal{Q}$-completion. This implies that the cohomology vanishes after $\mathcal{Q}$-completion.

Case 2: $2 = 0$ **in** $\Bbbk$. In this case the Jacobian ring is $\Lambda[x,y,z]/(z-1, x, y^3 - q) = \Lambda[y]/(y^3 - q)$. It is easy to see that $\mathcal{Q}$-completion does not change the cohomology.

Both cases are in agreement with Conjecture 1.20: if 2 is invertible, then $\mathrm{PD}(D) \star (-)$ is invertible, so $QH^*(M;\Lambda)_{\mathrm{crit}} = 0$. On the other hand $[D]$ is 2-divisible, so if $2 = 0$, then $QH^*(M;\Lambda)_{\mathrm{crit}} = QH^*(M;\Lambda)$.

This leads us to conjecture that $\mathbb{RP}^2 \subset \mathbb{CP}^2$ is SH-full if the characteristic of $\Bbbk$ is 2, but not otherwise (Entov's result that $\mathbb{RP}^2$ is $[\mathbb{CP}^2]$-superheavy over $\mathbb{Z}/2$ can be considered as further evidence for this conjecture). This would imply that $\mathbb{RP}^2$ is non-stably displaceable (which is known), and intersects any monotone Lagrangian which is Floer-theoretically essential over a field of characteristic 2 (note that this does *not* include the Chekanov torus, as can easily be seen from the superpotential computed in [3]).

Remark 1.41. We sketch some evidence for the mirror symmetry statement (1.5), in the case that $\mathrm{char}(\Bbbk) = 2$. Note that the completion of X is symplectomorphic to $T^*\mathbb{RP}^2$, so $SH^*(X;\Bbbk) \cong H^*(\mathcal{L}\mathbb{RP}^2;\Bbbk)$ by Viterbo's theorem

[1,43,44]. We can compute

$$H^*(\mathcal{L}\mathbb{RP}^2;\Bbbk) \cong H^*(\mathbb{RP}^2;\Bbbk) \oplus \bigoplus_{k>1} H^*(S(T\mathbb{RP}^2);\Bbbk)[k]$$

by [48], where the first factor comes from the manifold $\mathbb{RP}^2$ of constant loops, and the subsequent factors come from the manifolds $S(T\mathbb{RP}^2)$ of 'length-k' geodesics. Of course $H^*(\mathbb{RP}^2;\Bbbk) \cong \Bbbk[y]/y^3$ with $|y| = 1$, while $H^*(S(T\mathbb{RP}^2);\Bbbk)$ has rank 1 in degrees $0, 1, 2, 3$. On the other hand, one may compute that

$$\begin{aligned} H^*\left(\bigoplus_p \Lambda^p TY, [w_1, -]\right) &\cong \Bbbk[x, y, v]/(y^3, y^2x, y^2v) \\ &= \Bbbk[y]/y^3 \oplus \bigoplus_{k>1} \langle x^k, x^k y, x^{k-1} \cdot v, x^{k-1} y \cdot v\rangle, \end{aligned}$$

where $v = x\partial_x - y\partial_y$ is an anticommuting variable. We identify $\Bbbk[y]/y^3$ as corresponding to the constant loops, and the subsequent factors as corresponding to the length-k geodesics. The degrees match up (we observe that $|v| = 1$). We remark that $x = w_2$ is the basic loop around D, which corresponds to the family of length-1 geodesics, so it makes sense that multiplying by x takes us to the next k-value.

1.7.5. Fano hypersurfaces. We consider some examples motivated by [33]. They follow a similar philosophy to Remark 1.40, but are a bit different as they are obtained by partially compactifying an affine variety which is of log general type, rather than being log Calabi–Yau.

Let $M = M_{n,a}$ be a smooth hypersurface of degree $a \le n+1$ in $\mathbb{CP}^{n+1}$, and $D = D_{n,a,i}$ a union of $i \le n+2$ generic hyperplanes. This fits into the setup of Sect. 1.1, and we may take the weights all to be equal to $\lambda = \frac{2(n+2-a)}{i}$. In particular, Hypothesis A is satisfied if and only if $n+2-a \le i$. This corresponds to the variety $X_{n,a,i} = M_{n,a} \setminus D_{n,a,i}$ being log Calabi–Yau (in the case of equality) or log general type (otherwise). Hypothesis A is not satisfied precisely when $X_{n,a,i}$ is log Fano.

We conjecture that the mirror to $X_{n,a,i}$ is the Landau–Ginzburg model $(Y_{n,a,i}, W_{n,a,i})$, where

$$\begin{aligned} Y_{n,a,i} &= [\Bbbk^{n+2}/G_{n,a,i}] \qquad \text{is a stack, where} \\ G_{n,a,i} &= \ker\left(\mathbb{Z}_a^{n+1} \xrightarrow{\Sigma} \mathbb{Z}_a\right), \qquad \text{and} \\ W_{n,a,i} &= -z_1 \dots z_{n+2} + \sum_{j=i+1}^{n+2} z_j^a, \qquad \text{and furthermore that} \\ \beta_{n,a,i} &= q \cdot \sum_{j=1}^{i} z_j^a. \end{aligned}$$

Here we assume that $\Bbbk$ contains all ath roots of unity. The group $G_{n,a,i}$ acts torically, preserving $W_{n,a,i}$. The variables z_j have degree $(2-\lambda)/a$ for $j \le i$ and degree $2/a$ for $j > i$, and q has degree λ.

Now let us drop the n, a, i from the notation. By taking M to be a Fermat hypersurface, we obtain a natural action of the dual group G^* on M, respecting D. Restricting to the invariant pieces of the relevant group actions, mirror symmetry predicts that

$$SH^*(X;\Bbbk)^{G^*} \cong H^*\left(\Bbbk[z_1,\ldots,z_{n+2},\partial/\partial z_1,\ldots,\partial/\partial z_{n+2}],[W,-]\right)^G,$$

and in fact that there is an underlying quasi-isomorphism of L_∞ algebras. In accordance with Conjecture 1.25, this gives us

$$H^*\left(\widetilde{SC}_\Lambda,\partial\right)^{G^*} \cong H^*\left(\Lambda[z_1,\ldots,z_{n+2},\partial/\partial z_1,\ldots,\partial/\partial z_{n+2}],[W+\beta,-]\right)^G,$$

and hence

$$QH^*(M;\Lambda)^{G^*} \cong Jac(W+\beta)^G.$$

The Jacobian ring has relations

$$\begin{aligned} \frac{z_1\ldots z_{n+2}}{z_j} &= qz_j^{a-1} \qquad \text{for } j\le i\\ \frac{z_1\ldots z_{n+2}}{z_j} &= z_j^{a-1} \qquad \text{for } j> i. \end{aligned}$$

Multiplying them together we get that

$$(z_1\ldots z_{n+2})^{n+1} = q^i(z_1\ldots z_{n+2})^{a-1}.$$

This allows us to compute that

$$Jac(W+\beta)^G = \Lambda[H]/\left(H^{n+1}-q^iH^{a-1}\right),$$

where $H = z_1\ldots z_{n+2}$.

The class H corresponds to the hyperplane class (except for the case $n+2-a=1$, when it corresponds to the hyperplane class plus $a!q^i$). One can check that this is the correct answer for $QH^*(M;\Lambda)^{G^*}$, see [17,20]. This is in agreement with Theorem B.

We can also check Conjecture 1.20 in this case. We can factor the defining relation in the Jacobian ring as:

$$H^{n+1}-q^iH^{a-1} = H^{a-1}\prod_{\zeta^{n+2-a}=1}\left(H-\zeta q^{\frac{i}{n+2-a}}\right).$$

Note that we have $H = z_1\ldots z_{n+2} = qz_1^a$, from the first relation in the Jacobian ring. Thus $\mathcal{Q}(H)=1$. On the other hand, $\mathcal{Q}(q^{i/(n+2-a)}) = i/(n+2-a)$. Therefore, precisely when Hypothesis A is not satisfied, the factors $(H-\zeta q^{i/(n+2-a)})$ become invertible in the $\mathcal{Q}$-completion, as argued in Sect. 1.6.2.

The result is that the $\mathcal{Q}$-completion gives $\Lambda[H]/H^{a-1}$, which corresponds to the zero generalized eigenspace (note that Hypothesis A is satisfied for all $i\ge 1$ in the anomalous case $n+2-a=1$, when this corresponds to the $-a!q^i$ generalized eigenspace.)

1.8. Outline

In Sect. 2, we examine the structure of our symplectic manifold in a neighbourhood of the divisor D. In particular, we introduce the notion of a 'system of commuting Hamiltonians near D', and say what it means for a Liouville one-form to be 'adapted' to such a system. This completes the statement of the results in Sect. 1.3, where these notions were used without being defined.

In Sect. 3, we establish our conventions for Hamiltonian Floer theory and relative symplectic cohomology in M, and explain how they are related to symplectic cohomology of the exact symplectic manifold X. In particular, we establish that the map (1.2) respects index and action; and we prove the 'positivity of intersection'-type result which is used to prove Proposition 1.12.

In Sect. 4, we construct the functions ρ^R which are smoothings of ρ^0. We consider degenerate Hamiltonians of the form $h \circ \rho^R$, explain how to perturb them to obtain non-degenerate time-dependent Hamiltonians H, and give estimates for the index and action of their orbits.

In Sect. 5, we prove our main results.

History of work This paper started with M.S.B. and N.S. trying to prove Conjecture 1.25. A solution was announced in 2015, but never appeared. The project languished, until U.V. joined the collaboration in 2019 and pushed it to completion in its current form. M.S.B. and N.S. apologize for the long delay between announcement and appearance of the work.

2. Symplectic divisors

2.1. Basics

We recall some notions from [36, Section 2.1]. Let (M,ω) be a $2n$-dimensional closed symplectic manifold and let $D = \cup_{i=1}^N D_i$ be a symplectic divisor in (M,ω). This means that for each i, $D_i \subset M$ is a connected smooth closed submanifold with real codimension two and for each subset $I \subset [N]$ the intersection $\bigcap_{i\in I} D_i$ is transverse and

$$D_I := \bigcap_{i\in I} D_i \subset M$$

is a symplectic submanifold. Since the D_i intersect transversally, for each $I \subset [N]$ there is an isomorphism of vector bundles

$$N_M D_I \xrightarrow{\sim} \bigoplus_{i\in I} N_M D_i|_{D_I} \tag{2.1}$$

over D_I, induced by the inclusions $TD_I \subset TD_i|_{D_I}$. Recall the normal bundle $N_M D$ for any symplectic submanifold $D \subset (M,\omega)$ has a *symplectic orientation* induced by the symplectic orientations of TD and TM.

Definition 2.1. A symplectic divisor $D \subset (M,\omega)$ is

(i) a *simple crossings (SC) divisor* if (2.1) is an isomorphism of oriented vector bundles, where each normal bundle is given its symplectic orientation, for all I.

(ii) *orthogonal* if for all $i \neq j$ and $x \in D_i \cap D_j$ the ω-normal bundle $(T_xD_i)^\omega \subset T_xM$ is contained in T_xD_j.

Remark 2.2. In [23, Section 5] McLean proved that any SC divisor $D \subset (M, \omega)$ can be smoothly isotoped in the space of SC divisors to an orthogonal SC divisor $D' \subset (M, \omega)$; and that $X' = M \backslash D'$ is convex deformation equivalent to $X = M \backslash D$. This implies that $SH^*(X; \Bbbk) \cong SH^*(X'; \Bbbk)$, by [23, Lemma 4.11]. These results mean that it suffices to prove Theorems B and C under the assumption that D is orthogonal.

Setting $X = M\backslash D$, by Lefschetz duality (e.g. Proposition 7.2 of [7]) we have

$$H_2(M, X) \cong \mathbb{Z}^N \quad \text{where} \quad A \mapsto (A \cdot D_i)_{i=1}^N. \tag{2.2}$$

The inverse is given by mapping the ith basis vector to a disk $u_i : (\mathbb{D}, \partial\mathbb{D}) \to (M, X)$ that is disjoint from the other D_j and with intersection number $u_i \cdot D_i = 1$. The dual basis vectors of $H^2(M, X) \cong \mathbb{Z}^N$ are what we called $\mathrm{PD}^{rel}(D_i)$ in Sect. 1.1.

Assume that

$$\boldsymbol{\kappa} = \sum_i \kappa_i \mathrm{PD}^{rel}[D_i] \in H^2(M, X; \mathbb{R})$$

is a lift of $[\omega]$ under the map $H^2(M, X; \mathbb{R}) \to H^2(M; \mathbb{R})$ with $\kappa_i \in \mathbb{R}$.

Remark 2.3. In the setup from Sect. 1.1, $\boldsymbol{\kappa}$ will be $\kappa\boldsymbol{\lambda}$.

Now consider a de Rham representative (ω, θ) for $\boldsymbol{\kappa}$ consisting of the symplectic form ω together with a one-form $\theta \in \Omega^1(X)$ satisfying $\mathrm{d}\theta = \omega|_X$, and

$$\kappa_i = \int_{u_i} \omega - \int_{\partial u_i} \theta.$$

Following McLean [23,24] we call κ_i the *wrapping numbers for D with respect to θ*, though we use the opposite sign convention than in [23].

2.2. Systems of commuting Hamiltonians

Definition 2.4. Let $D = \cup_i D_i$ be an SC divisor in a closed symplectic manifold (M, ω), and $R > 0$. A *system of commuting Hamiltonians (scH) near D, of radius R*, is a collection of open neighborhoods $UD_i \supset D_i$ and proper smooth functions $r_i : UD_i \to [0, R)$, for each i, with the following properties. For each i,

- r_i generates an $\mathbb{R}/\mathbb{Z}$ action on UD_i, and $r_i^{-1}(0) = D_i$.
- The fixed point set of the $\mathbb{R}/\mathbb{Z}$ action on UD_i is D_i.
- The $\mathbb{R}/\mathbb{Z}$ action on $UD_i \backslash D_i$ is free.

For all pairs i, j,

- $UD_i \cap UD_j$ is invariant under the $\mathbb{R}/\mathbb{Z}$ action generated by r_i.
- The Hamiltonians r_i and r_j Poisson commute on $UD_i \cap UD_j$.

We will denote a scH near D of radius R with the notation $\{r_i : UD_i \to [0, R)\}$.

Note that for any scH of radius R, we can 'shrink' it to an scH of radius $R' < R$ by replacing UD_i with $\{r_i < R'\}$ for each i.

Proposition 2.5. *Let D be an SC divisor in a closed symplectic manifold (M, ω). If D admits a scH, then it is orthogonal.*

Proof. Assume that $\{r_i : UD_i \to [0, R)\}$ is a scH near D. We need to show that for all $i \neq j$ and $x \in D_i \cap D_j$ the symplectic orthogonal $(T_x D_i)^\omega \subset T_x M$ is contained in $T_x D_j$.

We consider the action of $S := \mathbb{R}/\mathbb{Z}$ on $T_x M$ induced by r_i. The action on $T_x D_i \subset T_x M$ is trivial, since D_i is fixed pointwise under the action of S. The action on $UD_i \cap UD_j$ leaves $\{r_j = 0\} \cap UD_i \cap UD_j$ invariant by the Poisson commutativity property. Therefore, $T_x D_j$ is an invariant subspace of $T_x M$ under the S action. Finally, since the action of S on $T_x M$ preserves the symplectic pairing, $(T_x D_i)^\omega \subset T_x M$ is also an invariant subspace.

Note that the action of S on $(T_x D_i)^\omega$ cannot be trivial by the Bochner linearization theorem, as x does not have a neighborhood on which S acts trivially. Now we finish the proof with the following claim:

- Assume that V is a finite dimensional symplectic representation of S, which is the direct sum of two representations $W \oplus E$, where W is the trivial representation on a symplectic codimension 2 subspace, E is not the trivial representation, and E and W are symplectically orthogonal. Let W' be another codimension 2 symplectic subspace of V which is invariant under the action of S. Then if W' is transverse to W, it has to contain E.

The proof of this statement is as follows. There exists $w + e \in W'$ with $e \neq 0$, as W' is transverse to W. For any $\theta \in S$ we have $\theta \cdot (w + e) \in W'$; hence, $\theta \cdot (w + e) - (w + e) = \theta \cdot e - e \in W'$. We may choose θ so that $\theta \cdot e \neq e$, so $W' \cap E \neq \{0\}$. This implies that $E \subset W'$ as required. □

Definition 2.6. Let $D = \cup_{i=1}^N D_i$ be an SC divisor in a closed symplectic manifold (M, ω) and let $\{r_i : UD_i \to [0, R)\}$ be a scH near D. For all $I \subset [N]$, define $UD_I := \cap_{i \in I} UD_i$. We obtain a $(\mathbb{R}/\mathbb{Z})^I$ action on UD_I with a moment map

$$r_I : UD_I \to [0, R)^I$$

whose components are given by r_i, for $i \in I$.

Proposition 2.7. *Let D be an orthogonal SC divisor in a closed symplectic manifold (M, ω). Then D admits a scH.*

Proof. This is an immediate consequence of [23, Lemma 5.14], where for each i, we use the well-defined radial coordinate of the symplectic disk bundle over D_i in the statement as our r_i (the domain is the symplectic disk bundle of course). It is trivial to see that this gives a scH near D. □

Remark 2.8. It is natural to ask whether all systems of commuting Hamiltonians come from standard tubular neighborhoods in the sense of McLean. Even if this is the case, the extra choice of a standard tubular neighborhood on top of a system of commuting Hamiltonians is not needed for our constructions and arguments.

2.3. Adapted Liouville one-forms

Definition 2.9. Let D be an SC divisor in a closed symplectic manifold (M,ω) and let $\{r_i : UD_i \to [0,R)\}$ be an admissible scH near D. We call a one-form $\theta \in \Omega^1(M\backslash D)$ satisfying $\mathrm{d}\theta = \omega|_{M\backslash D}$ and with wrapping numbers $\kappa_i > 0$ *adapted* to $\{r_i : UD_i \to [0,R)\}$ if the Liouville vector field Z of θ satisfies

$$Z(r_i) = r_i - \kappa_i$$

over $UD_i\backslash D$, for all i.

Proposition 2.10. *Let D be an orthogonal SC divisor in a closed symplectic manifold (M,ω). Assume that*

$$[\omega] = \sum_i \kappa_i \cdot \mathrm{PD}(D_i) \qquad \textit{in } H_2(M,\mathbb{R}),$$

with $\kappa_i > 0$. Then there exists $\{r_i : UD_i \to [0,R)\}$ a scH near D for which there exists an adapted $\theta \in \Omega^1(M\backslash D)$ with wrapping numbers κ_i.

Proof. We use a scH as in the proof of Proposition 2.7. Then, a one-form θ on $M\backslash D$ produced by [23, Lemma 5.17] is adapted in the sense of Definition 2.9, as we show below. Note that by the relative de Rham isomorphism, there is a primitive θ' defined on $M\backslash D$ such that the relative cohomology class in $H^2(M, M\backslash D)$ defined by (ω,θ') is equal to $\sum \kappa_i \cdot \mathrm{PD}(D_i)$, which is why we can use McLean's lemma.

Using McLean's notation for the moment, on the fibers of the projections $\pi_I : UD_I \to D_I$ we have

$$\theta|_{F_I^*} = \sum_{i\in I} (r_i - \kappa_i)\,\mathrm{d}\phi_i, \tag{2.3}$$

where $F_I^* \cong \prod_{i\in I}(\mathbb{D}_R\backslash 0)$ is the product of punctured disks. Using (2.3), we have

$$Z(r_i) = \theta(X_{r_i}) = \theta(\partial_{\phi_i}) = r_i - \kappa_i,$$

as required. □

Remark 2.11. Again one could ask whether every Liouville one-form adapted to a system of commuting Hamiltonians is adapted to some compatible standard tubular neighborhood in the sense of McLean. Whatever the answer might be, the flexibility that we achieved in these two sections already shows itself in the toric examples of Sect. 1.7.1.

2.4. Admissibility

Definition 2.12. Let $D = \cup_{i=1}^N D_i$ be an SC divisor in a closed symplectic manifold (M,ω), and let $\{r_i : UD_i \to [0,R)\}$ be a scH near D. Given $I \subset [N]$, a *standard chart* (U,ϕ) in UD_I is an $(\mathbb{R}/\mathbb{Z})^I$-invariant open subset $U \subset UD_I$ and a $(\mathbb{R}/\mathbb{Z})^I$-equivariant symplectic embedding

$$\phi : U \to \mathbb{C}^I \times \mathbb{C}^{n-|I|},$$

where we use the action of $(\mathbb{R}/\mathbb{Z})^I$ on $\mathbb{C}^I \times \mathbb{C}^{n-|I|}$ given by

$$\theta\cdot((z_i)_{i\in I}, w) = ((e^{2\pi i\theta_i} z_i)_{i\in I}, w) \text{ for all } \theta \in (\mathbb{R}/\mathbb{Z})^I$$
$$\text{and } ((z_i)_{i\in I}, w) \in \mathbb{C}^I \times \mathbb{C}^{n-|I|}.$$

Lemma 2.13. *Let $D = \cup_{i=1}^N D_i$ be an SC divisor in a closed symplectic manifold (M, ω) and let $\{r_i : UD_i \to [0, R)\}$ be a scH near D. For every $I \subset [N]$ and $x \in D_I$, there exists a standard chart (U, ϕ) in UD_I containing x.*

Proof. This immediately follows from the equivariant Darboux theorem [18, Theorem 22.1]. □

We now choose an arbitrary Riemannian metric on M, and let $\mathrm{inj}(M)$ be the injectivity radius with respect to this metric. We call a standard chart (U, ϕ) in UD_I *admissible* if U is contractible and has diameter $< \mathrm{inj}(M)/2$. The significance of admissibility for us is that it guarantees uniqueness of caps:

Lemma 2.14. *If $\gamma : S^1 \to M$ is a loop contained in some admissible standard chart, then there exists a disc bounding γ, whose image is contained inside an admissible chart. Moreover, such a disc is independent of the choice of admissible chart containing γ, up to homotopy rel. boundary in M.*

Proof. The existence is clear, as admissible standard charts are contractible. The uniqueness follows as the union of two admissible standard charts containing γ has diameter $< \mathrm{inj}(M)$, hence is contained in a ball of radius $< \mathrm{inj}(M)$. As the ball is contractible, the caps in the two charts are homotopic rel. boundary in M. □

Definition 2.15. Let $D = \cup_{i=1}^N D_i$ be an SC divisor in a closed symplectic manifold (M, ω). We call a scH near D *admissible* if for every $I \subset [N]$ and $y \in UD_I$, there exists an admissible standard chart (U, ϕ) in UD_I with $y \in U$.

Lemma 2.16. *Let D be an SC divisor in a closed symplectic manifold (M, ω), and $\{r_i : UD_i \to [0, R)\}$ a scH near D. Then any sufficiently small shrinking of the scH is admissible.*

Proof. First note that any standard chart around $x \in D_I$ can be shrunk so that it is admissible. Therefore, we have a neighbourhood of D_I given by the union of all admissible standard charts. By shrinking the scH sufficiently, we may ensure that UD_I is contained in the neighbourhood, for all I. □

Remark 2.17. In Sect. 3.1, we will define a cap for a loop $\gamma : S^1 \to M$ to be an equivalence class of discs u bounding γ under the equivalence relation $u_1 \sim u_2$ if $\int u_1^* \omega = \int u_2^* \omega$. Therefore, we could get away with the following weaker notion of admissibility for the purposes of the present paper. We call a standard chart weakly admissible if it is simply connected. Assume that we have a loop γ inside UD_I that is the orbit of a point under the action of a one dimensional subgroup S of $(\mathbb{R}/\mathbb{Z})^I$. We claim that the symplectic area of a cap of γ that is contained inside a weakly admissible standard chart U (assuming such charts exist) only depends on γ, i.e. it is independent of U and the cap chosen inside of U. The reason is because we can then compute the symplectic area by transporting everything into $\mathbb{C}^I \times \mathbb{C}^{n-|I|}$ and see that it is equal to $l(0) - l(p)$, where $l : \mathbb{R}^I \to \mathbb{R}$ is a function whose pre-composition with r_I generates the action of S and p is the point of $\mathbb{R}^I_{\geq 0}$ above which γ lives. Hence, for such γ existence of a weakly admissible standard chart determines uniquely an equivalence class of caps. This would be enough for our purposes.

3. Quantum, Hamiltonian Floer, and symplectic cohomology

3.1. Quantum and Hamiltonian Floer cohomology

In this section, (M,ω) will be a closed symplectic manifold such that $2\kappa c_1(TM) = [\omega]$ on $\pi_2(M)$ for some $\kappa > 0$.

Let A' be the subgroup $\{2c_1(TM)(B) : B \in \pi_2(M)\} \subset \mathbb{Z}$ and set $\Lambda' = \Bbbk[A']$, graded by $i(e^a) = a$.

Let $\gamma : S^1 \to M$ be a nullhomotopic loop in M. A *cap* for γ is an equivalence class of disks $u : \mathbb{D} \to M$ bounding γ, where $u \sim u'$ if and only if the Chern number of the spherical class $[u-u']$ vanishes: $c_1(TM)(u-u') = 0$. The set of caps for γ is a torsor for A', which acts via

$$a \cdot (\gamma, u) = (\gamma, u\#C) \qquad \text{where} \qquad 2c_1(TM)(C) = a.$$

Given a non-degenerate Hamiltonian $F : S^1 \times M \to \mathbb{R}$, let $\mathcal{P}_F$ denote the set of contractible one-periodic orbits of F, and let $\tilde{\mathcal{P}}_F$ be the set of orbits equipped with a cap. Elements $\tilde{\gamma} = (\gamma, u) \in \tilde{\mathcal{P}}_F$ have a $\mathbb{Z}$-grading and an action

$$i(\gamma,u) = \mathrm{CZ}(\gamma,u) + \frac{\dim(M)}{2} \quad \text{and} \quad \mathcal{A}_F(\gamma,u) := \int_{S^1} F(t,\gamma(t))\,\mathrm{d}t + \int_{\mathbb{D}} u^*\omega,$$

and these are compatible with the action of A' in that

$$i(a\cdot(\gamma,u)) = i(\gamma,u) + a \qquad \text{and} \qquad \mathcal{A}(a\cdot(\gamma,u)) = \mathcal{A}(\gamma,u) + \kappa a\,.$$

Note that the 'mixed index'

$$i_{\text{mix}}(\gamma) := i(\gamma,u) - \kappa^{-1}\mathcal{A}(\gamma,u)$$

is independent of the cap u.

Define $CF^*(M,F)$ to be the free $\mathbb{Z}$-graded $\Bbbk$-module generated by $\tilde{\mathcal{P}}_F$. It is naturally a graded Λ'-module, via $e^a \cdot (\gamma,u) := a\cdot(\gamma,u)$. It also admits a Floer differential after the choice of a generic S^1-family of ω-compatible almost complex structures (which we suppress from the notation). The differential is Λ'-linear, increases the grading by 1, does not decrease action, and squares to zero.

One can also define continuation maps $CF(M,F_0) \to CF(M,F_1)$ in the standard way by choosing a smooth function $\mathcal{F} : \mathbb{R}_s \times S^1 \times M \to \mathbb{R}$, which is equal to F_0 for $s \ll 0$ and to F_1 for $s \gg 0$, as well as an $\mathbb{R}\times S^1$ dependent family of ω-compatible almost complex structures, which together satisfy a regularity condition. Continuation maps are Λ'-linear chain maps. If the continuation maps are defined using monotone Floer data, which means $\frac{\partial \mathcal{F}}{\partial s} \geq 0$, then the continuation map $CF(M,F_0) \to CF(M,F_1)$ does not decrease action.

Remark 3.1. We would like to stress that the discussion of Hamiltonian Floer theory that we gave here is slightly simpler than the general theory due to our positive monotonicity assumption. In particular, we did not need to complete

Λ or our Hamiltonian Floer groups, which is necessary in general for the potential infinite sums to make sense. For details, we refer the reader to [31]. Apart from the ones that we have explicitly stated above, our conventions for Hamiltonian Floer theory agree with (1), (2), (3) and (5) of Section 3.1 in [40].

Let $A \subset \mathbb{Q}$ be a subgroup such that $A' \subset A$. Let $\Lambda = \Bbbk[A]$, with the same grading convention $i(e^a) = a \in \mathbb{Q}$; then we have an inclusion $\Lambda' \subset \Lambda$. (Eventually we will take A and Λ to be as defined in the beginning of Sect. 1.2 but we choose to be more general for a while.)

Let us define the Λ-cochain complex

$$CF^*(M, F; \Lambda) := CF^*(M, F) \otimes_{\Lambda'} \Lambda.$$

We denote the cohomology of this cochain complex by $HF^*(M, F; \Lambda) := H^*(CF^*(M, F; \Lambda), \partial)$. There exists a natural PSS chain map:

$$C^*(M; \Bbbk) \otimes_{\Bbbk} \Lambda \to CF^*(M, F; \Lambda),$$

which is known to be a quasi-isomorphism [28]. The PSS map is well defined up to chain homotopy and compatible with chain level continuation maps up to chain homotopy.

We now introduce the notion of 'fractional caps' of orbits. A fractional cap for γ is a formal expression $u + a$, where u is a cap for γ and $a \in \mathbb{R}$, and we declare $u + a \sim u' + a'$ if and only if $a - a' \in A'$ and $u' = (a - a') \cdot u$. There is a well-defined index and action associated with a fractional cap:

$$i(\gamma, u + a) := i(\gamma, u) + a, \qquad \mathcal{A}(\gamma, u + a) := \mathcal{A}(\gamma, u) + \kappa a.$$

There is a natural bijection between the $\Bbbk$-basis $(\gamma, u) \otimes e^a$ of $CF^*(M, F; \Lambda)$, and the set of fractionally capped orbits $(\gamma, u + a)$ with $a \in A$.

3.2. Relative symplectic cohomology

Let $M, \omega, \kappa, \Lambda$ be as in Sect. 3.1. We now define relative symplectic cohomology for compact subsets of M over Λ, referring to [40] for the details. As briefly mentioned in the introduction (see Sect. 1.3, especially the footnote on pages 4-5), the construction below is slightly different than the one in [40]. Namely, here we use capped orbits (in particular we only consider contractible orbits) and keep track of the caps rather than weighting Floer solutions using a formal variable.

Let $K \subset M$ be compact. We call the following data a choice of acceleration data for K:

- $H_1 \leq H_2 \leq \cdots$ a monotone sequence of non-degenerate one-periodic Hamiltonians $H_i : S^1 \times M \to \mathbb{R}$ cofinal among functions satisfying $H\mid_{S^1\times K} < 0$. In other words, for every $(t, x) \in S^1 \times M$,

$$H_i(t, x) \xrightarrow[i \to +\infty]{} \begin{cases} 0, & x \in K, \\ +\infty, & x \notin K. \end{cases}$$

- A monotone homotopy of Hamiltonians $H_{i,i+1} : [i, i+1] \times S^1 \times M \to \mathbb{R}$, for all i, which is equal to H_i and H_{i+1} at the corresponding end points.
- A $\mathbb{R}_{\geq 1} \times S^1$-family of ω-compatible almost complex structures.

We denote the acceleration data as a single family of time-dependent Hamiltonians and almost complex structures (H_τ, J_τ), $\tau \in \mathbb{R}_{\geq 1}$. We also fix an non-decreasing surjective smooth map $(-\infty, \infty) \to [0, 1]$. Given a $[i, i+1]$-dependent family of Hamiltonians and almost complex structures, we use this map to write down a Floer equation for maps from $\mathbb{R} \times S^1$ to M. Let us call the resulting $\mathbb{R} \times S^1$-family of Hamiltonians and almost complex structures the associated Floer data.

We require the acceleration data (H_τ, J_τ) to satisfy the following two assumptions:

(1) For each $i \in \mathbb{N}$, (H_i, J_i) is regular.
(2) For each $i \in \mathbb{N}$, the Floer data associated with $(H_\tau, J_\tau)_{\tau \in [i,i+1]}$ is regular.

Given acceleration data (H_τ, J_τ), Hamiltonian Floer theory provides a 1-ray of Floer Λ-cochain complexes, called a *Floer 1-ray*:

$$\mathcal{C}(H_\tau, J_\tau) := CF^*(M, H_1; \Lambda) \to CF^*(M, H_2; \Lambda) \to \cdots .$$

The horizontal arrows are Floer continuation maps defined using the monotone homotopies appearing in the acceleration data. Recall that a cylinder u contributing to a Floer differential or a continuation map has non-negative topological energy

$$E_{\text{top}}(u) = \int_{S^1} \gamma_{\text{out}}^* H_{\text{out}} \, \mathrm{d}t - \int_{S^1} \gamma_{\text{in}}^* H_{\text{in}} \, \mathrm{d}t + \int_{\mathbb{R}\times S^1} u^*\omega \geq 0, \tag{3.1}$$

where γ_{out}, γ_{in} are the asymptotic orbits of u, and H_{out}, H_{in} are the Hamiltonians at the corresponding ends. (For Floer differentials, $H_{\text{out}} = H_{\text{in}} = H_i$ and for continuation maps, $H_{\text{out}} = H_{i+1}$, $H_{\text{in}} = H_i$ for some i.)

Remark 3.2. We also note that the inequality in (3.1) comes from the more general inequality

$$E_{\text{top}}(u) \geq \int_{\mathbb{R}\times S^1} \left(\frac{\partial H}{\partial s}\right) (u(s,t), s, t) \mathrm{d}s\mathrm{d}t, \tag{3.2}$$

where u is a solution of the Floer equation for an arbitrary $H : \mathbb{R} \times S^1 \times M \to \mathbb{R}$ which is s-independent at the ends.

From now on, we will use the terminology introduced in Sect. A.3 freely. We apologetically ask the reader to take a look at it before moving further. Using the grading and action considerations from Sect. 3.1, $\mathcal{C}(H_\tau, J_\tau)$ becomes a 1-ray in $FiltCh_\Lambda$. We define the Λ-cochain complexes $tel(\mathcal{C}(H_\tau, J_\tau))$ and $\widehat{tel}(\mathcal{C}(H_\tau, J_\tau))$ as in Sect. A.3. We can now repeat Section 3.3.2 of [40] in this set-up.

Proposition 3.3. *For two different choices of acceleration data for K, (H_τ, J_τ) and (H'_τ, J'_τ), there is a canonical isomorphism*

$$H^*\left(\widehat{tel}(\mathcal{C}(H_\tau, J_\tau))\right) \cong H^*\left(\widehat{tel}(\mathcal{C}(H'_\tau, J'_\tau))\right)$$

of $\mathbb{Q}$-graded Λ-modules.

Hence, we define

$$SH^*_M(K;\Lambda) := H^*\left(\widehat{tel}(\mathcal{C}(H_s,J))\right).$$

Proposition 3.4. *There are canonical restriction maps of $\mathbb{Q}$-graded Λ-modules for $K \subset K'$:*

$$SH^*_M(K';\Lambda) \to SH^*_M(K;\Lambda).$$

□

We finally list the three properties we will need of relative symplectic cohomology. Here is the first one.

Theorem 3.5. *Assume that $tel(\mathcal{C}(H_\tau, J_\tau))$ is degreewise complete. Then $SH^*_M(K;\Lambda) = QH^*(M;\Lambda)$.*

Proof. Follows from the basic properties of the PSS maps discussed at the end of Sect. 3.1 along with the diagram (A.2) and the fact that a direct limit of quasi-isomorphisms is a quasi-isomorphism. □

Before we state the second property, we note the following important statement from Hamiltonian Floer theory.

Let $H : S^1 \times M \to \mathbb{R}$ a non-degenerate Hamiltonian and J an S^1-dependent almost complex structure compatible with ω. Assume that (H, J) is regular and fix $\Delta \geq 0$.

- The Floer data $(H_s := H + \Psi(s)\Delta, J_s := J)$, where $\Psi : \mathbb{R} \to \mathbb{R}$ is a smooth function that is equal to 0 for $s < -1$ and to 1 for $s > 1$, is regular. This is a standard fact in Floer theory noting that adding $\Psi(s)\Delta$ does not change the Floer equation.
- The resulting continuation map

  $$c_\Psi : CF^*(M,H) \to CF^*(M,H+\Delta)$$

  is the naive map which sends each capped orbit to itself. Yet, note that the action of the capped orbit for $H + \Delta$ is Δ more than its action for H.

Let us fix a non-decreasing Ψ for the proof below. Let us denote the continuation map above for any H and Δ by c_Ψ by abuse of notation.

Theorem 3.6. *If K is stably displaceable, then $SH^*_M(K;\Lambda) = 0$.*

Sketch of proof. The proof is identical to that in the Section 4.2 of [39] up to minor modifications. We provide an overview of the proof for completeness.

Let us first prove the result when K is displaceable. Let (H_τ, J_τ) be a choice of acceleration data for K and $H : [0,1] \times M \to \mathbb{R}$ be a function whose time-1 Hamiltonian flow $\phi : M \to M$ displaces K. In fact, ϕ displaces a domain neighborhood D of K. Assume that H_τ's are so that ∂D is a level set of H_1 for all $t \in S^1$, and $H_\tau = H_1 + \tau - 1$ on $M - int(D)$ for all τ.

We recall an elementary construction for reparametrizing Hamiltonian flows . Let $I = [0, T]$ and $I' = [0, T']$ be closed intervals, and $\psi : I' \to I$ be a smooth map which sends 0 to 0 and T' to T. Then, the time T-flow of the time dependent Hamiltonian vector field X_t, $t \in I$ of $h : I \times M \to \mathbb{R}$ and the time-T' flow of X'_t, $t \in I'$ of $(h \circ (\psi \times id)) \cdot \frac{d\psi}{dt} : I' \times M \to \mathbb{R}$ are the same map $M \to M$.[11]

Let us fix a non-decreasing function $\psi : [0, 1/2] \to [0, 1]$, which is locally constant in a neighborhood of the endpoints of $[0, 1/2]$.

Using the reparametrization construction with ψ, starting with $H_L, H_R : M \times [0,1] \to \mathbb{R}$ we can cook up a new Hamiltonian $H_L \phi H_R : M \times \mathbb{R}/\mathbb{Z} \to \mathbb{R}$, such that the H_L and H_R parts are supported in $(1/2, 1)$ and $(0, 1/2)$ respectively. The Hamiltonian flow of $H_L \phi H_R$ is tangent to X_{H_R} first. After not moving for a short period, it arrives at $\phi^1_{H_R}$ in less than 1/2-time, and stops for a while. At some point after time 1/2, it starts moving again, this time being tangent to X_{H_L}, and reaches to $\phi^1_{H_L} \circ \phi^1_{H_R}$ before time-1. It then stops for a little until time 1, after which it repeats this flow.

We define $SH^*_M(K, H; \Lambda)$ via the family $H \phi H_s$ in the same way we defined $SH^*_M(K; \Lambda)$. Note that this construction does not use that H displaces K. In particular, we can define $SH^*_M(K, 0; \Lambda)$, and it follows from Lemma 4.2.1 of [39] that $SH^*_M(K, 0; \Lambda)$ is isomorphic (as a graded $\Lambda-$module) to $SH^*_M(K; \Lambda)$. Here and in the future, by abuse of notation, we denote the constant function $M \times [0,1] \to \mathbb{R}$, sending everything to $\Delta \in \mathbb{R}$ by Δ.

The next step is to show that $SH^*_M(K, H; \Lambda)$ is isomorphic to SH^*_M $(K, 0; \Lambda)$, which is true for arbitrary H. We can find a $\Delta \geq 0$ such that

$$-\Delta \leq H(x,t) \leq \Delta,$$

for all $(x,t) \in M \times [0,1]$. This implies that for any $G : M \times [0,1] \to \mathbb{R}$, we have

$$-c\Delta + 0\phi G \leq H\phi G \leq c\Delta + 0\phi G \leq 2c\Delta + H\phi G,$$

where $c > 0$ is a constant that depends on our choice of ψ.

Hence we obtain filtered chain maps

$$\begin{aligned} &tel(\mathcal{C}(-c\Delta + 0\phi H_s)) \to tel(\mathcal{C}(H\phi H_s)) \to tel(\mathcal{C}(c\Delta + 0\phi H_s)) \\ &\to tel(\mathcal{C}(2c\Delta + H\phi H_s)). \end{aligned}$$

The composition of the first two maps is filtered chain homotopic to the map obtained from $c'_\Psi s$ as explained right before the theorem using a filling in 3-slits argument. The same result is true for the composition of last two maps.

Using Lemma A.2's last statement and the second bullet point of Lemma A.3, we obtain that there is a chain of maps

$$SH^*_M(K, 0; \Lambda) \to SH^*_M(K, H; \Lambda) \to SH^*_M(K, 0; \Lambda) \to SH^*_M(K, H; \Lambda),$$

where the composition of the first two and the last two maps are isomorphisms. This implies the result.

The main point of the proof is to show that $SH^*_M(K, H; \Lambda) = 0$ for the displacing Hamiltonian H from the beginning of the argument. This uses Lemma A.5. The more detailed claim is that a slightly modified version of the family $H\phi H_s$ gives rise to a 1-ray that satisfies the conditions of Lemma

[11] We warn the reader that there is a typo in the relevant formula in [39].

A.5. The actual proof of this is too long to include here (see Section 4.2.3 of [39]). Let us instead explain the intuition behind the proof. Let $\gamma : S^1 \to M$ be a 1-periodic orbit of $H\phi H_s$ for some s. Because ϕ displaces D, either $\gamma(0)$ or $\gamma(1/2) = \phi^{-1}(\gamma(0))$ needs to lie outside of D. Then conservation of energy and that ∂D is a level set of H_s for all times shows that in fact we have $\gamma([0, 1/2]) \subset M\backslash D$. Now if we could use parametrized moduli spaces and cascades instead of continuation maps, we would have our proof. This relies in the fact that $\gamma([0, 1/2]) \subset M\backslash D$ holds for all 1-periodic orbits of all $H\phi H_s$ and that $\frac{\partial H_s}{\partial s} = 1$ in $M\backslash D$: the actions increase with a constant rate as we follow the orbits and accidental solutions can only further increase the action. There are technical difficulties in making this work, so we refer the reader to [39] for the actual proof.

We move on to the case when K is only stably displaceable. Let T^2 be a symplectic torus such that

$$\tilde{K} := K \times \gamma \subset M \times T^2$$

is displaceable inside $M \times T^2$, where γ is a meridian in T^2. Note that $M \times T^2$ also satisfies the conditions of our construction of relative symplectic cohomology over Λ as T^2 is aspherical.

We will prove that $SH^*_M(K;\Lambda)$ naturally injects into $SH^*_{M\times T^2}(\tilde{K};\Lambda)$, which finishes the proof. It is easy to see that acceleration data can be chosen for $\gamma \subset T^2$ where each Hamiltonian in the cofinal family has exactly 4 contractible orbits, and the differentials on each of the corresponding Hamiltonian Floer groups vanish. Using the the chain level Künneth isomorphism for Hamiltonian Floer theory and that completion commutes with tensor product with a finite dimensional Λ-module, we easily prove the desired claim. $\square$

We come to the third and final property of relative symplectic cohomology that we will discuss in this section. Recall from the introduction that a compact set $K \subset M$ is called *SH-invisible* if $SH^*_M(K;\Lambda) = 0$.

Theorem 3.7. *If a compact subset $K \subset M$ is SH-invisible, then any compact subset $K' \subset K$ is also SH-invisible.*

Proof. The proof is identical to that of Theorem 1.2 (4) in [38]. The key point (Proposition 2.5 of [38]) is that there is a distinguished element $1_K \in SH_M(K, \Lambda)$, called the unit, with the following properties.

- $SH_M(K, \Lambda) = 0$ if and only if $1_K = 0$.
- Restriction maps send units to units.

The element 1_K is constructed so that it is the unit of a pair-of-pants type product structure on $SH_M(K, \Lambda)$. The details are in Section 5 of [38]. $\square$

3.3. Towards the symplectic cohomology of the divisor complement

We return to the geometric setup of Sect. 1.1: (M, ω) will be a closed symplectic manifold that is monotone

$$2\kappa c_1^M = [\omega] \in H^2(M;\mathbb{R}) \quad \text{with} \quad \kappa > 0,$$

$D = \cup_{i=1}^N D_i \subset (M, \omega)$ will be a simple crossings divisor and $\lambda_1, \ldots, \lambda_n \in \mathbb{Q}_{>0}$ will be the weights. We will denote $X = M\backslash D$, $\boldsymbol{\lambda} \in H^2(M, X;\mathbb{R})$ will be the associated lift of $2c_1^M$, and $\theta \in \Omega^1(X)$ will be a primitive of $\omega|_X$ such that the relative de Rham cohomology class of (ω, θ) is $\kappa\boldsymbol{\lambda}$.

First we recall the action and index of orbits in the exact symplectic manifold (X, θ). Let $F : S^1 \times X \to \mathbb{R}$ be a Hamiltonian, and $\gamma : S^1 \to X$ a non-degenerate orbit of F. Its action is defined to be

$$\mathcal{A}_F(\gamma) := \int_{S^1} F(t, \gamma(t))\, \mathrm{d}t + \int_{S^1} \gamma^*\theta.$$

To associate an index to orbits, we require an additional piece of data: a homotopy class of trivializations η of $\Lambda^{\mathrm{top}}_{\mathbb{C}}(TX)^{\otimes 2N}$, for some integer $N > 0$. To define the index $i_\eta(\gamma)$ of an orbit γ, we first choose a trivialization Φ of γ^*TX; we denote the Conley–Zehnder index with respect to this trivialization by $\mathrm{CZ}(\gamma, \Phi)$. The trivialization Φ induces a trivialization of $\Lambda^{\mathrm{top}}_{\mathbb{C}}(\gamma^*TX)^{\otimes 2N}$, and we define $w(\Phi, \eta) \in \mathbb{Z}$ to be the winding number of

$$\eta^{-1} \circ \Lambda^{\mathrm{top}}_{\mathbb{C}}(\Phi)^{\otimes 2N} : S^1 \to \mathbb{C}^*.$$

We then define

$$i_\eta(\gamma) = \mathrm{CZ}(\gamma, \Phi) + \frac{dim(X)}{2} - \frac{w(\Phi, \eta)}{N}.$$

One easily checks that the index is independent of the trivialization Φ. Note that it is fractional: $i_\eta(\gamma) \in \frac{1}{N}\mathbb{Z}$.

In our setting, the relevant choice of trivialization η is determined by $\boldsymbol{\lambda}$. Let N be an integer such that $N\lambda_i \in \mathbb{Z}$ for all i. Then $\sum_i N\lambda_i[D_i]$ is Poincaré dual to $c_1(\Lambda^{\mathrm{top}}_{\mathbb{C}}(TM)^{\otimes 2N})$ by definition, so we may choose a section of $\Lambda^{\mathrm{top}}_{\mathbb{C}}(TM)^{\otimes 2N}$ which is non-vanishing over X, and vanishes with multiplicity $N\lambda_i$ along D_i. Restricting this section to X defines a homotopy class of trivializations of $\Lambda^{\mathrm{top}}_{\mathbb{C}}(TX)^{\otimes 2N}$, which we denote by $\eta_{\boldsymbol{\lambda}}$. We will write $i(\gamma)$ for $i_{\eta_{\boldsymbol{\lambda}}}(\gamma)$.

Now let $F : S^1 \times M \to \mathbb{R}$ be a Hamiltonian, and $\gamma : S^1 \to X$ a non-degenerate orbit of F which is contractible in M, and contained inside X. We define a canonical fractional cap u_{in} for γ, by setting $u_{\mathrm{in}} := u - u \cdot \boldsymbol{\lambda}$ for an arbitrary cap u; the result is clearly independent of u. One should think of u_{in} as a 'cap inside X': indeed, if u were a cap contained inside X, we would have $u_{\mathrm{in}} = u$.

Lemma 3.8. *We have*

$$i(\gamma) = i(\gamma, u_{\mathrm{in}}) \qquad \textit{and} \qquad \mathcal{A}_F(\gamma) = \mathcal{A}_F(\gamma, u_{\mathrm{in}}).$$

Proof. Let us choose an arbitrary $u : \mathbb{D} \to M$ capping γ. We start with the action. Directly from the definitions:

$$\mathcal{A}_F(\gamma) = \int_{S^1} F(t, \gamma(t))\, \mathrm{d}t + \int_{S^1} \gamma^*\theta$$
$$\text{and} \qquad \mathcal{A}_F(\gamma, u_{\mathrm{in}}) = \int_{S^1} F(t, \gamma(t))\, \mathrm{d}t + \int_{\mathbb{D}} u^*\omega - \kappa u \cdot \boldsymbol{\lambda}.$$

Therefore, the result follows from the assumption that the relative de Rham cohomology class of (ω, θ) is $\kappa\boldsymbol{\lambda}$.

Recalling definitions for indices:

$$i(\gamma) = \mathrm{CZ}(\gamma, \Phi) + \frac{\dim(X)}{2} - \frac{w(\Phi, \eta_{\boldsymbol{\lambda}})}{N},$$

where we choose Φ to be the trivialization of γ^*TX induced by the cap u, and

$$i(\gamma, u) = \mathrm{CZ}(\gamma, u) + \frac{\dim(M)}{2} - u \cdot \boldsymbol{\lambda}.$$

Therefore, we need to show that

$$w(\Phi, \eta_{\boldsymbol{\lambda}}) = Nu \cdot \boldsymbol{\lambda}.$$

This follows because $\eta_{\boldsymbol{\lambda}}$ actually induces a section of $\Lambda_{\mathbb{C}}^{\mathrm{top}}(u^*TM)^{\otimes 2N}$. Using any trivialization of $\Lambda_{\mathbb{C}}^{\mathrm{top}}(u^*TM)^{\otimes 2N}$, we can think of this section as a map $\mathbb{D} \to \mathbb{C}$, which does not vanish along the boundary. The degree of this map at $0 \in \mathbb{C}$ is easily computed to be $Nu \cdot \lambda$ using that $\eta_{\boldsymbol{\lambda}}$ vanishes with multiplicity $N\lambda_i$ along D_i. It is an elementary fact that the same degree is also equal to the winding number that we are interested in, so the result follows. □

3.4. Positivity of intersection

In this section, we prove a result based on Abouzaid–Seidel's 'integrated maximum principle'. We will later use it to prove Proposition 1.12, although the result is more broadly applicable.

Let (W, ω) be a symplectic manifold with a concave boundary modelled on the contact manifold (Y, θ). This means that $\partial W = Y$, and there is a symplectic embedding of the symplectization $(Y \times [c, c + \epsilon), d(\rho \cdot \theta))$ onto a neighbourhood of the boundary, where $\rho \in [c, c + \epsilon)$ is the Liouville coordinate. Note that as $\omega|_Y = c d\theta$, we have a relative de Rham cohomology class $[\omega; c\theta] \in H^2(W, Y)$. We will consider $u : (\Sigma, \partial\Sigma) \to (W, Y)$ satisfying the pseudoholomorphic curve equation for a certain class of almost-complex structures and Hamiltonian perturbations, and give a criterion guaranteeing that $[\omega; c\theta](u) \geq 0$, with equality if and only if $u \subset Y$.

To define our pseudoholomorphic curve equation, we choose a complex structure j on Σ, a family $\mathcal{J}$ of ω-compatible almost-complex structures J_z parametrized by $z \in \Sigma$, and a Hamiltonian-valued one-form $\mathcal{K} \in \Omega^1(\Sigma; C^\infty(W))$. Note that differential forms on $\Sigma \times W$ decompose into types:

$$\Omega^\bullet(\Sigma \times W) = \bigoplus_{j+k=\bullet} \Omega^j(\Sigma, \Omega^k(W)),$$

so we may interpret $\mathcal{K}$ as a one-form on $\Sigma \times W$. The de Rham differential decomposes as $d = d_\Sigma + d_W$, where

$$d_\Sigma : \Omega^j(\Sigma, \Omega^k(W)) \to \Omega^{j+1}(\Sigma, \Omega^k(W)) \qquad \text{and}$$
$$d_W : \Omega^j(\Sigma, \Omega^k(W)) \to \Omega^j(\Sigma, \Omega^{k+1}(W)).$$

The isomorphism $C^\infty(TW) \to \Omega^1(W)$ sending $v \mapsto \omega(v, -)$ allows us to turn $d_W\mathcal{K}$ into a Hamiltonian-vector-field-valued one-form $X_{\mathcal{K}} \in \Omega^1(\Sigma; C^\infty(TW))$. We will consider the pseudoholomorphic curve equation

$$(\mathrm{d}u - X_{\mathcal{K}})^{0,1} = 0.$$

Note that the $(0,1)$-projection of $v \in \Omega^1(\Sigma; C^\infty(TW))$ is given by $\frac{1}{2}(v + \mathcal{J} \circ v \circ j)$.

We introduce the geometric energy of a pseudoholomorphic curve u:

$$E_{\mathrm{geom}}(u) = \frac{1}{2}\int_\Sigma \|\mathrm{d}u - X_{\mathcal{K}}\|^2 .$$

It is manifestly non-negative. Let $\tilde{u} : \Sigma \to \Sigma \times W$ denote the graph of u. We have the standard computation (e.g. Equation (8.12) of [32]):

$$E_{\mathrm{geom}}(u) = \int_\Sigma u^*\omega + \tilde{u}^*\left(d_W\mathcal{K} + \{\mathcal{K},\mathcal{K}\}\right),$$

where the final term lives in $\Omega^2(\Sigma, C^\infty(W))$ and is defined by $\{\mathcal{K},\mathcal{K}\}(v,w) := \{\mathcal{K}(v),\mathcal{K}(w)\}$, where $\{-,-\}$ is the Poisson bracket.

We also introduce the topological energy

$$E_{\mathrm{top}}(u) := \int_\Sigma u^*\omega + \tilde{u}^*\mathrm{d}\mathcal{K}.$$

Note that

$$E_{\mathrm{top}}(u) = E_{\mathrm{geom}}(u) + \int_\Sigma \tilde{u}^*\left(\mathrm{d}_\Sigma\mathcal{K} - \{\mathcal{K},\mathcal{K}\}\right).$$

Proposition 3.9. *Suppose that*

(1) *J_z is of contact type along Y, for all $z \in \partial\Sigma$:*

$$\mathrm{d}\rho \circ J_z = -\rho\theta.$$

(2) *There exist one-forms $\alpha, \beta \in \Omega^1(\Sigma)$ such that $\mathcal{K} = \alpha \cdot \rho + \beta$ in a neighbourhood of Y*

(3) *We have $\mathrm{d}_\Sigma\mathcal{K} - \{\mathcal{K},\mathcal{K}\} - \mathrm{d}\beta \geq 0$.*[12]

Then any smooth map $u : (\Sigma, \partial\Sigma) \to (W, Y)$ satisfying $(\mathrm{d}u - X_{\mathcal{K}})^{0,1} = 0$, with $\partial\Sigma \neq \emptyset$, will satisfy $[\omega; c\theta](u) \geq 0$, with equality if and only if $u \subset Y$.

Proof. We have

$$\begin{aligned}
[\omega; c\theta](u) &= \int_\Sigma u^*\omega - \int_{\partial\Sigma} u^* c\theta \\
&= E_{\mathrm{geom}}(u) - \int_\Sigma \tilde{u}^*\left(\mathrm{d}_W\mathcal{K} + \{\mathcal{K},\mathcal{K}\}\right) - c\int_{\partial\Sigma} u^*\theta \\
&\geq \int_\Sigma \tilde{u}^*\left(-\mathrm{d}\mathcal{K} + \mathrm{d}_\Sigma\mathcal{K} - \{\mathcal{K},\mathcal{K}\}\right) - c\int_{\partial\Sigma} u^*\theta \qquad \text{as } E_{\mathrm{geom}}(u) \geq 0 \\
&\geq \int_\Sigma \tilde{u}^*\left(-\mathrm{d}\mathcal{K} + \mathrm{d}\beta\right) - c\int_{\partial\Sigma} u^*\theta \qquad \text{by hypothesis (3)} \\
&= \int_{\partial\Sigma} -\tilde{u}^*\mathcal{K} + \beta - c \cdot u^*\theta.
\end{aligned}$$

By hypothesis (2), the first and second terms combine to give

$$\int_{\partial\Sigma} -\tilde{u}^*(\alpha \cdot \rho + \beta) + \beta = -\int_{\partial\Sigma} c \cdot \alpha,$$

[12] Given $\xi \in \Omega^2(\Sigma, C^\infty(W))$, we say that $\xi \geq 0$ if for all $z \in \Sigma$, $v \in T_z\Sigma$, and $w \in W$, we have $\xi(v, jv)(w) \geq 0$.

as $\rho = c$ along Y.

We can analyse the remaining term using the argument in [2, Lemma 7.2]. Let $v \in T_z\partial\Sigma$ be a positively-oriented boundary vector. Using the Floer equation

$$(\mathrm{d}u - X_{\mathcal{K}})^{0,1} = 0,$$

we obtain

$$u_*(v) = -Ju_*j(v) + X_{\mathcal{K}}(v) + JX_{\mathcal{K}}j(v),$$

so

$$u^*\theta(v) = -\theta\left(Ju_*j(v)\right) + \theta\left(X_{\mathcal{K}}(v)\right) + \theta\left(JX_{\mathcal{K}}j(v)\right).$$

We analyse each term on the RHS. For the first, we note that $j(v)$ points into Σ. Therefore $u_*(j(v))$ points into W. Such vectors can be written as the sum of a non-negative multiple of the Liouville vector and a vector that is tangent to Y. Because J is of contact type, this implies that

$$\theta\left(Ju_*j(v)\right) \geq 0.$$

For the second, we note that hypothesis (2) ensures that $X_{\mathcal{K}}(v) = -\alpha(v) \cdot \mathcal{R}$, where $\mathcal{R}$ is the Reeb vector field on Y. Thus, $\theta(X_{\mathcal{K}}(v)) = -\alpha(v)$. For the third, hypothesis (2) again ensures that $X_{\mathcal{K}}(j(v))$ is a multiple of the Reeb vector field; because J is of contact type, $\theta(JX_{\mathcal{K}}j(v)) = 0$. Putting it all together, we have

$$u^*\theta(v) \leq -\alpha(v).$$

Combining, we finally obtain

$$[\omega; c\theta](u) \geq \int_{\partial\Sigma} -c \cdot \alpha + c \cdot \alpha = 0$$

as required.

If equality holds then we have $E_{\mathrm{geom}}(u) = 0$, which implies that $\mathrm{d}u = X_{\mathcal{K}}$. Hypothesis (2) then implies that $u_*(v) = X_{\mathcal{K}}(v)$ is a multiple of the Reeb vector field $\mathcal{R}$ in a neighbourhood of Y, for all v; as $\mathcal{R}$ is tangent to Y, this implies that u is contained in Y. □

Remark 3.10. Note that if $\mathcal{K}' = \mathcal{K} + \xi$, where $\xi \in \Omega^1(\Sigma)$, then $X_{\mathcal{K}} = X_{\mathcal{K}'}$, so the associated pseudoholomorphic curve equations are identical. Thus, we would expect that if the hypotheses of Proposition 3.9 hold for $\mathcal{K}$, then they should also hold for $\mathcal{K}'$. Indeed, Hypothesis (2) holds, as $\mathcal{K}' = \alpha \cdot \rho + \beta'$, where $\beta' = \beta + \xi$; and Hypothesis (3) also holds, because $\mathcal{K}' - \beta' = \mathcal{K} - \beta$.

Proposition 3.9 is designed to prove Proposition 1.12 (= Proposition 5.10), but there are other natural situations where Hypotheses (2) and (3) can be made to hold. The simplest, of course, is if $\mathcal{K}$ vanishes in a neighbourhood of Y. Alternatively, similarly to [2], we may have $\mathcal{K} = H \cdot \gamma$ where H is independent of $z \in \Sigma$, $H = a\rho + b$ in a neighbourhood of Y, $H \geq b$ over W, and $d\gamma \geq 0$.

4. Special Hamiltonian

Our goal in this section is to construct the special functions $\rho^R : M \to \mathbb{R}$, defined for $R > 0$ sufficiently small, as mentioned Sect. 1.5. Recall their key properties:

- ρ^R is continuous on M, and smooth on the complement of the skeleton $\mathbb{L}$;
- $\rho^R|_{\mathbb{L}} = 0$ and $\rho^R|_D \approx 1$;
- we have $Z(\rho^R) = \rho^R$ on $X \backslash \mathbb{L}$, where Z is the Liouville vector field on (X, θ);
- $\rho^R \to \rho^0$ as $R \to 0$.

Having constructed the functions ρ^R, we use them to construct the Hamiltonians on M which we use in our main arguments; and we compute the action and index of the orbits of these Hamiltonians. The results are expressed in Lemmas 4.21 and 4.24.

We use the geometric setup of Sect. 1.1 with slight modifications in light of Sect. 2. Let us spell this out fully. We have a closed symplectic manifold (M, ω) which is monotone

$$2\kappa c_1^M = [\omega] \in H^2(M; \mathbb{R}) \quad \text{with} \quad \kappa > 0,$$

$D = \cup_{i=1}^N D_i \subset (M, \omega)$ is an orthogonal simple crossings divisor and $\lambda_1, \ldots, \lambda_N \in \mathbb{Q}_{>0}$ is a choice of weights. We denote $X = M \backslash D$ and $\boldsymbol{\lambda} \in \mathbb{R}^N \cong H^2(M, X; \mathbb{R})$ is the associated lift of $2c_1^M$. We also choose an admissible system of commuting Hamiltonians $\{r_i : UD_i \to [0, R_0)\}$ near D and a primitive $\theta \in \Omega^1(X)$ of $\omega|_X$ such that the relative de Rham cohomology class of (ω, θ) is $\kappa\boldsymbol{\lambda}$. We assume that θ is adapted to $\{r_i : UD_i \to [0, R_0)\}$ and that

$$R_0 < \kappa\lambda_i, \text{ for all } i. \tag{4.1}$$

The last condition can be achieved by shrinking the ascH (as explained in Sect. 2.2).

In fact, we will consider the $(0, R_0)$-family of such data obtained by shrinking the ascH to radius $R \in (0, R_0)$, while keeping all else fixed. The parameter R will also[13] be used as the 'smoothing parameter' for ρ^R. In Sect. 5, we will want R to be sufficiently small for certain arguments to work. The approximations in this section (such as $\rho^R|_D \approx 1$) will be more and more accurate as R tends to 0. The dependence on R of our constructions below should be understood in this light.

4.1. Overview of the construction of ρ^R

The first step in the construction is to enlarge the sets UD_i via the Liouville flow. This gives us open sets $UD_i^{\max}$, together with toric moment maps $r_i^{\max} : UD_i^{\max} \to [0, \kappa\lambda_i)$, such that $\cup_i UD_i^{\max} = M \backslash \mathbb{L}$. For $I \subset [N]$, we define $UD_I^{\max} = \cap_{i \in I} UD_i^{\max}$, and we have toric moment maps $r_I^{\max} : UD_I^{\max} \to \prod_{i \in I} [0, \kappa\lambda_i)$.

[13] We note that this is for notational convenience only.

Now for each non-empty $I \subset [N]$, we define open subsets $\mathring{U}D_I^{\max} \subset UD_I^{\max}$ so that $\cup_I \mathring{U}D_I^{\max} = M \setminus \mathbb{L}$. We will define $\rho^R|_{\mathring{U}D_I^{\max}} = \tilde{\rho}_I^R \circ r_I^{\max}$, for smooth functions

$$\tilde{\rho}_I^R : \prod_{i \in I} [0, \kappa\lambda_i) \to \mathbb{R}$$

carefully chosen so that the definition agrees on the overlaps and ρ^R satisfies the desired key properties. In fact, $\tilde{\rho}_I^R$ will be well defined on the larger region

$$V_I := \mathbb{R}^I \setminus \prod_{i \in I} [\kappa\lambda_i, \infty).$$

Let us briefly discuss how we will ensure that ρ^R thus defined satisfies $Z(\rho^R) = \rho^R$. We translate this into a property of the functions $\tilde{\rho}_I^R$. We denote the standard projection by $\mathrm{pr}_I : \mathbb{R}^N \to \mathbb{R}^I$, and set $\boldsymbol{\lambda}_I := \mathrm{pr}_I(\boldsymbol{\lambda})$. We consider the (Euler-type) vector field $\tilde{Z}_I$ on $\mathbb{R}^I$ defined by

$$\left(\tilde{Z}_I\right)_r := \sum_{i \in I} (r_i - \kappa\lambda_i) \frac{\partial}{\partial r_i}.$$

Lemma 4.1. *For all $x \in UD_I \setminus D$,*

$$(r_I)_* Z_x = \left(\tilde{Z}_I\right)_{r_I(x)}.$$

Proof. Follows from the fact that $Z(r_i) = r_i - \kappa\lambda_i$, as θ is adapted to the scH. □

In fact, $UD_I^{\max}$ and $r_I^{\max}$ are constructed so that Lemma 4.1 also holds if we put *max* superscripts on the r_I and UD_I (Lemma 4.4). This gives us

Corollary 4.2. *The function $\rho^R := \tilde{\rho}_I^R \circ r_I^{\max}$ satisfies $Z(\rho^R) = \rho^R$ if and only if $\tilde{Z}_I(\tilde{\rho}_I^R) = \tilde{\rho}_I^R$.*

Note that a function $f : V_I \to \mathbb{R}$ satisfies $\tilde{Z}_I(f) = f$ if and only if it is linear along the rays emanating from $\kappa\boldsymbol{\lambda}_I$, converging to 0 at that point.

The functions $\tilde{\rho}_I^R$ will be constructed roughly as follows. We will choose a hypersurface $\tilde{Y}_I^R \subset V_I \cap \mathbb{R}^I_{\geq 0}$ which is a smoothing of $\tilde{Y}_I^0 := \partial\mathbb{R}^I_{\geq 0}$, satisfying certain properties (see Lemma 4.8). Then, we will define $\tilde{\rho}_I^R$ as the function that is linear along the rays emanating from $\kappa\boldsymbol{\lambda}_I$, converging to zero at that point, and takes the value 1 on $\tilde{Y}_I^R$.

For the other key properties of ρ^R let us mention the following slightly sketchy point to orient the reader. Recall that ρ^R is supposed to be a smoothing of the continuous function $\rho^0 : M \to \mathbb{R}$ introduced in Sect. 1.3, which has all of the properties we need (e.g., it satisfies $\rho^0|_{\mathbb{L}} = 0$, $\rho^0|_D = 1$ and $Z(\rho^0) = \rho^0$), except it is not smooth. We now give an alternative description of the function ρ^0, which is parallel with the construction of ρ^R. We extend the function $\frac{\kappa\lambda_i - r_i^{\max}}{\kappa\lambda_i} : UD_i^{\max} \to \mathbb{R}$ to M by defining it to be 0 everywhere outside of its original domain of definition. Let us momentarily denote this extension with the same notation. Then we have

$$\rho^0 = \max_i \frac{\kappa\lambda_i - r_i^{\max}}{\kappa\lambda_i}.$$

In particular, on $UD_I^{\max}$, we have $\rho^0 = \tilde{\rho}_I^0 \circ r_I^{\max}$, where $\tilde{\rho}_I^0(r) = \max_{i\in I} \frac{\kappa\lambda_i - r_i}{\kappa\lambda_i}$. Note that $\tilde{\rho}_I^0$ is equal to 0 at $\kappa\boldsymbol{\lambda}_I$, linear along the rays emanating from this point, and equal to 1 along $\tilde{Y}_I^0$. The functions $\tilde{\rho}_I^R$ mentioned above will be consistently chosen smoothings of the functions $\tilde{\rho}_I^0$.

Remark 4.3. We would like to warn the reader of an abuse of notation we already committed a couple of times above and will continue with below. We will use r_i both as the function $r_i : UD_i \to [0, R)$ and also the i^{th} coordinate function on $\mathbb{R}^I$ with $i \in I$. We believe that this will not cause too much confusion, partly because often we will actually need to be using $r_i^{\max} : UD_i^{\max} \to [0, \kappa\lambda_i)$ in place of the former anyway.

4.2. Construction of $UD_I^{\max}$, $r_I^{\max}$, $\mathring{U}D_I^{\max}$

Note that $UD_i \backslash D_i$ is closed under the positive Liouville flow as long as the flow is defined, by Lemma 4.1 and Eq. (4.1). Let us define $UD_i^{\max} \subset M$ as the union of UD_i with the set of points in X that enter into UD_i under the positive Liouville flow in finite time. Of course we have $UD_i \subset UD_i^{\max}$. Note that $UD_i^{\max}$ depends on R just as UD_i does (unless $D = D_i$ is smooth); nevertheless we suppress R from the notation.

We extend r_i to

$$r_i^{\max} : UD_i^{\max} \to \mathbb{R}_{\geq 0}$$

by first flowing into UD_i with the Liouville flow in some time $T \geq 0$, applying r_i, and then flowing with $\tilde{Z}_{\{i\}}$ for time $-T$. This is well defined and smooth by Lemma 4.1.

Let us also define $UD_I^{\max} := \cap_{i\in I} UD_i^{\max}$ and

$$r_I^{\max} : UD_I^{\max} \to \mathbb{R}^I_{\geq 0}.$$

Then the following is true by construction:

Lemma 4.4. *Lemma 4.1 holds if we put max superscripts on the r_I and UD_I.*

Now, recall that $\tilde{Y}_I^0 := \partial\mathbb{R}^I_{\geq 0}$. Define the projection-from-$\kappa\boldsymbol{\lambda}_I$ map

$$P_I : V_I \to \tilde{Y}_I^0,$$

which flows a point along $\tilde{Z}_I$ until it intersects $\tilde{Y}_I^0$.

Now, let us define $UD_i^{1/2} := \{r_i \leq R/2\} \subset UD_i = \{r_i < R\}$. Define $UD_i^{1/2,\max}$ to be the union of $UD_i^{1/2}$ with the set of points in X that enter into $UD_i^{1/2}$ under the positive Liouville flow in finite time.

Definition 4.5. For $I \subset [N]$, define

$$\mathring{U}D_I^{\max} := UD_I^{\max} \backslash \bigcup_{j\notin I} UD_j^{1/2,\max}.$$

(Fig. 1 may help the reader visualize these sets.)

Lemma 4.6. *The sets $\left\{\mathring{U}D_I^{\max}\right\}_{\emptyset\neq I\subset[N]}$ form an open cover of $M\backslash\mathbb{L}$.*

Proof. Because $UD_i^{1/2,\max}$ is closed in $M\backslash\mathbb{L}$, and contained in $UD_i^{\max}$, the sets $UD_i^{\max}$ and $\left(UD_i^{1/2,\max}\right)^c$ form an open cover of $M\backslash\mathbb{L}$ for all i. Taking the intersection of these open covers over all i gives us an open cover by the sets

$$\bigcap_{i\in I} UD_i^{\max} \cap \bigcap_{i\notin I} \left(UD_i^{1/2,\max}\right)^c = \mathring{U}D_I^{\max}$$

for $I \subset [N]$. It remains to check that $\mathring{U}D_\emptyset^{\max} = \emptyset$. This follows from the fact that $\cup_i UD_i^{1/2,\max} = M\backslash\mathbb{L}$ (because every flowline of the Liouville vector field in $X\backslash\mathbb{L}$ ultimately enters $\cup_i UD_i^{1/2}$). □

The following is an easy consequence of Lemma 4.1 and the construction of $UD_I^{\max}$ and $UD_i^{1/2,\max}$:

Lemma 4.7. *If $i \in I$, then*

$$UD_I^{\max}\backslash UD_i^{1/2,\max} = (P_I \circ r_I^{\max})^{-1}\left(\{r_i > R/2\}\right).$$

4.3. Construction of $\tilde{Y}_I^R$

For any $R_0 > R > 0$ (as always in this section), let $q^R : \mathbb{R} \to \mathbb{R}$ be a function satisfying:

- $q^R(r) = 0$ for $r \geq R/2$;
- $(q^R)'(r) < 0$ for $r < R/2$;
- $q^R(0) = 1$.

Consider

$$Q_I^R : \mathbb{R}^I \to \mathbb{R}$$
$$Q_I^R(r) := \sum_{i\in I} q^R(r_i).$$

Now define

$$\tilde{Y}_I^R := \{Q_I^R = 1\}$$

(see Fig. 1).

Lemma 4.8. *The hypersurfaces $\tilde{Y}_I^R \subset \mathbb{R}^I$ have the following properties:*

1. *$\tilde{Y}_I^R$ is contained in the region $V_{I,\geq 0} := V_I \cap \mathbb{R}^I_{\geq 0}$.*
2. *Every flowline of $\tilde{Z}_I$ in V_I crosses $\tilde{Y}_I^R$ transversely at a unique point.*
3. *If $\tilde{\nu}_I^R : \tilde{Y}_I^R \to \mathbb{R}^I$ is a normal vector field (pointing towards the component containing $\kappa\boldsymbol{\lambda}$), then $\tilde{\nu}_{I,i}^R \geq 0$ for all i. (Here $\tilde{\nu}_{I,i}^R$ is the ith component of $\tilde{\nu}_I^R$.)*
4. *For any $J \subset I$, $\tilde{Y}_I^R$ coincides with $\tilde{Y}_J^R \times \mathbb{R}^{I\backslash J}$ over the region $\cap_{i\in I\backslash J}$ $\{\tilde{\nu}_{I,i}^R = 0\}$.*
5. *The region $\{\tilde{\nu}_{I,i}^R = 0\}$ contains $P_I^{-1}\left(\{r_i > R/2\}\right)$.*

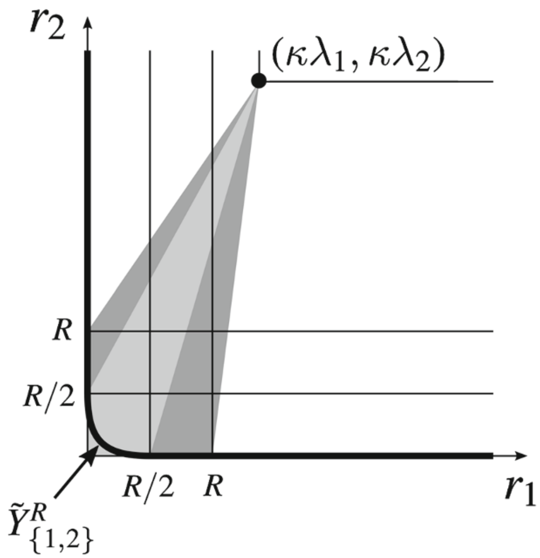


FIGURE 1. The hypersurface $\tilde{Y}^R_{\{1,2\}}$. The image of $r^{\max}_{\{1,2\}}$ is shaded. The images of the regions $\mathring{U}D^{\max}_{\{1\}} \cap \mathring{U}D^{\max}_{\{1,2\}}$ and $\mathring{U}D^{\max}_{\{2\}} \cap \mathring{U}D^{\max}_{\{1,2\}}$ are shaded darker

Proof. Property (1) follows from the fact that $Q^R_I \geq 1$ if any $r_i \leq 0$ and $Q^R_I = 0$ if all $r_i \geq R/2$.

To prove property (2), we first observe any flowline of $\tilde{Z}_I$ in V_I starts at $\kappa\boldsymbol{\lambda}_I$, where $Q^R_I = 0$, and ends up outside $V_{I,\geq 0}$, where $Q^R_I \geq 1$, so it must cross $\tilde{Y}^R_I$ somewhere. Furthermore, we have that $\tilde{Z}_I(Q^R_I) \geq 0$ for any i: we have

$$\tilde{Z}_I(q^R(r_i)) = (r_i - \kappa\lambda_i) \cdot (q^R)'(r_i),$$

where $(q^R)'(r_i) \leq 0$, and $r_i - \kappa\lambda_i < R/2 - \kappa\lambda_i < 0$ wherever $(q^R)'(r_i) \neq 0$. Finally, we have $\tilde{Z}_I(Q^R_I) > 0$ along $\tilde{Y}^R_I$, because at any point on $\tilde{Y}^R_I$ we have $q^R(r_i) > 0$ and hence $(q^R)'(r_i) < 0$ for some i.

Property (3) follows from the fact that $\partial Q^R_I/\partial r_i \leq 0$ for all i. Property (4) follows from the fact that $(q^R)'(r_i) = 0$ if and only if $q^R(r_i) = 0$. Property (5) follows from the fact that

$$P_I^{-1}(\{r_i > R/2\}) \cap \tilde{Y}^R_I \quad \subset \quad P_I^{-1}(\{r_i > R/2\}) \cap V_{I,\geq 0} \quad \subset \quad \{r_i > R/2\},$$

and $(q^R)'(r_i) = 0$ for $r_i > R/2$. $\square$

Remark 4.9. The hypersurface $\tilde{Y}^R_I$ has the additional property (which we will not use, but which may help the reader to visualize the construction) that it coincides with $\tilde{Y}^0_I$ away from a neighbourhood of the singular locus of the latter. We can also choose q^a to be a convex function, which would imply that the component of $\mathbb{R}^I - \tilde{Y}^R_I$ that does not contain 0 is convex (which would in turn imply that $\tilde{\rho}^R_I$ is convex). Again, we do not need this property.

4.4. Construction of $\tilde{\rho}_I^R$

By property (2) of $\tilde{Y}_I^R$, there is a unique smooth function $\tilde{\rho}_I^R : V_I \to \mathbb{R}$ satisfying

$$\tilde{\rho}_I^R|_{\tilde{Y}_I^R} = 1 \qquad \text{and} \qquad \tilde{Z}_I(\tilde{\rho}_I^R) = \tilde{\rho}_I^R.$$

Recall that the second condition means that $\tilde{\rho}_I^R$ is linear along the rays emanating from $\kappa\boldsymbol{\lambda}_I$, converging to zero at $\kappa\boldsymbol{\lambda}_I$. In particular, the level sets of $\tilde{\rho}_I^R$ are scalings of $\tilde{Y}_I^R$ centred at $\kappa\boldsymbol{\lambda}_I$.

Lemma 4.10. *If $J \subset I$, then*

$$\tilde{\rho}_I^R = \tilde{\rho}_J^R \circ \mathrm{pr}_{IJ} \qquad \textit{over the region} \qquad \bigcap_{i \in I \setminus J} P_I^{-1}\left(\{r_i > R/2\}\right),$$

where $\mathrm{pr}_{IJ} : \mathbb{R}^I \to \mathbb{R}^J$ is the natural projection.

Proof. Follows from the fact that $\tilde{Y}_I^R$ coincides with $\tilde{Y}_J^R \times \mathbb{R}^{I\setminus J}$ in the given region, by properties (4) and (5) of $\tilde{Y}_I^R$. □

4.5. Construction of ρ^R

Lemma 4.11. *For any $\emptyset \neq I, J \subset [N]$, we have*

$$\tilde{\rho}_I^R \circ r_I^{\max} = \tilde{\rho}_J^R \circ r_J^{\max} \qquad \textit{over} \qquad \mathring{U}D_I^{\max} \cap \mathring{U}D_J^{\max}.$$

Proof. First note that $\mathring{U}D_I^{\max} \cap \mathring{U}D_J^{\max} \subset UD_{I\cup J}^{\max}$. We have

$$\mathring{U}D_I^{\max} \cap \mathring{U}D_J^{\max} = \bigcap_{k \notin I \cap J} \left(P_{I\cup J} \circ r_{I\cup J}^{\max}\right)^{-1}\left(\{r_k > R/2\}\right),$$

as an immediate consequence of Lemma 4.7. Over this set, we have

$$\tilde{\rho}_I^R \circ r_I^{\max} = \tilde{\rho}_I^R \circ \mathrm{pr}_{I\cup J, I} \circ r_{I\cup J}^{\max} = \tilde{\rho}_{I\cup J}^R \circ r_{I\cup J}^{\max}$$

by Lemma 4.10. The result now follows by applying the same argument to $\tilde{\rho}_J^R \circ r_J^{\max}$. □

Lemmas 4.6 and 4.11 allow us to define:

Definition 4.12. We define $\rho^R : M \to \mathbb{R}$ to be equal to $\tilde{\rho}_I^R \circ r_I^{\max}$ over each $\mathring{U}D_I^{\max}$, and equal to 0 over $\mathbb{L}$.

To check that ρ^R is continuous along $\mathbb{L}$, we use the fact that $Z(\rho^R) = \rho^R$ on $M \setminus \mathbb{L}$ by Corollary 4.2, and the level sets of ρ^R are compact submanifolds disjoint from $\mathbb{L}$. It follows that $\rho^R \to 0$ as we go towards $\mathbb{L}$, so ρ^R is continuous along $\mathbb{L}$.

Definition 4.13. Because $Z(\rho^R) = \rho^R$, and $\rho^R|_D \geq 1$, the subset $K_\sigma^R := \{\rho^R \leq \sigma\}$ is a Liouville subdomain of X for any $\sigma \in (0,1)$. The contact manifold $Y_\sigma^R = \partial K_\sigma^R$ with contact form $\sigma^{-1}\iota_{Y_\sigma^R}^*\theta$ is independent of σ.

For the remainder of this section, we will drop R from the notation: so we write ρ instead of ρ^R, etc.

4.6. The Hamiltonian and its orbits

Let $h : \mathbb{R} \to \mathbb{R}$ be a smooth function which is constant on a neighbourhood of 0. It is clear that the function $h\circ\rho$ is smooth on M. We denote its Hamiltonian flow by $\Phi_t^{h\circ\rho}$. To describe the orbits of $h\circ\rho$, we first compute $\mathrm{d}\rho$.

Lemma 4.14. *There exist smooth functions $\nu_i : M\backslash\mathbb{L} \to \mathbb{R}_{\geq 0}$, supported in $UD_i^{\max}$, such that*

$$(\mathrm{d}\rho)_m = -\sum_i \nu_i(m)\cdot (dr_i^{\max})_m\,.$$

(Here the LHS denotes the value of the one-form $\mathrm{d}\rho$ at the point m. The RHS is well defined, even though $dr_i^{\max}$ is only defined over $UD_i^{\max}$, because ν_i vanishes outside $UD_i^{\max}$.)

Proof. For any i, I, we define the following function on V_I:

$$\tilde{\nu}_{I,i} := \begin{cases} -\partial\tilde{\rho}_I/r_i & \text{if } i\in I \\ 0 & \text{else} \end{cases}$$

Note that it is non-negative by property (3) of $\tilde{Y}_I$. We claim that for any i, and any $J\subset I$, we have

$$\tilde{\nu}_{I,i} = \tilde{\nu}_{J,i}\circ \mathrm{pr}_{IJ} \qquad \text{over the region} \qquad \bigcap_{i\in I\backslash J} P_I^{-1}\left(\{r_i > R/2\}\right).$$

If $i\in I$, this follows by Lemma 4.10 (there are two cases: $i\in J$ and $i\in I\backslash J$). If $i\notin I$, it is obvious as both functions are 0. This allows us to mimic the construction of ρ: we set $\nu_i = \tilde{\nu}_{I,i}\circ r_I^{\max}$ over $\mathring{U}D_I^{\max}$. We finally observe that $d\tilde{\rho}_I = -\sum_i \tilde{\nu}_{I,i}dr_i$, which completes the proof. ☐

For any $m\in M-\mathbb{L}$, we define $I(m) := \{i : \nu_i(m)\neq 0\}$. We have $m\in UD_{I(m)}^{\max}$.

We define $\nu : M\backslash\mathbb{L}\to\mathbb{R}^N$ to be the smooth function with coordinates $(\nu_1,\ldots,\nu_N)$. We note that the function $h'(\rho)\cdot\nu : M\backslash\mathbb{L}\to\mathbb{R}^N$ extends smoothly to M, and we denote this extension by $\nu^h : M\to\mathbb{R}^N$. Note that ν^h is constant along orbits of $h\circ\rho$, so we have a well-defined $\nu^h(\gamma)\in\mathbb{R}^N$ associated with such an orbit γ. We can interpret $\nu_i^h(\gamma)$ as 'the number of times γ wraps around D_i' (it is an integer unless γ is contained in D_i, see Lemma 4.16 below). We define

$$I(\gamma) := \{i : \nu_i^h(\gamma)\neq 0\}\subset [N].$$

Note that if $h'(\rho)\geq 0$, then $\nu^h(\gamma)\in\mathbb{R}^N_{\geq 0}$.

Corollary 4.15. *For any $m\in M\backslash\mathbb{L}$, we have $\Phi_1^{h\circ\rho}(m) = \nu^h(m)\cdot m$. To explain the notation, $\nu^h(m)\in\mathbb{R}^{I(m)}$ gets projected to $(\mathbb{R}/\mathbb{Z})^{I(m)}$, which then acts on $m\in UD_{I(m)}^{\max}$ by the Hamiltonian torus action.*

Lemma 4.16. *We have $\Phi_1^{h\circ\rho}(m) = m$ if and only if for all i, either $m\in D_i$ or $\nu_i^h(m)\in\mathbb{Z}$.*

Proof. For $m\in\mathbb{L}$, the claim is obvious, as $h\circ\rho$ is constant and ν^h vanishes. For $m\notin\mathbb{L}$, the claim follows from Corollary 4.15. ☐

4.7. Perturbing to achieve nondegeneracy

Now let us suppose that for some $\epsilon > 0$, we have that

- $h(\rho)$ is constant for $\rho \leq \epsilon$;
- $h(\rho)$ is linear for $\rho \geq 1 - \epsilon$;
- On any interval on which $h(\rho)$ is linear, except $(-\infty, \epsilon]$, the slope is not a Reeb period of Y.

Then the orbits of $h \circ \rho$ come in families parametrized by manifolds with corners.

The families are indexed by a set

$$P = \coprod_{I \subset [N]} P_I,$$

where P_I consists of families of orbits γ with $I(\gamma) = I$. The two cases $I = \varnothing, I \neq \varnothing$ must be treated differently. To describe $P_\varnothing$, let us suppose that ϵ' is maximal so that h is constant on $(-\infty, \epsilon']$. Then

$$P_\varnothing = \{0\} \cup \{\rho > \epsilon' : h'(\rho) = 0\}.$$

Associated with $p \in P_\varnothing$ is a set of constant orbits C_p, which can be identified with a subset of M:

$$C_0 = \{\rho \leq \epsilon'\}, \qquad C_p = \{\rho = p\} \quad \text{for } p \in P_\varnothing \backslash \{0\}.$$

On the other hand, for $I \neq \varnothing$ we have

$$\begin{aligned} P_I = \{p \in \operatorname{im}(r_I^{\max}) :\ & \text{for each } i \in I \text{ we either have} \\ & p_i = 0 \text{ or } h'(\tilde{\rho}_I(p)) \cdot \tilde{\nu}_{I,i}(p) \in \mathbb{Z} \backslash \{0\}\}. \end{aligned}$$

Associated with each $p \in P_I$, we define a subset of M:

$$C_p := \{m \in UD_I^{\max} : r_I^{\max}(m) = p, \nu_k(m) = 0 \text{ for } k \notin I\}.$$

For each $p \in P$, C_p is a manifold-with-corners on which the flow of $h \circ \rho$ is 1-periodic, yielding a manifold-with-corners of orbits which is diffeomorphic to C_p.

We now perturb $h \circ \rho$, in such a way as to make the orbits nondegenerate.

Lemma 4.17. *Given $\epsilon > 0$, there exists a perturbation H of $h \circ \rho$ with non-degenerate orbits, such that for any capped orbit (γ, u) of H, there exists a capped orbit $(\bar{\gamma}, \bar{u})$ of $h \circ \rho$, such that*

1. $|\mathcal{A}(\gamma, u) - \mathcal{A}(\bar{\gamma}, \bar{u})| < \epsilon$;
2. $|\operatorname{CZ}(\gamma, u) - \operatorname{CZ}(\bar{\gamma}, \bar{u})| \leq k(\bar{\gamma})/2$, *where* CZ *denotes the Conley–Zehnder index,*[14] *and* $k(\bar{\gamma}) := \dim\ker \left(D\Phi^1_{h\circ\rho} - \mathrm{id} \right)_{\bar{\gamma}(0)}$.

Proof. Note that the subsets $C_p \subset M$, $p \in P$ are closed, disjoint, preserved by the flow of $h \circ \rho$, and the flow is one-periodic on them. We will choose disjoint neighbourhoods N_p of C_p, and perturb in each N_p separately, i.e., $H = h \circ \rho + \delta$ where $\delta = \sum_p \delta_p$ with δ_p supported in N_p.

[14] The definition is due to Robbin–Salamon in the case of the possibly-degenerate orbit γ.

We fix a Riemannian metric on M for the duration of this proof. In particular, whenever we say that a function is 'C^k-small', we mean with respect to this metric.

Note that $d(\Phi_1^{h\circ\rho}(m), m) > \eta$ for some $\eta > 0$ over the compact set $M \setminus \cup_p N_p$. By making δ C^1-small, we can make $\Phi_t^{h\circ\rho+\delta}$ C^0-close to $\Phi_t^{h\circ\rho}$ for all $t \in [0,1]$; in particular we can ensure that all fixed points of $\Phi_1^{h\circ\rho+\delta}$ lie in some N_p. By taking a generic such δ, we can ensure that all orbits of $h\circ\rho+\delta$ are nondegenerate. By taking N_p small, we may ensure that any orbit γ of $h\circ\rho+\delta$ is C^0-close to an orbit $\bar{\gamma}$ of $h\circ\rho$. When the orbits are sufficiently C^0-close, we can construct a cylinder $v : S^1 \times [0,1] \to M$ stretching between γ and $\bar{\gamma}$, so that $v(\cdot, t)$ is the unique geodesic from $\gamma(t)$ to $\bar{\gamma}(t)$; concatenating with this cylinder defines a natural bijection between caps for γ and $\bar{\gamma}$. To arrange (1) we must bound the symplectic area of the cylinder. This is achieved by observing that

$$\int_{S^1\times[0,1]} v^*\omega = \int_{S^1\times[0,1]} \omega\left(\frac{\partial v}{\partial s}, \frac{\partial v}{\partial t}\right),$$

and $\partial v/\partial s$ can be made arbitrarily small while $\partial v/\partial t$ is bounded.

Now we arrange (2). Recall that $\mathrm{CZ}(\bar{\gamma}, \bar{u})$ is by definition that Conley–Zehnder index of the path of symplectic matrices $\Psi_t(D\Phi_t^{h\circ\rho})\Psi_t^{-1}$, where Ψ_t is a trivialization of $\bar{\gamma}^*TM$ induced by the cap $\bar{u}$ and $\mathrm{CZ}(\gamma, u)$ is the Conley–Zehnder index of the corresponding path of symplectic matrices. By making δ C^2-small, we can make $\Phi_t^{h\circ\rho+\delta}$ C^1-close to $\Phi_t^{h\circ\rho}$ for all $t \in [0,1]$;[15] this implies that the aforementioned paths of symplectic matrices can be made C^0-close; the result now follows by [24, Corollary 4.9]. $\square$

Remark 4.18. Our approach to perturbing degenerate orbits follows [24]. With more effort one can prove a more precise result: one can find a Morse–Bott perturbation H, whose orbits are precisely the orbits of $h\circ\rho$ corresponding to critical points of a Morse function defined on the manifold with corners (and increasing at the boundary), and are nondegenerate. The technique for doing this goes back to [6, Proposition 2.2], see also [26, Section 3.3] and [21]. These references all deal with closed manifolds of orbits; the case of manifolds with corners is addressed in [14], in a setting closely related to ours.

4.8. Action computation

We start with a preliminary lemma which will be used in our action computation below. We state this lemma in a much more general setup than we need and after the proof make some comments to explain how we will specialize it.

Lemma 4.19. *Let (M, ω) be a symplectic manifold and $\pi : M \to \mathbb{R}^k$ be a smooth map. Let $f : \mathbb{R}^k \to \mathbb{R}$ be a smooth function. Let ϕ_t be the Hamiltonian*

[15] This means that given $\eta > 0$, we may choose δ so that for all $(m, v) \in TM$ with $|v| \leq 1$, we have

$$d\left(D\Phi_t^{h\circ\rho+\delta}(m, v), D\Phi_t^{h\circ\rho}(m, v)\right) < \eta$$

for an a priori fixed Riemannian metric on TM.

flow of $\tilde{f} := \pi^* f$*. Consider a map*

$$u : [0,1] \times [0,1] \to M$$

such that for all $(t,s) \in [0,1] \times [0,1]$*,*

$$u(t,s) = \phi_t(u(0,s)).$$

Moreover, we assume that $u([0,1] \times \{0\}) = \{A\}$ *and* $u([0,1] \times \{1\}) = \{B\}$*, where* A *and* B *are points in* $\mathbb{R}^k$*. We orient* $[0,1] \times [0,1]$ *is so that* ∂_t, ∂_s *is a positive basis.*

Then, the symplectic area of u *is equal to* $f(B) - f(A)$*.*

Proof. This is an elementary computation.

$$\begin{aligned}
\int_{[0,1]\times[0,1]} u^*\omega &= \int_0^1 \int_0^1 \omega(u_*\partial_t, u_*\partial_s)\mathrm{d}s\mathrm{d}t \\
&= \int_0^1 \int_0^1 \omega(X_{\tilde{f}}, u_*\partial_s)\mathrm{d}s\mathrm{d}t \\
&= \int_0^1 \left(\int_0^1 \mathrm{d}f(\pi_* u_* \partial_s)\mathrm{d}s \right) \mathrm{d}t \\
&= \int_0^1 \left(\int_{\{t\}\times[0,1]} (\pi \circ u \circ \iota_t)^* \mathrm{d}f \right) \mathrm{d}t \\
&= \int_0^1 (f(B) - f(A))\, \mathrm{d}t \\
&= f(B) - f(A)
\end{aligned}$$

as required. □

Note that the assumption on the boundary of u is automatic if π is involutive; even more specifically, when π is a moment map for a Hamiltonian torus action. Also note that if f is an affine function, then $f(B) - f(A)$ is equal to the linear part of f evaluated at the vector $\overrightarrow{AB}$ considered as an element of $\mathbb{R}^k$. If π is a moment map for a Hamiltonian $(\mathbb{R}/\mathbb{Z})^k$-action, and f is integral affine, then u as in the statement of the lemma satisfies

$$u(0,s) = u(1,s), \quad \text{for all } s \in [0,1].$$

We will only use this special case of the lemma below, where u can also be thought of as a map $\mathbb{R}/\mathbb{Z} \times [0,1] \to M$. As a final remark that will be relevant, note that the blow down map

$$\mathbb{R}/\mathbb{Z} \times [0,1] \to \mathbb{D} \subset \mathbb{C}, \quad \text{where } (t,s) \mapsto s e^{2\pi i t}$$

is orientation reversing, where we use the standard orientation of $\mathbb{C}$.

Let us now get back to the action computation that we wanted to undertake, continuing the notation used in the previous section.

There is a canonical cap u_{out} associated with any orbit γ of $h \circ \rho$, which we now describe. If $I(\gamma) = \emptyset$, then γ is a constant orbit. We define u_{out} to be the constant cap in this case. Otherwise, γ is contained in $UD^{\max}_{I(\gamma)}$. If γ is contained in $UD_{I(\gamma)}$ then it is contained in an admissible standard chart, and we define u_{out} to be the cap contained in that chart. Note that u_{out} is well-defined by Lemma 2.14.

Note that if γ is an orbit on D, it is contained in $UD_{I(\gamma)}$. For an orbit γ not contained in D, we define u_{out} to be the union of the cylinder swept by γ along the Liouville flow taking it into $UD_{I(\gamma)}$, with the canonical cap in an admissible chart.

At this point the reader might also benefit from looking at Remark 2.17, which gives a simpler version of admissibility and suffices for the purposes of this paper. It works because of the following Lemma.

Lemma 4.20. *The action of the* 1*-periodic orbit* γ *of* $h \circ \rho$ *with respect to the outer cap is given by*

$$\mathcal{A}(\gamma, u_{\text{out}}) = h(\rho(\gamma)) + \sum_i \nu_i^h(\gamma) \cdot r_i^{\max}(\gamma).$$

Proof. The action is

$$\mathcal{A}(\gamma, u_{\text{out}}) = \int_{S^1} h(\rho(\gamma(t)) + \int_{u_{\text{out}}} \omega.$$

The first term is $h(\rho(\gamma))$, because $h(\rho(\gamma(t)) = h(\rho(\gamma))$ is constant along γ. We claim that the second term is

$$\omega(u_{\text{out}}) = \sum_i \nu_i^h(\gamma) \cdot r_i^{\max}(\gamma).$$

Consider the map

$$f : \mathbb{R}^I \to \mathbb{R}$$
$$f(r) = \sum_{i \in I} -\nu_i^h(\gamma) \cdot r_i.$$

Notice that γ is a one periodic orbit of the Hamiltonian vector field of $\tilde{f} := f \circ r_I^{\max}$ (see Lemma 4.14).

We break u_{out} into two pieces: the piece $u_{\text{out},1}$ lying in an admissible chart, and the piece $u_{\text{out},2} = \cup_{t \in [0,T]} \varphi_t(\gamma)$ swept out by the Liouville flow. Assume that the boundary of $u_{\text{out},1}$ is contained in $r_I^{-1}((a_i)_{i \in I})$.

Using the symplectic embedding of the admissible chart into $\mathbb{C}^I \times \mathbb{C}^{n-|I|}$, we see that $\int_{u_{\text{out},1}} \omega$ is equal to the symplectic area of an arbitrary cap of a 1-periodic orbit of $X_{\tilde{f}}$ contained inside the fiber above $(a_i)_{i \in I}$ of the moment map $\mathbb{C}^I \times \mathbb{C}^{n-|I|} \to \mathbb{R}^I$. Choosing the cap obtained by radially scaling the loop to the origin inside the slice $\mathbb{C}^I \times \{c\}$ that it is contained in, we immediately obtain (e.g. using Lemma 4.19):

$$\int_{u_{\text{out},1}} \omega = -\left(\sum_{i \in I} -\nu_i^h(\gamma) \cdot a_i \right).$$

For the area of the second piece, we use Lemma 4.19 for the map $r_I^{\max}$, function f and map $u_{\text{out},2}$ to obtain:

$$\int_{u_{\text{out},2}} \omega = \sum_{i \in I} \nu_i^h(\gamma) \cdot (r_i^{\max}(\gamma) - a_i).$$

Note that here we used the $r_I^{\max}$-relatedness of the Liouville vector field and the Euler vector field (i.e., Lemma 4.4).

Putting the computations together, we get the desired result. $\square$

We define the fractional inner cap $u_{\text{in}} := u_{\text{out}} - \nu^h(\gamma) \cdot \boldsymbol{\lambda}$ as in Sect. 3.3. Strictly speaking we do not need the following result for our argument, but we thought it was informative. Note that it is a slight generalization of the well-known formula in [45, Section 1.2], which gives the result for SH-type orbits.

Lemma 4.21. *The action of the orbit γ of $h \circ \rho$ with respect to the inner cap is given by*

$$\mathcal{A}(\gamma, u_{\text{in}}) = h(\rho(\gamma)) - h'(\rho(\gamma)) \cdot \rho(\gamma).$$

Proof. By Lemma 4.20, setting $\rho = \rho(\gamma)$, we have

$$\begin{aligned}
\mathcal{A}(\gamma, u_{\text{in}}) &= h(\rho) + \sum_i \nu_i^h(\gamma) \cdot r_i^{\max}(\gamma) - \nu^h(\gamma) \cdot \boldsymbol{\lambda} \\
&= h(\rho) - h'(\rho) \sum_i \nu_i(\gamma) \cdot (r_i(\gamma) - \lambda_i) \\
&= h(\rho) - h'(\rho) \cdot \tilde{Z}_I \left(\tilde{\rho}_I\right)_{r_I^{\max}(\gamma)} \\
&= h(\rho) - h'(\rho) \cdot \rho,
\end{aligned}$$

where the last step follows as $\tilde{Z}_I(\tilde{\rho}_I) = \tilde{\rho}_I$ and $\tilde{\rho}_I \circ r_I^{\max} = \rho$. $\square$

4.9. Index computation

Lemma 4.22. *Let γ be an orbit of $h \circ \rho$, with $J := \{j \in I(\gamma) : r_j^{\max}(\gamma) \neq 0\}$. Define the $|J| \times |J|$ matrix*

$$Hess_\gamma := \left(\frac{\partial^2 (h \circ \tilde{\rho}_I)}{\partial r_i \partial r_j} (r_I(\gamma)) \right)_{i,j \in J}.$$

Then the Conley–Zehnder index of the orbit γ of $h \circ \rho$ with respect to the outer cap is given by

$$\mathrm{CZ}(\gamma, u_{\text{out}}) = 2 \sum_i \left\lceil \nu_i^h(\gamma) \right\rceil + \frac{1}{2} \mathrm{sign}\left(Hess_\gamma \right).$$

Proof. For constant orbits the result is easy, so we assume that γ is nonconstant. We may assume that γ and u_{out} lie in an admissible chart $\mathbb{C}^{I(\gamma)} \times \mathbb{C}^{n-|I(\gamma)|}$, as the index does not change as we flow along the Liouville flow. The flow of $h \circ \rho$ in the admissible chart decomposes as a product of the flow

$$\varphi_t(r, \theta) = (r, \theta + 2\pi t \tilde{\nu}^h(r))$$

on $\mathbb{C}^{I(\gamma)}$ (written in action-angle coordinates) with the trivial flow on $\mathbb{C}^{n-|I(\gamma)|}$. Thus $\mathrm{CZ}(\gamma, u^{\text{out}}) = \mathrm{CZ}(D\varphi_t)$. We have

$$\begin{aligned}
\mathrm{CZ}(D\varphi_t) &= \mathrm{CZ}\left(diag\left(e^{2\pi i t \cdot \nu^h(z)}\right) \cdot \left(\mathbf{1} + 2\pi i t \cdot \left[\frac{\partial \nu_j^h(z)}{\partial z_i} \right] \right) \right) \\
&= \mathrm{CZ}\left(diag\left(e^{2\pi i t \cdot \nu^h(z)}\right)\right)
\end{aligned}$$

$$+\,\mathrm{CZ}\left(diag\left(e^{2\pi i\cdot\nu^h(z)}\right)\cdot\left(\mathbf{1}+2\pi it\cdot\left[\frac{\partial\nu_j^h(z)}{\partial z_i}\right]\right)\right)$$

by a standard argument (c.f. [26, Section 3.3]). The first term is equal to $2\sum_i\left\lceil\nu_i^h(\gamma)\right\rceil$ (see [26, Section 3.2]). For the second, we decompose $\mathbb{C}^{I(\gamma)}=\mathbb{C}^J\oplus\mathbb{C}^{I(\gamma)\setminus J}$. Note that $\partial\nu_j^h/\partial z_i=0$ for $i\notin J$, because r_i has vanishing derivative along $\{z_i=0\}$, where our orbit is contained. Also note that $e^{2\pi i\cdot\nu_i^h(z)}=1$ for $i\in J$. Putting these together, one finds that the second term is equal to the Conley–Zehnder index of the path $\mathbf{1}_J+2\pi it\cdot[\partial\nu_j^h/\partial z_i]_{i,j\in J}$. Writing this in the basis given by action-angle coordinates (i.e., $(r_i\partial/\partial r_i,\partial/\partial\theta_i)_{i\in J}$), we see that it takes the form of a symplectic shear, whose Conley–Zehnder index is equal to

$$\mathrm{CZ}\begin{pmatrix}\mathbf{1} & -2\pi t\cdot Hess_\gamma\\ 0 & \mathbf{1}\end{pmatrix}=\frac{1}{2}\mathrm{sign}\left(Hess_\gamma\right)$$

by the 'normalization' property of the Conley–Zehnder index, see [30, Theorem 4.1].[16] □

Lemma 4.23. *Let γ be an orbit of H which corresponds to an orbit $\bar\gamma$ of $h\circ\rho$ as in Lemma 4.17. Then we have*

$$i(\gamma,u_{\mathrm{out}})=2\sum_i\left\lceil\nu_i^h(\bar\gamma)\right\rceil+\delta(\gamma),$$

where $0\le\delta(\gamma)\le 2n$.

Proof. We apply Lemmas 4.17 and 4.22. Continuing the notation from the proof of the latter, we have

$$\ker\left(D\varphi_1-\mathrm{id}\right)=\mathbb{C}^{n-|I(\gamma)|}\oplus\mathbb{C}^{I(\gamma)\setminus J}\oplus\langle\partial/\partial\theta_j\rangle_{j\in J}\oplus\ker\left(Hess_{\bar\gamma}\right).$$

Recall that $k(\bar\gamma)$ is, by definition, the dimension of this space. Thus we have

$$k(\bar\gamma)+|\mathrm{sign}\left(Hess_{\bar\gamma}\right)|\le 2n.$$

Combining the stated Lemmas, we have

$$\left|\mathrm{CZ}(\gamma,u^{\mathrm{out}})-2\sum_i\left\lceil\nu_i^h(\bar\gamma)\right\rceil-\frac{1}{2}\mathrm{sign}\left(Hess_{\bar\gamma}\right)\right|\le\frac{k(\bar\gamma)}{2}$$

$$\Rightarrow\left|\mathrm{CZ}(\gamma,u^{\mathrm{out}})-2\sum_i\left\lceil\nu_i^h(\bar\gamma)\right\rceil\right|\le\frac{2n}{2}=n.$$

Recalling that $i(\gamma,u^{\mathrm{out}}):=n+\mathrm{CZ}(\gamma,u^{\mathrm{out}})$, the result is immediate. □

Lemma 4.24. *Let γ be an orbit of H which corresponds to an orbit $\bar\gamma$ of $h\circ\rho$ as in Lemma 4.17, and suppose that $h'(\rho)\ge 0$ everywhere, so that $\nu_i^h(\bar\gamma)\ge 0$ for all i. Then we have*

$$i(\gamma,u_{\mathrm{in}})\ge\sum_i(2-\lambda_i)\cdot\nu_i^h(\bar\gamma).$$

In particular, when Hypothesis A is satisfied, we have $i(\gamma,u_{\mathrm{in}})\ge 0$.

[16] The signature of a symmetric matrix is the number of positive eigenvalues minus the number of negative eigenvalues.

Proof. By Lemma 4.23, we have

$$\begin{aligned} i(\gamma, u_{\text{in}}) &\geq \sum_i 2\left\lceil \nu_i^h(\bar{\gamma})\right\rceil - \lambda_i \cdot \nu_i^h(\bar{\gamma}) \\ &\geq \sum_i (2-\lambda_i)\cdot \nu_i^h(\bar{\gamma}) \end{aligned}$$

as required. □

Lemma 4.25. *Let γ be an orbit of H which corresponds to an orbit $\bar{\gamma}$ of $h\circ\rho$ as in Lemma 4.17. Then we have*

$$i_{\text{mix}}(\gamma) = \sum_i (2-\kappa^{-1}r_i^{\max}(\bar{\gamma}))\cdot \nu_i^h(\bar{\gamma}) - \kappa^{-1}h(\rho(\bar{\gamma})) + D(\gamma),$$

where $D(\gamma)$ is bounded: in particular, the lower bound is $D(\gamma) \geq -\kappa^{-1}\epsilon(\gamma)$, where $\epsilon(\gamma)$ is as in Lemma 4.17.

Proof. The equality follows using the outer cap to compute the mixed index, via Lemmas 4.17, 4.20, and 4.23. □

5. Proofs

In this section, we prove Theorems B, C, and D. We will assume throughout that the divisor D is orthogonal, although that is not a hypothesis of Theorems B and C; the general results follow using Remark 2.2.

Because D is orthogonal (and in particular admits an admissible system of commuting Hamiltonians), we can make all of the constructions from the previous section, whose notation and assumptions (e.g. Equation 4.1) we continue. Right before Sect. 4.6 we had started omitting the dependence on $R \in (0, R_0)$ from the notation for brevity, now we bring it back.

5.1. Properties of $\tilde{\rho}_I^R$

When we talk about a property (n) of $\tilde{Y}_I^R$ below, we mean the properties from Lemma 4.8.

Lemma 5.1. *There is a continuous function $\epsilon_1 : [0,, R_0) \to \mathbb{R}_{\geq 0}$, with $\epsilon_1(0) = 0$, such that for all $R \in (0, R_0)$, all I, and all $r \in \tilde{Y}_I^0$, we have*

$$1 \leq \tilde{\rho}_I^R(r) \leq 1 + \epsilon_1(R).$$

Proof. Note that $\tilde{Y}_I^R$ is sandwiched between $\tilde{Y}_I^0$ and $(R/2, \ldots, R/2) + \tilde{Y}_I^0$; hence, it is also sandwiched between $\tilde{Y}_I^0$ and $\alpha\cdot\kappa\boldsymbol{\lambda} + \tilde{Y}_I^0$, where $\alpha = R/(2\kappa \min \lambda_i)$. It follows that

$$1 \leq \tilde{\rho}_I^R(r) \leq \frac{1}{1-\alpha}$$

for $r \in \tilde{Y}_I^0$, which gives the desired result. □

Lemma 5.2. *There is a continuous function $\epsilon_2 : [0, R_0) \to \mathbb{R}_{\geq 0}$, with $\epsilon_2(0) = 0$, such that for all $R \in (0, R_0)$, all I, and all $r \in \tilde{Y}_I^0$, we have*

$$1 \leq \sum_i \kappa\lambda_i \cdot \tilde{\nu}_{I,i}^R(r) \leq 1 + \epsilon_2(R). \tag{5.1}$$

Proof. Because $\tilde{Z}_I\left(\tilde{\rho}_I^R\right)=\tilde{\rho}_I^R$ by construction, we have

$$\sum_i(\kappa\lambda_i-r_i)\cdot\tilde{\nu}_{I,i}^R(r)=\tilde{\rho}_I^R(r). \tag{5.2}$$

Thus Lemma 5.1 gives

$$1\le\tilde{\rho}_I^R(r)=\sum_i(\kappa\lambda_i-r_i)\cdot\tilde{\nu}_{I,i}^R(r)\le\sum_i\kappa\lambda_i\cdot\tilde{\nu}_{I,i}^R(r),$$

where the last step uses the fact that $\tilde{\nu}_{I,i}^R(r)\ge 0$ by property (3) of $\tilde{Y}_I^R$, and $r_i\ge 0$ for all i.

For the right-hand bound, observe that $\tilde{\nu}_{I,i}^R(r)=0$ whenever $r_i>R/2$, by property (5) of $\tilde{Y}_I^R$; as $\tilde{\nu}_{I,i}^R\ge 0$ this implies that

$$\sum_i(\kappa\lambda_i-R/2)\cdot\tilde{\nu}_{I,i}^R(r)\le\sum_i(\kappa\lambda_i-r_i)\cdot\tilde{\nu}_{I,i}^R(r)=\tilde{\rho}_I^R(r)\le 1+\epsilon_1(R).$$

Thus we have

$$\sum_i\kappa\lambda_i\cdot\tilde{\nu}_{I,i}^R(r)\le\left(\max_i\frac{\kappa\lambda_i}{\kappa\lambda_i-R/2}\right)\cdot(1+\epsilon_1(R))$$

where the RHS converges to 1 as $R\to 0$, as required. □

The following Lemma will be used in the proof of Theorem B:

Lemma 5.3. *There exists a continuous function $\sigma_{\text{crit}}^B:[0,R_0)\to\mathbb{R}_{\ge 0}$, with $\sigma_{\text{crit}}^B(0)=\sigma_{\text{crit}}$ (recall Definition 1.9), such that for all $R\in(0,R_0)$, all I, and all $r\in\tilde{Y}_I^0$, we have*

$$\sum_i\left(2-\kappa^{-1}r_i\right)\cdot\tilde{\nu}_{I,i}^R(r)-\kappa^{-1}\left(\tilde{\rho}_I^R(r)-\sigma_{\text{crit}}^B(R)\right)>0.$$

Proof. Note that by property (4) of $\tilde{Y}_I^R$, if $\tilde{\nu}_{I,i}^R(r)\ne 0$ and $r\in\tilde{Y}_I^0$ then $r_i\le R/2$. Combining this observation with Lemma 5.1, we have

$$\sum_i\left(2-\kappa^{-1}r_i\right)\cdot\tilde{\nu}_{I,i}^R(r)-\kappa^{-1}\tilde{\rho}_I^R(r)\ge\sum_i\left(2-\frac{R}{2\kappa}\right)\cdot\tilde{\nu}_{I,i}^R(r)-\kappa^{-1}\cdot(1+\epsilon_1(R)).$$

Dividing the left-hand bound in (5.1) by $\max_i\kappa\lambda_i$ immediately gives

$$\sum_i\left(2-\frac{R}{2\kappa}\right)\cdot\tilde{\nu}_{I,i}^R(r)\ge\frac{2-\frac{R}{2\kappa}}{\max_i\kappa\lambda_i}.$$

Thus, we may take

$$\tilde{\sigma}_{\text{crit}}^B(R)=1+\epsilon_1(R)-\frac{2-\frac{R}{2\kappa}}{\max_i\lambda_i},\qquad\sigma_{\text{crit}}^B(R)=\max\left(0,\tilde{\sigma}_{\text{crit}}^B(R)\right),$$

which clearly has the desired properties. □

The following Lemma will be used in the proof of Theorem D:

Lemma 5.4. *There exists a continuous function* $\sigma^D_{\text{crit}} : [0, R_0) \to \mathbb{R}_{\geq 0}$, *with* $\sigma^D_{\text{crit}}(0) = \sigma_{\text{crit}}$, *and a positive function* $\eta : (0, R_0) \to \mathbb{R}_{>0}$, *such that for all* $R \in (0, R_0)$, *all* I, *all* $i \in I$, *and all* $r \in V_I$ *satisfying* $\tilde{\rho}^R_I(r) > \sigma^D_{\text{crit}}(R)$ *and* $\tilde{\nu}^R_{I,i}(r) \neq 0$, *we have*

$$2 - \kappa^{-1} r_i \geq \eta(R) \cdot \left(\tilde{\rho}^R_I(r) - \sigma^D_{\text{crit}}(R)\right).$$

Proof. Suppose that the flowline of $\tilde{Z}_I$ passing through r exits $\tilde{Y}^0_I$ at r'. Because both $\tilde{\rho}^R_I$ and $r_i - \kappa\lambda_i$ vary linearly along flowlines of $\tilde{Z}_I$, we have

$$\frac{\tilde{\rho}^R_I(r)}{\tilde{\rho}^R_I(r')} = \frac{r_i - \kappa\lambda_i}{r'_i - \kappa\lambda_i},$$

and, therefore,

$$2 - \kappa^{-1} r_i = 2 - \lambda_i + \frac{\tilde{\rho}^R_I(r)}{\tilde{\rho}^R_I(r')} \cdot \left(\lambda_i - \frac{r'_i}{\kappa}\right).$$

Now by property (5) of $\tilde{Y}^I_R$, if $\tilde{\nu}^R_{I,i}(r) \neq 0$ then r lies in the region $P_I^{-1}(\{r_i \leq R/2\})$, and therefore $r'_i \leq R/2$. We also have $\tilde{\rho}^R_I(r') \leq 1 + \epsilon_1(R)$ by Lemma 5.1. It follows that

$$2 - \kappa^{-1} r_i \geq 2 - \lambda_i + \frac{\tilde{\rho}^R_I(r) \cdot \left(\lambda_i - \frac{R}{2\kappa}\right)}{1 + \epsilon_1(R)}.$$

Now let us set

$$\tilde{\sigma}^D_{\text{crit}}(R) = \max_i \frac{(\lambda_i - 2) \cdot (1 + \epsilon_1(R))}{\lambda_i - \frac{R}{2\kappa}};$$

then we find that the functions

$$\sigma^D_{\text{crit}}(R) = \max\left(0, \tilde{\sigma}^D_{\text{crit}}(R)\right),$$
$$\eta(R) = \min_i \frac{\lambda_i - \frac{R}{2\kappa}}{1 + \epsilon_1(R)}$$

have the desired properties. □

5.2. Proof of Theorem B

Let $R \in (0, R_0)$ be sufficiently small that $\sigma^B_{\text{crit}}(R) < 1$. Let $\sigma = \sigma^B_{\text{crit}}(R) + 2\delta < 1$, for some $\delta > 0$. The proof will rely on a special choice of acceleration data for K^R_σ (see Definition 4.13) which we now describe. Fix $0 < \ell_1 < \ell_2 < \cdots$ such that the Reeb flow on $Y^R_\sigma = \partial K^R_\sigma$ has no ℓ_n-periodic orbits for all n, and $\ell_n \to \infty$ as $n \to \infty$. (Here we take the contact form from Definition 4.13.)

We now choose smooth functions $h_n : \mathbb{R} \to \mathbb{R}$ approximating max $(0, \ell_n(\rho - \sigma))$. We require that they each satisfy the conditions from Sect. 4.7, and furthermore,

- $h_1 < h_2 < \cdots$ (pointwise);
- $h'_n(\rho) \geq 0$;
- $h_n(\rho) = 2\delta_n$ for $\rho \leq \sigma/2$;
- $h_n(\rho) = \ell_n(\rho - \sigma) + \delta_n$ for $\rho \geq \sigma$,

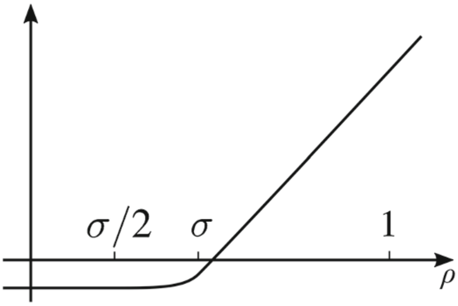


FIGURE 2. The function h_n

where $\delta_n < 0$ converges monotonically to 0 as $n \to \infty$, and furthermore $\ell_n\sigma - \delta_n < \ell_{n+1}\sigma - \delta_{n+1}$ for all n (the latter condition will be used in the proof of Proposition 5.10). See Fig. 2. Note that h_n converges monotonically to 0 on $(-\infty, \sigma]$ and $+\infty$ outside it. We extend $(h_n)_{n\in\mathbb{Z}_{\geq 1}}$ to $(h_\tau)_{\tau\in[1,\infty)}$ by convex interpolation: $h_\tau = (n+1-\tau)h_n + (\tau - n)h_{n+1}$ for $\tau \in [n, n+1]$. We choose our acceleration data (H_τ, J_τ) for $K_\sigma^R \subset M$, where H_n is a perturbation of $h_n \circ \rho^R$ as in Lemma 4.17, where the parameter ϵ in the Lemma is chosen smaller than $\ell_n\delta$, and H_τ is a corresponding perturbation of $h_\tau \circ \rho^R$. We further require that, over a 'neck region' $\{\sigma \leq \rho^R \leq \sigma + \epsilon\}$ (where $\sigma < \sigma + \epsilon < 1$), we have $H_\tau = h_\tau \circ \rho^R$ and J_τ is of contact type.

We denote by $\mathcal{C} = \mathcal{C}(H_\tau, J_\tau)$ the corresponding Floer 1-ray

$$CF^*(M, H_1; \Lambda) \to CF^*(M, H_2; \Lambda) \to \cdots,$$

so that $SC^*_M(K_\sigma^R; \Lambda) = \widehat{tel}(\mathcal{C})$. By restricting (H_τ, J_τ) to K_σ^R, we obtain acceleration data appropriate for defining the symplectic cochain complex of K_σ^R. We denote by $\mathcal{C}_{SH} := \mathcal{C}(H_\tau|_{K_\sigma^R}, J_\tau|_{K_\sigma^R})$ the corresponding Floer 1-ray

$$CF^*(K_\sigma^R, H_1|_{K_\sigma^R}; \Bbbk) \to CF^*(K_\sigma^R, H_2|_{K_\sigma^R}; \Bbbk) \to \cdots,$$

so that $SC^*(K_\sigma^R; \Bbbk) = tel(\mathcal{C}_{SH})$.

By construction, the orbits of H_n are either contained in K_σ^R (in which case we say they are of SH-type), or contained in $M \backslash K_\sigma^R$ (in which case we say they are of D-type). We have a corresponding direct sum decomposition of Λ-modules:

$$tel(\mathcal{C}) = tel(\mathcal{C})_{SH} \oplus tel(\mathcal{C})_D.$$

Let us denote $SC_\Lambda := (SC^*(K_\sigma^R; \Bbbk) \otimes_\Bbbk \Lambda, d \otimes id_\Lambda)$. We have the isomorphism

$$\iota : SC_\Lambda \to tel(\mathcal{C})_{SH},$$
$$\iota(\gamma \otimes e^a) = (\gamma, a \cdot u_{\text{in}}).$$

By Lemma 3.8, this map respects action and index.

Lemma 5.5. *If γ is a D-type orbit of H_n, then $i_{\text{mix}}(\gamma) \geq \kappa^{-1}\delta\ell_n$.*

Proof. By Lemma 4.25, we have

$$i_{\mathrm{mix}}(\gamma) \geq \sum_i \left(2 - \kappa^{-1} r_i^{\max}(\bar{\gamma})\right) \cdot \nu_i^h(\bar{\gamma}) - \kappa^{-1}\left(h\left(\rho^R(\bar{\gamma})\right) + \epsilon(\gamma)\right).$$

Note that as γ is a D-type orbit, we have $\rho^R(\bar{\gamma}) \geq 1$ by Lemma 5.1, and therefore,

$$h\left(\rho^R(\bar{\gamma})\right) \leq \ell_n\left(\rho^R(\bar{\gamma}) - \sigma\right) \qquad \text{and}$$
$$h'\left(\rho^R(\bar{\gamma})\right) = \ell_n.$$

Thus, we have $\nu_i^h(\bar{\gamma}) = \ell_n \cdot \tilde{\nu}_{I,i}^R(r_I^{\max}(\bar{\gamma}))$. Setting $r = r_I^{\max}(\bar{\gamma})$ (which lies in $\tilde{Y}_I^0$ because $\bar{\gamma}$ is a D-type orbit), and recalling that we chose $\epsilon(\gamma) < \ell_n\delta$, we obtain

$$\begin{aligned} i_{\mathrm{mix}}(\gamma) &\geq \ell_n \cdot \sum_i (2 - \kappa^{-1} r_i) \cdot \tilde{\nu}_{I,i}^R(r) - \kappa^{-1}\ell_n(\tilde{\rho}_I^R(r) - \sigma) - \kappa^{-1}\ell_n\delta \\ &\geq \kappa^{-1}\ell_n \cdot \left(\sigma - \sigma_{\mathrm{crit}}^B(R) - \delta\right) = \kappa^{-1}\delta\ell_n, \end{aligned}$$

where the second inequality follows from Lemma 5.3. □

We now consider the filtration on $tel(\mathcal{C})$ associated with the filtration map

$$\mathcal{F}'(\gamma, u) := \frac{\mathcal{A}(\gamma, u) + \delta\ell_n}{\kappa},$$

if γ is a 1-periodic orbit of H_n. It is clear that this is a filtration map, because the differential increases action, and it also increases n and hence ℓ_n by the definition of the telescope complex. We define the corresponding filtration on SC_Λ, associated with the filtration map

$$\mathcal{F}(\gamma \otimes e^a) := \frac{\mathcal{A}(\gamma) + \kappa a + \delta\ell_n}{\kappa}.$$

(Note that because ι respects index and action, we have $\mathcal{F}' \circ \iota = \mathcal{F}$.) For any cochain complex C^*, we define the quotient complex $\sigma_{<p}C^* := \oplus_{*<p}C^*$ with the induced differential.

Lemma 5.6. *For any p, ι induces an isomorphism of graded $\mathcal{Q}_{\geq 0}\Lambda$-modules*

$$\iota : \sigma_{<p}\mathcal{F}_{\geq p}SC_\Lambda \xrightarrow{\sim} \sigma_{<p}\mathcal{F}'_{\geq p}tel(\mathcal{C}).$$

Proof. It suffices to show that $\sigma_{<p}\mathcal{F}'_{\geq p}tel(\mathcal{C})$ does not include any D-type orbits. Indeed, for any generator (γ, u) of this complex, we have

$$i(\gamma, u) < p \leq \frac{\mathcal{A}(\gamma, u) + \delta\ell_n}{\kappa},$$

which means γ cannot be of D-type, by Lemma 5.5. □

We now recall that SC_Λ comes equipped with the $\mathcal{Q}$-filtration, induced by the $\mathcal{Q}$-filtration on Λ (c.f. Eq. (1.1)).

Lemma 5.7. *For any p, the inclusions of the following subcomplexes are quasi-isomorphisms:*

$$\mathcal{F}'_{\geq p}tel(\mathcal{C}) \subset tel(\mathcal{C}), \qquad \text{and} \qquad \mathcal{F}_{\geq p}\mathcal{Q}_{\geq q}SC_\Lambda \subset \mathcal{Q}_{\geq q}SC_\Lambda$$

(the latter for any $q \in \mathbb{Z} \cup \{-\infty\}$).

Proof. Observe that $\mathcal{F}'_{\geq p}tel(\mathcal{C})$ is the telescope complex of the sub-1-ray of Floer groups

$$\mathcal{A}_{\geq \kappa p-\delta \ell_n} CF^*(M, H_n; \Lambda) \subset CF^*(M, H_n; \Lambda).$$

As the action filtration on each $CF^*(M, H_n; \Lambda)$ is exhaustive, continuation maps increase action, and $\kappa p - \delta \ell_n \to -\infty$ as $n \to \infty$, the result follows by Lemma A.1. The argument for $\mathcal{F}_{\geq p}\left(tel(\mathcal{C}_{SH}) \otimes \mathcal{Q}_{\geq q}\Lambda\right) \subset tel(\mathcal{C}_{SH}) \otimes \mathcal{Q}_{\geq q}\Lambda$ is identical. □

Lemma 5.8. *For any p, we have*

$$H^j\left(\sigma_{<p}\mathcal{F}'_{\geq p}tel(\mathcal{C})\right) \cong QH^j(M; \Lambda) \qquad \textit{for } j < p-1.$$

Proof. Applying Lemma 5.7 and the PSS isomorphism, we have

$$\begin{aligned} &H^j(\mathcal{F}_{\geq p}tel(\mathcal{C})) = H^j(tel(\mathcal{C})) = \varinjlim_n HF^j(M, H_n; \Lambda) \\ &= \varinjlim_n QH^j(M; \Lambda) = QH^j(M; \Lambda). \end{aligned}$$

The result now follows as the degree truncation $\sigma_{<p}$ does not affect cohomology in degrees $< p-1$. □

We now denote

$$\left(SC_\Lambda^{(p)}, d^{(p)}\right) := \sigma_{<p}\mathcal{F}_{\geq p}(SC_\Lambda, d \otimes id_\Lambda).$$

For any $p > q$, we have a natural chain map $SC_\Lambda^{(p)} \to SC_\Lambda^{(q)}$, induced by the inclusion $\mathcal{F}_{\geq p} \subset \mathcal{F}_{\geq q}$ and the projection $\sigma_{<p} \twoheadrightarrow \sigma_{<q}$. In particular, we obtain an inverse system $\mathcal{SC}_*$ of graded filtered $\mathcal{Q}_{\geq 0}\Lambda$-modules. We consider the 'homotopy inverse limit'

$$\widetilde{SC}_\Lambda := \underleftarrow{tel}(\mathcal{SC}_*),$$

(see Sect. A.4 for the notation). We denote the differential by $\widetilde{d}$, and equip it with the filtration $\widetilde{\mathcal{Q}}$ induced by $\mathcal{Q}_{\geq \bullet}$ (see Remark A.8).

We now make precise the notion of 'filtered quasi-isomorphism' appearing in Theorem B (1). We consider the category of $\mathbb{Q}$-graded filtered $\mathcal{Q}_{\geq 0}\Lambda$-cochain complexes $(M, d, \mathcal{Q}_{\geq \bullet})$, where multiplication by e^a increases the degree and the filtration level by a. Morphisms are $\mathcal{Q}_{\geq 0}\Lambda$-linear filtered chain maps. A morphism in this category is called a filtered quasi-isomorphism if it induces a quasi-isomorphism on each associated graded. Objects M and N are said to be filtered quasi-isomorphic if there exists a zigzag of filtered quasi-isomorphisms between them. This implies, in particular, that we have isomorphisms $H^j(\mathrm{Gr}_k^{\mathcal{Q}} M) \cong H^j(\mathrm{Gr}_k^{\mathcal{Q}} N)$ for all j, k.

Lemma 5.9. *The filtered complex $(\widetilde{SC}_\Lambda, \widetilde{d}, \widetilde{\mathcal{Q}}_{\geq \bullet})$ is filtered quasi-isomorphic to $(SC_\Lambda, d \otimes id_\Lambda, \mathcal{Q}_{\geq \bullet})$ in the above sense.*

Proof. We have maps of inverse systems

$$\begin{tikzcd}
SC_\Lambda & SC_\Lambda \arrow[l, "id"'] & SC_\Lambda \arrow[l, "id"'] & \cdots \arrow[l] \\
\mathcal{F}_{\geq 0}SC_\Lambda \arrow[u] \arrow[d] & \mathcal{F}_{\geq 1}SC_\Lambda \arrow[l] \arrow[u] \arrow[d] & \mathcal{F}_{\geq 2}SC_\Lambda \arrow[l] \arrow[u] \arrow[d] & \cdots \arrow[l] \\
\sigma_{<0}\mathcal{F}_{\geq 0}SC_\Lambda & \sigma_{<1}\mathcal{F}_{\geq 1}SC_\Lambda \arrow[l] & \sigma_{<2}\mathcal{F}_{\geq 2}SC_\Lambda \arrow[l] & \cdots, \arrow[l]
\end{tikzcd}$$

both of which induce a filtered quasi-isomorphism on the corresponding inverse telescope complex. For the upper map, this follows from Lemma 5.7. The lower map requires a little more argument. We first observe that $H^j(\mathrm{Gr}_k\sigma_{<p}\mathcal{F}_{\geq p}SC_\Lambda) \cong H^j(\mathrm{Gr}_k SC_\Lambda)$ for $j < p-1$. It follows easily that for each j, the inverse system $H^j(\mathrm{Gr}_k\sigma_{<p}\mathcal{F}_{\geq p}SC_\Lambda)$ satisfies the Mittag-Leffler condition, so its $\varprojlim^1$ vanishes. Therefore, the cohomology of the kth-associated graded of the inverse telescope of the bottom inverse system is

$$\varprojlim H^j(\mathrm{Gr}_k\sigma_{<p}\mathcal{F}_{\geq p}SC_\Lambda) = H^j(\mathrm{Gr}_k SC_\Lambda),$$

by Lemma A.7. This completes the argument.

Finally, we observe that there is a filtered quasi-isomorphism from the inverse telescope of the top inverse system to SC_Λ. Indeed, we take the composition

$$\underleftarrow{tel}\left(SC_\Lambda \xleftarrow{id} SC_\Lambda \xleftarrow{id} \dots\right) \to \prod_{p\in\mathbb{N}} SC_\Lambda \to SC_\Lambda$$

where the first map is the natural one (i.e., the one appearing in the proof of Lemma A.7), and the second map is given by projecting to any of the identical factors. Because this inverse system clearly satisfies the Mittag-Leffler condition, the proof of Lemma A.7 shows that the induced map on cohomology is the obvious isomorphism

$$\varprojlim_p H^*(SC_\Lambda) \cong H^*(SC_\Lambda).$$

Therefore, the chain map is a quasi-isomorphism, and applying the same argument to the associated graded pieces shows that it is a filtered quasi-isomorphism. This completes the necessary zig-zag of filtered quasi-isomorphisms. □

Proposition 5.10. (= Proposition 1.12) *For any Floer solution u that contributes to $\mathcal{C}(H_\tau, J_\tau)$ with both ends asymptotic to SH-type orbits, we have $u \cdot \boldsymbol{\lambda} \geq 0$. In case of equality, u is contained in K^R_σ.*

Proof. Let $u : \mathbb{R} \times S^1 \to M$ be a pseudoholomorphic curve contributing to $\mathcal{C}(H_\tau, J_\tau)$, with both ends asymptotic to SH-type orbits. We choose $\epsilon > 0$ so that u is transverse to $\partial K^R_{\sigma+\epsilon}$, and in a neighbourhood of $\partial K^R_{\sigma+\epsilon}$ we have that $H_\tau = h_\tau \circ \rho^R$ and J_τ is of contact type. We will apply Proposition 3.9 to the part of u that lies in $\{\rho^R \geq \sigma + \epsilon\}$, to show that u is contained in

$K^R_{\sigma+\epsilon}$; applying the same argument to a sequence of such ϵ converging to 0 will show that $u \subset K^R_\sigma$ as required.

We check the hypotheses of Proposition 3.9 one by one. First recall that we chose J_τ to be of contact type along $\partial K^R_{\sigma+\epsilon}$, so hypothesis (1) is satisfied.

Now we check hypothesis (2). We have $H_\tau = h_\tau \circ \rho^R$ in a neighbourhood of $\partial K^R_{\sigma+\epsilon}$. Thus, $\mathcal{K} = \left(h_{\psi(s)} \circ \rho^R\right) \mathrm{d}t$ in this region, where $\psi(s)$ is either constant in the case of a Floer differential, or $\psi(s) = n$ for $s \ll 0$ and $\psi(s) = n+1$ for $s \gg 0$, in the case of a continuation map. Now observe that $h_n(\rho)$ is a linear function of ρ for $\rho \geq \sigma$, and h_τ is obtained by linear interpolation from the h_n, hence is also linear in ρ; this establishes hypothesis (2).

Finally, we check hypothesis (3). We have $\mathcal{K} = H_{\psi(s)}(t)\mathrm{d}t$, so $d_\Sigma \mathcal{K} = \partial_s H_{\psi(s)}(t)\mathrm{d}t \geq 0$ as H_τ is increasing. Furthermore, we have $\{\mathcal{K}, \mathcal{K}\}(\partial_s, \partial_t) = \{0, H(s,t)\} = 0$. It remains to address the term $\mathrm{d}\beta$ appearing in the hypothesis. Observe that $h_n(\rho) = \ell_n(\rho - \sigma) + \delta_n = \alpha_n \rho + \beta_n$, where we have arranged that the 'constant terms' $\beta_n = -\ell_n \sigma + \delta_n$ are decreasing. We can extend β_n to β_τ by linear interpolation, just as we did for h_τ; this will clearly be a decreasing function of τ. We then have $\beta = \beta_{\psi(s)}\mathrm{d}t$, and it is clear that $\mathrm{d}\beta \leq 0$. Putting the three terms together,

$$d_\Sigma \mathcal{K} - \{\mathcal{K}, \mathcal{K}\} - \mathrm{d}\beta \geq 0,$$

verifying hypothesis (3). The result now follows by Proposition 3.9. □

Proof of Theorem B. Item (1) holds by Lemma 5.9. For item (2), we observe that $SC^{(p)}_\Lambda$ comes equipped with another differential, namely the pullback of the differential on $\sigma_{<p}\mathcal{F}'_{\geq p} tel(\mathcal{C})$ under the isomorphism of Lemma 5.6, which we denote by $\partial^{(p)}$. The difference $d^{(p)} - \partial^{(p)}$ does not decrease the $\mathcal{Q}$-filtration, by Proposition 5.10. Any Floer solution u contributing to the part of $\partial^{(p)}$ which preserves the $\mathcal{Q}$-filtration must satisfy $u \cdot \boldsymbol{\lambda} = 0$, and hence be contained in K^R_σ by Proposition 5.10. These are precisely the Floer solutions contributing to $d^{(p)}$, so in fact $d^{(p)} - \partial^{(p)}$ strictly increases the $\mathcal{Q}$-filtration. The maps in the inverse system are clearly chain maps for the differentials $\partial^{(p)}$, so $\widetilde{SC}_\Lambda$ admits a corresponding differential, which we denote by ∂; and $\widetilde{d} - \partial$ strictly increases the $\widetilde{\mathcal{Q}}$-filtration, by the corresponding property of $d^{(p)} - \partial^{(p)}$.

For item (3), we observe that Lemma 5.8 implies that the inverse system $H^j(\sigma_{<p}\mathcal{F}'_{\geq p} tel(\mathcal{C}))$ has the Mittag–Leffler property, and in particular has $\varprojlim^1 = 0$. Therefore we have

$$\begin{aligned} H^j(\widetilde{SC}_\Lambda, \partial) &\cong \varprojlim H^j(\sigma_{<p}\mathcal{F}'_{\geq p} tel(\mathcal{C})) \qquad \text{by Lemma } A.7 \\ &\cong QH^j(M; \Lambda) \qquad \text{by Lemma 5.8 again.} \end{aligned}$$

□

5.3. Proof of Theorem C

To fit with the standard terminology for spectral sequences, in which filtrations are assumed to be increasing (see [46, Chapter 5]), we turn the decreasing filtrations $\widetilde{\mathcal{Q}}_{\geq \bullet}$ into increasing ones by setting $\bar{\mathcal{Q}}_j = \widetilde{\mathcal{Q}}_{\geq -j}$.

Lemma 5.11. *Suppose that Hypothesis* A *holds. Then the $\bar{\mathcal{Q}}$-filtration on $\widetilde{SC}_\Lambda$ is bounded below. (Recall that this means that for each i, there exists $q(i)$ such that $\bar{\mathcal{Q}}_{q(i)}\widetilde{SC}^i_\Lambda = 0$.)*

Proof. If $i(\gamma \otimes e^a) = i$, then

$$a_0\widetilde{\mathcal{Q}}(\gamma \otimes e^a) = a = i(\gamma \otimes e^a) - i(\gamma) \leq i$$

by Lemma 4.24. Thus we may take $q(i) = \lfloor -i/a_0 \rfloor - 1$. □

Proof of Theorem C. We start by establishing that the inclusion

$$\left(\bigcup_q \bar{\mathcal{Q}}_q\widetilde{SC}_\Lambda, \partial\right) \subset \left(\widetilde{SC}_\Lambda, \partial\right)$$

is a quasi-isomorphism. This follows as

$$\begin{aligned} &H^*\left(\bigcup_q \bar{\mathcal{Q}}_q\widetilde{SC}_\Lambda, \partial\right) \\ &\quad = \varinjlim_q H^*\left(\bar{\mathcal{Q}}_q\widetilde{SC}_\Lambda, \partial\right) \qquad \text{as direct limit commutes with cohomology} \\ &\quad = H^*\left(\widetilde{SC}_\Lambda, \partial\right) \qquad \text{by Lemma 5.13 below.} \end{aligned}$$

The spectral sequences induced by these filtered complexes are identical (this follows immediately from the construction). The $\bar{\mathcal{Q}}$-filtration on $\bigcup_q \bar{\mathcal{Q}}_q\widetilde{SC}_\Lambda$ is exhaustive by construction, and bounded below by Lemma 5.11. Therefore, the corresponding spectral sequence converges to $H^*\left(\widetilde{SC}_\Lambda, \partial\right)$ by [46, Theorem 5.5.1]; and this is isomorphic to $QH^*(M;\Lambda)$ by Theorem B (3).

Now we identify the E_1 page. By definition we have $E_0^{j,k} = \mathrm{Gr}_j^{\bar{\mathcal{Q}}}\widetilde{SC}^{j+k}_\Lambda$, and $d_0^{j,k}$ is the differential induced by ∂. The latter is equal to the differential induced on the associated graded by $\widetilde{d}$, by Theorem B (2) (combined with the fact that any cylinder u satisfying $u \cdot \boldsymbol{\lambda} > 0$ satisfies $u \cdot \boldsymbol{\lambda} \geq a_0$). Therefore, $\left(E_0^{j,k}, d_0^{j,k}\right) = \left(\mathrm{Gr}_j^{\bar{\mathcal{Q}}}\widetilde{SC}^{j+k}_\Lambda, \widetilde{d}\right)$, which is quasi-isomorphic to $\left(\mathrm{Gr}_j^{\bar{\mathcal{Q}}}SC^{j+k}_\Lambda, d \otimes id_\Lambda\right)$ by Lemma 5.9. Observe that $\mathrm{Gr}_j^{\bar{\mathcal{Q}}}\Lambda$ is spanned by q^{-j}, and hence is concentrated in degree $-ja_0$. It follows that $E_1^{j,k} = SH^{j(1+a_0)+k}(K^R_\sigma;\Bbbk) \otimes_\Bbbk \Bbbk \cdot q^{-j}$ as claimed. □

Lemma 5.12. *The map*

$$H^j\left(\mathcal{Q}_{\geq q}SC^{(p)}_\Lambda, \partial^{(p)}\right) \to H^j\left(\mathcal{Q}_{\geq q}SC^{(r)}_\Lambda, \partial^{(r)}\right)$$

is an isomorphism, for all $p \geq r > j+2$.

Proof. Let (C, ∂) be the cone of the chain map $\left(\mathcal{Q}_{\geq q}SC^{(p)}_\Lambda, \partial^{(p)}\right) \to \left(\mathcal{Q}_{\geq q}SC^{(r)}_\Lambda, \partial^{(r)}\right)$. The $\mathcal{Q}$-filtration on C is bounded below by Lemma 5.11, and it is clearly bounded above by q. Therefore, the corresponding spectral sequence converges to the cohomology of (C, ∂) by [46, Theorem 5.5.1].

The E_1 page is the cohomology of the cone of the chain map $\left(\mathcal{Q}_{\geq q} SC_\Lambda^{(p)}, d^{(p)}\right) \to \left(\mathcal{Q}_{\geq q} SC_\Lambda^{(r)}, d^{(r)}\right)$. This cone coincides with the cone of the chain map $\mathcal{Q}_{\geq q}\mathcal{F}_{\geq p} SC_\Lambda \to \mathcal{Q}_{\geq q}\mathcal{F}_{\geq r} SC_\Lambda$ in degrees $< r-1$. The latter cone is acyclic, by Lemma 5.7. Therefore, $E_1^{j,k} = 0$ for $j+k < r-2$. Because the spectral sequence converges, this means $H^j(C,\partial) = 0$ for $j < r-2$. This implies the result. □

Lemma 5.13. *The natural map*

$$\varinjlim_q H^*\left(\bar{\mathcal{Q}}_q \widetilde{SC}_\Lambda, \partial\right) \to H^*\left(\widetilde{SC}_\Lambda, \partial\right)$$

is an isomorphism.

Proof. Note that

$$\left(\mathcal{Q}_{\geq q}\widetilde{SC}_\Lambda, \partial\right) = \underleftarrow{tel}\left(\mathcal{Q}_{\geq q} SC^{(p)}, \partial^{(p)}\right).$$

The inverse system $H^j(\mathcal{Q}_{\geq q} SC^{(p)}, \partial^{(p)})$ has the Mittag-Leffler property for all j, q, by Lemma 5.12, so

$$\begin{aligned} H^j\left(\mathcal{Q}_{\geq q}\widetilde{SC}_\Lambda, \partial\right) &\cong \varprojlim_p H^j\left(\mathcal{Q}_{\geq q} SC^{(p)}, \partial^{(p)}\right) \qquad \text{by Lemma } A.7 \\ &\cong H^j\left(\mathcal{Q}_{\geq q} SC^{(p)}, \partial^{(p)}\right) \qquad \text{for any } p > j \\ &\quad + 2, \text{ by Lemma } 5.12. \end{aligned}$$

A similar argument, using Lemma 5.8, shows that

$$H^j\left(\widetilde{SC}_\Lambda, \partial\right) = H^j\left(SC_\Lambda^{(p)}, \partial^{(p)}\right) \qquad \text{for any } p > j+1.$$

Therefore, we have an identification

$$\begin{array}{ccc} \varinjlim H^j\left(\bar{\mathcal{Q}}_q\widetilde{SC}_\Lambda, \partial\right) & \longrightarrow & H^j\left(\widetilde{SC}_\Lambda, \partial\right) \\ \| & & \| \\ \varinjlim H^j\left(\mathcal{Q}_{\geq q} SC_\Lambda^{(p)}, \partial^{(p)}\right) & \longrightarrow & H^j\left(SC_\Lambda^{(p)}, \partial^{(p)}\right), \end{array}$$

for any $p > j+2$. The bottom map is an isomorphism, because the $\mathcal{Q}$-filtration on $SC^{(p)}$ is exhaustive. □

5.4. Proof of Theorem D

The key to the proof is the following:

Proposition 5.14. *Let $\sigma^D_{\mathrm{crit}}(R) < \sigma_1 < \sigma_2 < 1$. Then there exists an isomorphism*

$$SH^*_M\left(\overline{M\backslash K^R_{\sigma_1}}; \Lambda\right) \cong SH^*_M\left(\overline{M\backslash K^R_{\sigma_2}}; \Lambda\right).$$

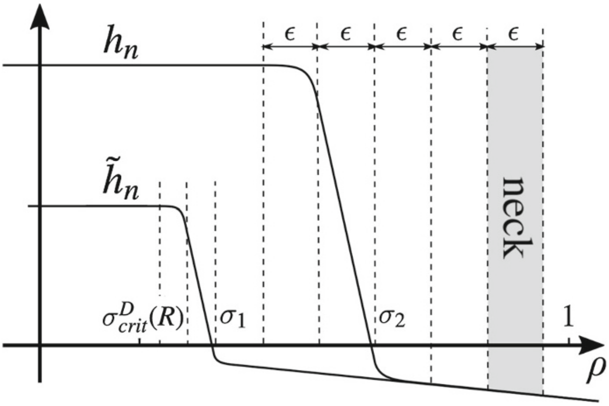


FIGURE 3. The functions h_n and $\tilde{h}_n$

The proof relies on the Contact Fukaya Trick of [38, Section 4], with which we assume some familiarity.

We first describe a special choice of acceleration data for $\overline{M\backslash K^R_{\sigma_2}} \subset M$. Let $\epsilon > 0$ be such that $0 < \sigma_2 - 2\epsilon$, $\sigma_2 + 3\epsilon < 1$, and $(\sigma_1/\sigma_2)\cdot(\sigma_2 - 2\epsilon) > \sigma^D_{\mathrm{crit}}(R)$. Let $0 < \ell_1 < \ell_2 < \cdots$ and $\delta_1 < \delta_2 < \cdots < 0$ be reals such that the Reeb flow on $Y^R_{\sigma_2} = \partial K^R_{\sigma_2}$ has no ℓ_n-periodic orbits or δ_n-periodic orbits for all n, and $\ell_n \to \infty$, $\delta_n \to 0$ as $n \to \infty$.

We choose smooth functions $h_n : \mathbb{R} \to \mathbb{R}$ satisfying:

- $h_1 < h_2 < \cdots$ (pointwise);
- $h_n'(\rho) \leq 0$;
- $h_n(\rho) = \ell_n\epsilon$ for $\rho \leq \sigma_2 - 2\epsilon$;
- $h_n(\rho) = -\ell_n(\rho - \sigma_2) + \delta_n$ for $\sigma_2 - \epsilon \leq \rho \leq \sigma_2$;
- $h_n(\rho) = \delta_n\rho$ for $\rho \geq \sigma_2 + \epsilon$.

Note that h_n converges monotonically to 0 on $[\sigma_2, \infty)$ and to $+\infty$ outside it. We define $\tilde{h}_n(\rho) = \frac{\sigma_1}{\sigma_2} h_n\left(\frac{\sigma_2}{\sigma_1}\rho\right)$, and observe that $\tilde{h}_n$ converges monotonically to 0 on $[\sigma_1, \infty)$ and to $+\infty$ outside it. See Fig. 3.

We extend h_n to h_τ by linear interpolation as before, and make a choice of acceleration data (H_τ, J_τ) for $\overline{M\backslash K^R_{\sigma_2}}$ such that H_n is a perturbation of $h_n \circ \rho^R$ in accordance with Lemma 4.17. We assume that $H_n > 0$ over the region $\{\rho^R \leq \sigma_2 - \epsilon\}$ (we can arrange this so long as $h_n \circ \rho^R > 0$ over this region, which is true so long as the δ_n are chosen sufficiently small). We need to make some special assumptions over the 'neck' region $\{\sigma_2 + 2\epsilon \leq \rho^R \leq \sigma_2 + 3\epsilon\}$, which make the contact Fukaya trick work: first, we assume that J_τ is of contact type over the neck (this includes the assumption that J_τ is invariant under translation by the Liouville vector field); second, we assume that the perturbation term $H_\tau - h_\tau \circ \rho^R$ vanishes over the neck, which is possible as $h_n \circ \rho^R = \delta_n\rho^R$ has no periodic orbits over this region.

We now choose a smooth function $f : \mathbb{R} \to \mathbb{R}$ satisfying:[17]

[17]Our function f corresponds to the function g^{-1} from [38].

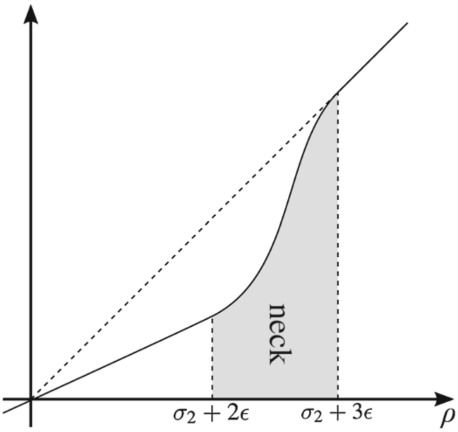


FIGURE 4. The function f

- $f'(\rho) > 0$;
- $f(\rho) \le \rho$;
- $f(\rho) = \frac{\sigma_1}{\sigma_2} \cdot \rho$ for $\rho \le \sigma_2 + 2\epsilon$;
- $f(\rho) = \rho$ for $\rho \ge \sigma_2 + 3\epsilon$.

See Fig. 4. We then define a diffeomorphism $\phi : M \to M$ by

$$\phi(m) = \begin{cases} \varphi_{\log\left(\frac{f(\rho^R(m))}{\rho^R(m)}\right)}(m) & \text{for } m \in X; \\ m & \text{for } \rho^R(m) > \sigma_2 + 3\epsilon, \end{cases}$$

where $\varphi_t : X \to X$ denotes the time-t Liouville flow. The definition is chosen so that $\rho^R(\phi(m)) = f(\rho^R(m))$. Note that ϕ sends $K^R_{\sigma_2}$ to $K^R_{\sigma_1}$ via the time-$\log(\sigma_1/\sigma_2)$ Liouville flow.

We now define acceleration data $(\tilde{H}_\tau, \tilde{J}_\tau)$ for $\overline{M \backslash K^R_{\sigma_1}}$ by taking

$$\tilde{J}_\tau = \phi_* J_\tau;$$
$$\tilde{H}_\tau = \begin{cases} \frac{\sigma_1}{\sigma_2} \phi_* H_\tau & \text{on } \phi\left(\{\rho^R \le \sigma_2 + 3\epsilon\}\right); \\ \phi_* H_\tau & \text{on } \phi\left(\{\rho^R \ge \sigma_2 + 2\epsilon\}\right). \end{cases}$$

Note that the definition of H_τ agrees on the overlaps, using the fact that $H_\tau = \delta_\tau \rho^R$ and $\phi = \varphi_{\sigma_1/\sigma_2}$ over this region. Furthermore, we observe that $\phi_* X_{H_\tau} = X_{\tilde{H}_\tau}$. (This is relatively easy to check on the complement of the image of the neck region $\phi(\{\sigma_2 + 2\epsilon \le \rho^R \le \sigma_2 + 3\epsilon\})$; over the neck region it uses the fact that both H_τ and $\tilde{H}_\tau$ are equal to $\delta_\tau \rho^R$.) The fact that J_τ is of contact type over the neck ensures that $\phi_* J_\tau$ is ω-compatible.

Thus we have constructed acceleration data for $\overline{M \backslash K^R_{\sigma_2}}$ and $\overline{M \backslash K^R_{\sigma_1}}$ leading to Floer 1-rays

$$\mathcal{C}_{\sigma_2} := \mathcal{C}(H_\tau, J_\tau) \qquad \text{and} \qquad \mathcal{C}_{\sigma_1} := \mathcal{C}(\tilde{H}_\tau, \tilde{J}_\tau),$$

such that the map $(\gamma, u) \mapsto \phi(\gamma, u) := (\phi \circ \gamma, \phi \circ u)$ defines an isomorphism $\mathcal{C}_{\sigma_2} \xrightarrow{\sim} \mathcal{C}_{\sigma_1}$, which, however, need not respect the action filtrations. We want to prove that this map of 1-rays induces an isomorphism of the completed telescopes

$$\widehat{tel}(\mathcal{C}_{\sigma_2}) \to \widehat{tel}(\mathcal{C}_{\sigma_1}).$$

Lemma 5.15. *There exist constants B, $\eta > 0$, C such that*

$$-i_{\text{mix}}(\gamma) - B \geq -i_{\text{mix}}(\phi(\gamma)) \geq \eta \cdot (-i_{\text{mix}}(\gamma)) + C$$

for any orbit γ of H_n, and the corresponding orbit $\phi(\gamma)$ of $\tilde{H}_n$.

Proof. First, we show that $-i_{\text{mix}}(\gamma) - B \geq -i_{\text{mix}}(\phi(\gamma))$, for some $B > 0$ that we specify below. Note that $i(\gamma, u_{\text{out}}) = i(\phi(\gamma), \phi(u)_{\text{out}})$, so it suffices to show $\mathcal{A}(\gamma, u_{\text{out}}) - B \geq \mathcal{A}(\phi(\gamma), \phi(u)_{\text{out}})$. Let $(\bar{\gamma}, \bar{u}_{\text{out}})$ be a capped orbit of $h_n \circ \rho^R$ corresponding to (γ, u_{out}) under Lemma 4.17. Then we have

$$\mathcal{A}(\gamma, u_{\text{out}}) = h_n(\rho^R(\bar{\gamma})) + \sum_i \nu_i^h(\bar{\gamma}) \cdot r_i^{\max}(\bar{\gamma}) + \epsilon(\gamma) \tag{5.3}$$

by Lemma 4.20, where $\epsilon(\gamma)$ is bounded, and similarly for $\phi(\gamma, u_{\text{out}})$. We consider the first term on the RHS. Note that orbits occur either in the region $\{\rho^R \leq \sigma_2 - \epsilon\}$, in which case $\tilde{h}_n(\rho^R(\phi(\bar{\gamma}))) = \frac{\sigma_1}{\sigma_2} h_n(\rho^R(\bar{\gamma})) < h_n(\rho^R(\bar{\gamma}),$ because $h_n(\rho^R(\bar{\gamma})) > 0$ (we ensured this positivity when choosing our perturbation); or in the region $\{\rho^R \geq \sigma_2\}$, where both $h_n(\rho^R(\bar{\gamma})) = \delta_n \rho^R(\bar{\gamma})$ and $\tilde{h}_n(\rho^R(\phi(\bar{\gamma}))) = \delta_n \rho^R(\phi(\bar{\gamma}))$ lie in the bounded interval $(\delta_1 \cdot (1 + \epsilon_1(R)), 0)$. In either case, we have $h_n(\rho^R(\bar{\gamma})) \geq \tilde{h}_n(\rho^R(\phi(\bar{\gamma}))) + B'$ for some fixed B'. For the second term on the RHS of (5.3), note that $h'(\rho) \leq 0$, so $\nu_i^h(\bar{\gamma}) \leq 0$. We have $\nu_i^h(\bar{\gamma}) = \nu_i^h(\phi(\bar{\gamma}))$, and $r_i^{\max}(\bar{\gamma}) \leq r_i^{\max}(\phi(\bar{\gamma}))$ (here we use our assumption that $f(\rho) \leq \rho$, as well as the fact that $Z(r_i^{\max}) < 0$). Together this yields

$$\sum_i \nu_i^h(\bar{\gamma}) \cdot r_i^{\max}(\bar{\gamma}) \geq \sum_i \nu_i^h(\phi(\bar{\gamma})) \cdot r_i^{\max}(\phi(\bar{\gamma})).$$

Adding the bounds together, and taking $B > B' + 2|\epsilon(\gamma)|$ for all γ, gives the result.

Now we consider the $-i_{\text{mix}}(\phi(\gamma)) \geq \eta \cdot (-i_{\text{mix}}(\gamma)) + C$ part of the statement. By Lemma 4.25, we have

$$-i_{\text{mix}}(\gamma) = \sum_i -\nu_i^h(\bar{\gamma}) \cdot (2 - \kappa^{-1} r_i^{\max}(\bar{\gamma})) - \kappa^{-1} h_n(\rho^R(\bar{\gamma})) + D(\gamma), \tag{5.4}$$

where $|D(\gamma)|$ is bounded. We focus on the first term on the RHS. We start by recalling that $-\nu_i^h(\phi(\bar{\gamma})) = -\nu_i^h(\bar{\gamma}) \geq 0$. Note that if $\bar{\gamma}$ is non-constant, then $\rho^R(\bar{\gamma}) > \sigma_2 - 2\epsilon$, so $\rho^R(\phi(\bar{\gamma})) > \frac{\sigma_1}{\sigma_2}(\sigma_2 - 2\epsilon) > \sigma_{\text{crit}}^D(R)$. Therefore, by Lemma 5.4, whenever $\nu_i^h(\bar{\gamma}) \neq 0$ we have

$$2 - \kappa^{-1} r_i^{\max}(\phi(\bar{\gamma})) > 2\eta,$$

where $2\eta = \eta(R) \cdot (\frac{\sigma_1}{\sigma_2}(\sigma_2 - 2\epsilon) - \sigma_{\text{crit}}^D(R)) > 0$. As a result we have

$$\sum_i -\nu_i^h(\phi(\bar{\gamma})) \cdot (2 - \kappa^{-1} r_i^{\max}(\phi(\bar{\gamma}))) \geq \eta \cdot \sum_i -\nu_i^h(\bar{\gamma}) \cdot (2 - \kappa^{-1} r_i^{\max}(\bar{\gamma})).$$

Note that this inequality also holds for the constant orbits, as then we have $\nu_i^h(\bar{\gamma}) = \nu_i^h(\phi(\bar{\gamma})) = 0$.

Now we focus on the second term on the RHS of (5.4). We saw in the first part of the proof that $\tilde{h}_n(\rho^R(\phi(\bar{\gamma}))) = \frac{\sigma_1}{\sigma_2} \cdot h_n(\rho^R(\bar{\gamma})) > 0$ for orbits with $\rho^R(\bar{\gamma}) < \sigma_2 - \epsilon$. Decreasing η if necessary so that it is less than σ_1/σ_2, and recalling that $D(\gamma)$ is bounded, we obtain the desired bound for such orbits. For the remaining orbits we recall from the first part of the proof that both $\tilde{h}_n(\rho^R(\phi(\bar{\gamma})))$ and $h_n(\rho^R(\bar{\gamma}))$ are bounded. Therefore, decreasing C if necessary, we obtain the desired bound for the remaining orbits. □

Lemma 5.16. *If (γ_j, u_j) is a sequence of capped orbits of H_{n_j} such that $i(\gamma_j, u_j) = i$ is constant, then*

$$\mathcal{A}(\gamma_j, u_j) \to +\infty \qquad \Leftrightarrow \qquad \mathcal{A}(\phi(\gamma_j, u_j)) \to +\infty.$$

Proof. Lemma 5.15 gives

$$\begin{aligned} &\kappa^{-1}\mathcal{A}(\gamma_j, u_j) - i - B \geq \kappa^{-1}\mathcal{A}(\phi(\gamma_j, u_j)) - i \geq \eta \cdot (\kappa^{-1}\mathcal{A}(\gamma_j, u_j) - i) + C \\ &\quad\Rightarrow \qquad \mathcal{A}(\gamma_j, u_j) - \kappa B \geq \mathcal{A}(\phi(\gamma_j, u_j)) \geq \eta \cdot \mathcal{A}(\gamma_j, u_j) + \kappa C + (1-\eta)\kappa i, \end{aligned}$$

where $\eta > 0$, from which the result follows. □

Remark 5.17. Notice that the contact Fukaya trick that we presented here is simpler than the one in [38] (compare Fig. 3 above with Figure 2 in [38]). We would like to stress that it is possible to use this simpler version because we are in a different situation.

Proof of Proposition 5.14. By Lemma 5.16, the isomorphism $\mathcal{C}_{\sigma_1} \cong \mathcal{C}_{\sigma_2}$ induces an isomorphism of the corresponding degreewise-action-completed telescope complexes; so

$$SC^*_M\left(\overline{M \backslash K^R_{\sigma_1}}; \Lambda\right) \cong SC^*_M\left(\overline{M \backslash K^R_{\sigma_2}}; \Lambda\right),$$

and the result follows by taking cohomology. □

We continue with the following observation of McLean:

Proposition 5.18. (see Proposition 6.20 of [25]) *Let D be an SC divisor in a symplectic manifold M. Then D is stably displaceable.* □

Proof of Theorem D. It follows from Proposition 5.18 that a neighbourhood of our divisor D is stably displaceable. Suppose that R is sufficiently small that the domains UD_i of our system of commuting Hamiltonians $\{r_i : UD_i \to [0, R)\}$ are contained in this stably displaceable neighbourhood, for all i. This ensures that $\overline{M \backslash K^R_\sigma}$ is contained in this neighbourhood for σ sufficiently close to 1. In particular,

$$SH^*_M\left(\overline{M \backslash K^R_\sigma}; \Lambda\right) = 0$$

for such σ, by Theorem 3.6. By Proposition 5.14, we see that in fact we have the same result for any $\sigma^D_{\text{crit}}(R) < \sigma < 1$. This completes the proof using Theorem 3.7, as the sets $\left\{\overline{M \backslash K^R_\sigma}\right\}_{R>0, \sigma > \sigma^D_{\text{crit}}(R)}$ exhaust $M \backslash K_{\text{crit}}$ (this follows from the fact that $\rho^R \to \rho^0$ and $\sigma^D_{\text{crit}}(R) \to \sigma_{\text{crit}}$ as $R \to 0$). □

Acknowledgements

N.S. is supported by a Royal Society University Research Fellowship, ERC Starting Grant 850713—HMS, and the Simons Collaboration on Homological Mirror Symmetry. U.V was partially supported by ERC Starting Grant 850713—HMS and is currently supported by TUBITAK CoCirculation2 Grant number 121C034. We are grateful to Yoel Groman, Ana Rita Pires, Paul Seidel, Ivan Smith, and Dmitry Tonkonog for helpful conversations; and to Dan Pomerleano for explaining Conjecture 1.20 to us.

A. Algebraic background

A.1. Filtration maps

In this section, we present an elementary framework to better deal with the type of filtrations that we encounter in this paper, which are in particular indexed by real numbers.

A filtration map on an abelian group A is a map $\rho : A \to \mathbb{R} \cup \{\infty\}$ satisfying the inequality

$$\rho(x+y) \geq \min\left(\rho(x), \rho(y)\right),$$

equality $\rho(x) = \rho(-x)$, and sending 0 to ∞. A filtration map defines a filtration by the subgroups

$$F_{\geq \rho_0} A := \{\alpha \in A \mid \rho(a) \geq \rho_0\}.$$

Note that if (V_α, ρ_α) are abelian groups equipped with filtration maps indexed by a set $\alpha \in I$, then $\bigoplus_{\alpha \in I} V_\alpha$ is equipped with a filtration map given by

$$\rho\left(\sum v_i\right) := \min\left(\rho_i(v_i)\right).$$

Let us call this the *min* construction.

We can define a pseudometric on an abelian group A with a filtration map ρ by $d(a, a') := e^{-\rho(a-a')}$. The completion $\widehat{A}$ of A is defined by taking the abelian group of Cauchy sequences in A and modding out by the subgroup of sequences which converge to 0. $\widehat{A}$ is equipped with a canonical filtration map:

$$\rho((a_i)_{i\in\mathbb{N}}) = \varinjlim \rho(a_i).$$

We call A complete, if the natural map $A \to \widehat{A}$ is bijective.

We define filtration maps $\mathcal{Q} : \Lambda \to \mathbb{R}$ by setting $\mathcal{Q}(q^a) = a$ and using the *min* construction.

A filtration map on a $\mathbb{Z}$-graded Λ-module A is a filtration map for each A^i which in addition is additive for the module action by homogenous elements of Λ. A filtration map on a Λ-cochain complex C is a filtration map on the underlying $\mathbb{Z}$-graded Λ-module, which satisfies the condition that the differential does not decrease the filtration map. Let $F_{\geq\rho_0}C := \bigoplus_{i\in\mathbb{Z}} F_{\geq\rho_0}C^i$, which is of course nothing but the filtration associated with the filtration map on C constructed by the min construction.

Filtered chain maps between Λ-cochain complexes equipped with filtration maps are defined to be chain maps that do not decrease the values of the filtration maps. Filtered chain homotopies between filtered chain maps are defined in the same fashion.

A.2. Quasi-isomorphic subcomplexes of the telescope

Let

$$\mathcal{C} := \mathcal{C}_1 \xrightarrow{f_1} \mathcal{C}_2 \xrightarrow{f_2} \mathcal{C}_3 \xrightarrow{f_3} \dots \tag{A.1}$$

be a 1-ray of $\mathbb{Q}$-graded chain complexes.

The telescope $tel(\mathcal{C})$ of $\mathcal{C}$ is defined to be the cone of the chain map

$$id - f : \bigoplus_{i=1}^{\infty} \mathcal{C}_i \to \bigoplus_{i=1}^{\infty} \mathcal{C}_i.$$

Assume that we have a commutative diagram

$$\begin{tikzcd} \mathcal{C}'_1 \arrow[r] \arrow[d] & \mathcal{C}'_2 \arrow[r] \arrow[d] & \mathcal{C}'_3 \arrow[r] \arrow[d] & \dots, \\ \mathcal{C}_1 \arrow[r] & \mathcal{C}_2 \arrow[r] & \mathcal{C}_3 \arrow[r] & \dots \end{tikzcd}$$

where the vertical maps are inclusions of subcomplexes. We call the top 1-ray $\mathcal{C}'$.

We obtain the commutative diagram of $\mathbb{Q}$-graded abelian groups

$$\begin{tikzcd} H(tel(\mathcal{C}')) \arrow[r] \arrow[d] & H(tel(\mathcal{C})) \arrow[d] \\ H\left(\varinjlim(\mathcal{C}'_i)\right) \arrow[r] & H\left(\varinjlim(\mathcal{C}_i)\right) \end{tikzcd}, \tag{A.2}$$

where the vertical maps are isomorphisms (see [40, Lemma 2.2.2] for the proof).

Lemma A.1. *Assume that every element* γ *of* $\mathcal{C}_i$ *lands inside* $\mathcal{C}'_{i+N(\gamma)}$ *for some* $N(\gamma) > 0$. *Then,*

$$tel(\mathcal{C}') \to tel(\mathcal{C})$$

is a quasi-isomorphism.

Proof. Because direct limits commute with quotients, we have

$$\varinjlim \mathcal{C}_i / \varinjlim \mathcal{C}'_i \simeq \varinjlim \mathcal{C}_i/\mathcal{C}'_i,$$

as $\mathbb{Q}$-graded chain complexes. A basic property of filtered direct limits is that any element in $\varinjlim \mathcal{C}_i/\mathcal{C}'_i$ is in the image of the canonical map $\mathcal{C}_i/\mathcal{C}'_i \to \varinjlim \mathcal{C}_i/\mathcal{C}'_i$ for some $i > 0$. This and the given condition implies that the direct limit on the RHS is zero. In particular, the lower horizontal map in Diagram (A.2) is also an isomorphism. This finishes the proof. □

A.3. Completed telescopes

Let $FiltCh_\Lambda$ be the category of free $\mathbb{Q}$-graded Λ-cochain complexes equipped with a filtration map and morphisms given by filtered chain maps.

Let

$$\mathcal{C} := \mathcal{C}_1 \xrightarrow{f_1} \mathcal{C}_2 \xrightarrow{f_2} \mathcal{C}_3 \xrightarrow{f_3} \dots \tag{A.3}$$

be a 1-ray in $FiltCh_\Lambda$. Let us equip $tel(\mathcal{C})$ with the filtration map obtained from the min construction. If we define

$$\mathcal{C}_{A_0} = F_{\geq A_0}\mathcal{C}_1 \to F_{\geq A_0}\mathcal{C}_2 \to ...,$$

then by construction

$$F_{\geq A_0} tel(\mathcal{C}) = tel(\mathcal{C}_{A_0}). \tag{A.4}$$

We define maps between two 1-rays in $FiltCh_\Lambda$ as diagrams

$$\begin{tikzcd} \mathcal{C}_1 \arrow[r] \arrow[d] & \mathcal{C}_2 \arrow[r] \arrow[d] & \mathcal{C}_3 \arrow[r] \arrow[d] & \dots , \\ \mathcal{C}'_1 \arrow[r] & \mathcal{C}'_2 \arrow[r] & \mathcal{C}'_3 \arrow[r] & \dots \end{tikzcd} \tag{A.5}$$

where the horizontal arrows are filtered chain maps and each square is equipped with a map $\mathcal{C}_i \to \mathcal{C}'_{i+1}$, which is a filtered chain homotopy between the two filtered chain maps $\mathcal{C}_i \to \mathcal{C}'_{i+1}$ obtained by composing the arrows at the edges of the square. The resulting category we call 1-*ray*-Ch_Λ.

In 1-*ray*-Ch_Λ, we also have a notion of two morphisms being *equivalent*, defined by the existence of a homotopy of maps of 1-rays. The definition is identical to [40] except that here we require all the homotopy maps to not decrease the filtration values, instead of requiring them to be $\Lambda_{\geq 0}$-module maps.

The following is a direct analogue of the second bullet point of Lemma 2.1.9 in [40] for $n = 1$. The proof is omitted.

Lemma A.2. *Let us start with a morphism in* 1*-ray-*Ch_Λ

$$\begin{tikzcd} \mathcal{C}_1 \arrow[r] \arrow[d] & \mathcal{C}_2 \arrow[r] \arrow[d] & \mathcal{C}_3 \arrow[r] \arrow[d] & \dots \\ \mathcal{C}'_1 \arrow[r] & \mathcal{C}'_2 \arrow[r] & \mathcal{C}'_3 \arrow[r] & \dots \end{tikzcd} \tag{A.6}$$

Then there is an induced filtered chain map $tel(\mathcal{C}) \to tel(\mathcal{C}')$. *Hence, the telescope construction is a functor*

$$tel : 1\text{-}ray\text{-}Ch_\Lambda \to FiltCh_\Lambda.$$

Moreover, equivalent morphisms in 1*-ray-*Ch_Λ *gives rise to filtered homotopy equivalent chain maps.*

Degreewise completion defines a functor

$$\widehat{\cdot} : FiltCh_\Lambda \to FiltCh_\Lambda.$$

Let us call a chain map $C \to C'$ between Λ-cochain complexes equipped with filtration maps a strong filtered quasi-isomorphism if it induces a quasi-isomorphism

$$F_{\geq \rho_0} C \to F_{\geq \rho_0} C',$$

for every $\rho_0 \in \mathbb{R}$. Because the filtrations are exhaustive, a strong filtered quasi-isomorphism is a quasi-isomorphism.

Lemma A.3. *Under the degreewise completion functor*

- *a strong filtered quasi-isomorphism is sent to a strong filtered quasi-isomorphism.*
- *a filtered chain homotopy is sent to a filtered chain homotopy.*

Proof. The first bullet point follows from a spectral sequence comparison theorem. Precisely, we must show that the chain map $F_{\geq \rho_0}\widehat{C} \to F_{\geq \rho_0}\widehat{C}'$ is a quasi-isomorphism, for all $\rho_0 \in \mathbb{R}$. We consider the spectral sequences associated with these filtered complexes, and the map of spectral sequences between them associated with the strong filtered quasi-isomorphism. We observe that this map is an isomorphism on the E_1 page. To see this, we first observe that the map $\mathrm{Gr}_i^F C \to \mathrm{Gr}_i^F C'$ is a quasi-isomorphism for all i, using the long exact sequence associated with a short exact sequence of chain complexes. We have $\mathrm{Gr}_i^F \widehat{C} = \mathrm{Gr}_i^F C$, and similarly for C', so the map $\mathrm{Gr}_i^F \widehat{C} \to \mathrm{Gr}_i^F \widehat{C}'$ is also a quasi-isomorphism; it then follows by construction that the map is an isomorphism on the E_1 page. The filtrations are both complete and exhaustive by construction, so the Eilenberg–Moore Comparison Theorem [46, Theorem 5.5.11] gives the result.

The second bullet point follows from the fact that the completion is an additive functor. □

Remark A.4. The first bullet point of Lemma A.3 is not explicitly used in the present paper. It would be an input in the proof of the well definedness of relative symplectic cohomology (Proposition 3.3), which we omitted.

Lemma A.5. *Let $\mathcal{C} = \mathcal{C}_1 \to \mathcal{C}_2 \to \mathcal{C}_3 \to \cdots$ be a 1-ray in $FiltCh_\Lambda$ with the following property: for any integer $i \geq 1$ and real number r there exists a positive integer N such that the composition of the Λ-chain maps from the 1-ray $C_i \to C_{i+N}$ increases the filtration map by at least r for any element of C_i. Then, $\widehat{tel}(\mathcal{C})$ is acyclic.*

Proof. As in the proof of Lemma A.3, it suffices by the Eilenberg–Moore Comparison Theorem to show that $\mathrm{Gr}_i tel(\mathcal{C})$ is acyclic for all i. Using Equation A.4 and elementary homological algebra, we obtain that $\mathrm{Gr}_i tel(\mathcal{C})$ is quasi-isomorphic to the telescope of

$$\mathrm{Gr}_i\mathcal{C}_1 \to \mathrm{Gr}_i\mathcal{C}_2 \to \cdots .$$

Because the homology of the telescope is isomorphic to the homology of the direct limit, it suffices to show the acyclicity of the direct limit of this diagram. It is easy to see that the direct limit is in fact trivial on the nose (i.e. at the chain level). $\square$

A.4. Homotopy inverse limit

Definition A.6. Let $\mathcal{C}$ be an inverse system of cochain complexes and cochain maps:

$$C_0^* \xleftarrow{i_{01}} C_1^* \xleftarrow{i_{12}} \cdots .$$

We define the cochain complex $\prod_p C_p^*$ to be the degreewise direct product of the C_p^*. There is a natural chain map $id - i : \prod_p C_p^* \to \prod_p C_p^*$, sending $(c_p) \mapsto (c_p - i_{p,p+1}(c_{p+1}))$. We define the inverse telescope complex

$$\underleftarrow{tel}(\mathcal{C}) := \mathrm{Cone}\left(\prod_p C_p^* \xrightarrow{id-i} \prod_p C_p^*\right)[-1].$$

The following recovers the Milnor exact sequence if $\mathcal{C}$ satisfies the Mittag-Leffler condition. We believe that it is standard, but we could not locate it in the literature.

Lemma A.7. *There is a short exact sequence*

$$0 \to \varprojlim{}^1 H^{j-1}\left(C_p^*\right) \to H^j\left(\underleftarrow{tel}(\mathcal{C})\right) \to \varprojlim H^j\left(C_p^*\right) \to 0.$$

Proof. The long exact sequence associated with the short exact sequence of cochain complexes

$$0 \to \prod_p C_p^*[-1] \to \underleftarrow{tel}(\mathcal{C}) \to \prod_p C_p^* \to 0$$

gives an exact sequence

$$H^{j-1}\left(\prod_p C_p^*\right) \xrightarrow{[id-i]} H^{j-1}\left(\prod_p C_p^*\right) \to H^j\left(\underleftarrow{tel}(\mathcal{C})\right)$$
$$\to H^j\left(\prod_p C_p^*\right) \xrightarrow{[id-i]} H^j\left(\prod_p C_p^*\right).$$

This gives the desired short exact sequence, as the $\varprojlim$ is defined to be the kernel of $[id - i]$, and $\varprojlim^1$ is defined to be the cokernel. $\square$

Remark A.8. If $\mathcal{C}$ is an inverse system of *filtered* cochain complexes with *filtered* cochain maps, then the inverse telescope complex acquires a filtration by

$$F_{\geq p}\left(\underleftarrow{tel}(\mathcal{C})\right) := \underleftarrow{tel}(F_{\geq p}\mathcal{C})$$

(it is clear how to regard the RHS as a subcomplex of $\underleftarrow{tel}(\mathcal{C})$).

References

[1] Abbondandolo, A., Schwarz, M.: On the Floer homology of cotangent bundles. Commun. Pure Appl. Math. **59**(2), 254–316 (2006)

[2] Abouzaid, M., Seidel, P.: An open string analogue of Viterbo functoriality. Geom. Topol. **14**(2), 627–718 (2010)

[3] Auroux, D.: Mirror symmetry and T-duality in the complement of an anti-canonical divisor. J. Gökova Geom. Topol. GGT **1**, 51–91 (2007)

[4] Batyrev, V.V.: Quantum cohomology rings of toric manifolds. Astérisque **218**, 9–34 (1993)

[5] Cho, C.-H., Yong-Geun, O.: Floer cohomology and disc instantons of Lagrangian torus fibers in Fano toric manifolds. Asian J. Math. **10**(4), 773–814 (2006)

[6] Cieliebak, K., Floer, A., Hofer, H., Wysocki, K.: Applications of symplectic homology. II. Stability of the action spectrum. Math. Z. **223**(1), 27–45 (1996)

[7] Dold, A.: Lectures on algebraic topology, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 200, 2nd edn. Springer, Berlin (1980)

[8] Eliashberg, Y., Polterovich, L.: Symplectic quasi-states on the quadric surface and Lagrangian submanifolds. Preprint arXiv:1006.2501 (2010)

[9] Elzein, F., Némethi, A.: On the weight filtration of the homology of algebraic varieties: the generalized Leray cycles. Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **1**(4), 869–903 (2002)

[10] Entov, M.: Quasi-morphisms and quasi-states in symplectic topology. In: Proceedings of the International Congress of Mathematicians—Seoul 2014, vol. II, pp. 1147–1171. Kyung Moon Sa, Seoul (2014)

[11] Evans, J.D., Lekili, Y.: Generating the Fukaya categories of Hamiltonian G-manifolds. J. Am. Math. Soc. **32**(1), 119–162 (2019)

[12] Fabert, O.: Higher algebraic structures in Hamiltonian Floer theory. Adv. Geom. **20**(2), 179–215 (2020)

[13] Fabert, O., Salchow, J.-D.: The L_∞-structure on symplectic cohomology. Preprint arXiv:1903.12143

[14] Ganatra, S., Pomerleano, D.: Symplectic cohomology rings of affine varieties in the topological limit. Geom. Funct. Anal. **30**(2), 334–456 (2020)

[15] Ganatra, S., Pomerleano, D.: A log PSS morphism with applications to Lagrangian embeddings. J. Topol. **14**(1), 291–368 (2021)

[16] Getzler, E.: Lie theory for nilpotent L_∞-algebras. Ann. Math. (2) **170**(1), 271–301 (2009)

[17] Givental, A.B.: Equivariant Gromov–Witten invariants. Int. Math. Res. Not. **13**, 613–663 (1996)

[18] Guillemin, V., Sternberg, S.: Symplectic Techniques in Physics. Cambridge University Press, Cambridge (1984)

[19] Harder, A.: Torus fibers and the weight filtration. Preprint arXiv:1908.05110

[20] Jinzenji, M.: On quantum cohomology rings for hypersurfaces in CP^{N-1}. J. Math. Phys. **38**(12), 6613–6638 (1997)

[21] Kwon, M., van Koert, O.: Brieskorn manifolds in contact topology. Bull. Lond. Math. Soc. **48**(2), 173–241 (2016)

[22] McDuff, D.: Displacing Lagrangian toric fibers via probes. Low-dimensional and symplectic topology **82**, 131–160 (2011)

[23] McLean, M.: The growth rate of symplectic homology and affine varieties. Geom. Funct. Anal. **22**(2), 369–442 (2012)

[24] McLean, M.: Reeb orbits and the minimal discrepancy of an isolated singularity. Invent. Math. **204**(2), 505–594 (2016)

[25] McLean, M.: Birational Calabi–Yau manifolds have the same small quantum products. Ann. Math. (2) **191**(2), 439–579 (2020)

[26] Oancea, A.: A survey of Floer homology for manifolds with contact type boundary or symplectic homology. In: Symplectic Geometry and Floer Homology. A Survey of the Floer Homology for Manifolds with Contact Type Boundary or Symplectic Homology, Ensaios Mat., vol. 7, pp. 51–91. Soc. Brasil. Mat., Rio de Janeiro (2004)

[27] Pomerleano, D.: Intrinsic mirror symmetry and categorical crepant resolutions. Preprint arXiv:2103.01200

[28] Piunikhin, S., Salamon, D., Schwarz, M.: Symplectic Floer–Donaldson theory and quantum cohomology. Contact and symplectic geometry (Cambridge, 1994), vol. 8, pp. 171–200 (1996)

[29] Ritter, A.F.: Floer theory for negative line bundles via Gromov–Witten invariants. Adv. Math. **262**, 1035–1106 (2014)

[30] Robbin, J., Salamon, D.: The Maslov index for paths. Topology **32**(4), 827–844 (1993)

[31] Salamon, D.: Lectures on Floer homology. Symplectic Geometry and Topology (Park City, Utah, 1997), IAS/Park City Math. Ser., vol. 7, pp. 143–229 (1999)

[32] Seidel, P.: Fukaya Categories and Picard–Lefschetz Theory. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich (2008)

[33] Sheridan, N.: On the Fukaya category of a Fano hypersurface in projective space. Publ. Math. Inst. Hautes Études Sci. **124**, 165–317 (2016)

[34] Sheridan, N.: Versality of the relative Fukaya category. Geom. Topol. **24**(2), 747–884 (2020)

[35] Sun, Y.: Index bounded relative symplectic cohomology. Preprint arXiv:2109.06146

[36] Tehrani, M.F., McLean, M., Zinger, A.: Normal crossings singularities for symplectic topology. Adv. Math. **339**, 672–748 (2018)

[37] Tonkonog, D.: From symplectic cohomology to Lagrangian enumerative geometry. Adv. Math. **352**, 717–776 (2019)

[38] Tonkonog, D., Varolgunes, U.: Super-rigidity of certain skeleta using relative symplectic cohomology. Preprint arXiv:2003.07486

[39] Varolgunes, U.: Mayer–Vietoris property for relative symplectic cohomology. PhD thesis, Massachusetts Institute of Technology (2018)

[40] Varolgunes, U.: Mayer–Vietoris property for relative symplectic cohomology. Geom. Topol. **25**(2), 547–642 (2021)

[41] Venkatesh, S.: The quantitative nature of reduced Floer theory. Adv. Math. **383**, Paper No. 107682, 80 (2021)

[42] Vezzosi, G.: Basic structures on derived critical loci. Differ. Geom. Appl. **71**, 101635, 11 (2020)

[43] Viterbo, C.: Generating functions, symplectic geometry, and applications. In: Proceedings of the International Congress of Mathematicians, vols. 1, 2 (Zürich, 1994), pp. 537–547. Birkhäuser, Basel (1995)

[44] Viterbo, C.: Functors and computations in Floer cohomology. II. Prépublication Orsay 98–15. http://www.math.ens.fr/~viterbo/FCFH.II.2003.pdf (1998)

[45] Viterbo, C.: Functors and computations in Floer homology with applications. I. Geom. Funct. Anal. **9**(5), 985–1033 (1999)

[46] Weibel, C.A.: An introduction to homological algebra. Cambridge Studies in Advanced Mathematics, vol. 38. Cambridge University Press, Cambridge (1994)

[47] Weiwei, W.: On an exotic Lagrangian torus in $\mathbb{C}P^2$. Compos. Math. **151**(7), 1372–1394 (2015)

[48] Ziller, W.: The free loop space of globally symmetric spaces. Invent. Math. **41**(1), 1–22 (1977)

Matthew Strom Borman
Department of Mathematics
University of Illinois
1409 W. Green Street
Urbana
IL 61801
USA

Nick Sheridan and Umut Varolgunes
School of Mathematics
University of Edinburgh
Edinburgh EH9 3FD
UK
e-mail: nick.sheridan@ed.ac.uk

Accepted: April 20, 2021.

J. Fixed Point Theory Appl. (2022) 24:40
https://doi.org/10.1007/s11784-022-00961-w
Published online April 19, 2022

Journal of Fixed Point Theory and Applications

A symplectic embedding of the cube with minimal sections and a question by Schlenk

Fabian Ziltener

Abstract. I prove that the open unit cube can be symplectically embedded into a longer polydisc in such a way that the area of each section satisfies a sharp bound and the complement of each section is path-connected. This answers a variant of a question by F. Schlenk.

Mathematics Subject Classification. 53D05 Symplectic manifolds, general.

Keywords. Symplectic embedding, section of embedding, wrapping construction.

1. The main result

Let $n \geq 2$. By $q^1, p_1, \dots, q^n, p_n$ we denote the standard coordinates in $\mathbb{R}^{2n}$, and we equip $\mathbb{R}^{2n}$ with the standard symplectic form $\omega_0 := \sum_{i=1}^n \mathrm{d}q^i \wedge \mathrm{d}p_i$.[1] We denote by B_r^m resp. $\overline{B}_r^m$ the open resp. closed ball in $\mathbb{R}^m$ of radius r around 0. M. Gromov's famous non-squeezing theorem [2, Corollary, p. 310] implies that $\overline{B}_r^{2n}$ does not symplectically embed into the closed unit symplectic cylinder $\overline{B}_1^2 \times \mathbb{R}^{2n-2}$ if $r > 1$. In [5] F. Schlenk investigated how flexible symplectic embeddings are in the case $r \leq 1$. More precisely, for every $z \in \mathbb{R}^{2n-2}$, we define

$$\iota_z : \mathbb{R}^2 \to \mathbb{R}^{2n}, \quad \iota_z(y) := (y, z).$$

Answering a question of D. McDuff [4], in [5, Theorem 1.1] Schlenk proved that for every $a > 0$, there exists a symplectic embedding φ of $\overline{B}_1^{2n}$ into

This article is part of the topical collection "Symplectic geometry—A Festschrift in honour of Claude Viterbo's 60th birthday" edited by Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk.

[1]Following the physicists' convention, I use an upper index for the i-th coordinate of a point q in the base manifold $\mathbb{R}^n$ and lower index for the i-th coordinate of a covector $p \in \mathbb{R}^n = T_q^*\mathbb{R}^n$.

[2] This means two-dimensional Lebesgue measure.

$\overline{B}_1^2 \times \mathbb{R}^{2n-2}$, such that for every $z \in \mathbb{R}^{2n-2}$, the section $\iota_z^{-1}\big(\varphi\big(\overline{B}_1^{2n}\big)\big)$ has area[2] at most a.

Schlenk's lifting method [6, Sect. 8.4] also shows that for every positive integer k and every $a > \frac{1}{k}$, there exists a symplectic embedding of the open cube $(0,1)^{2n}$ into the open polydisc $(0,1)^{2n-1} \times (0,k)$, whose sections have area at most a. The main result of the present article answers the following two questions:

Question 1. *Is this statement true with the integer k replaced by a general real number $c \geq 1$?*

Question 2. *Can the bound a on the areas of the sections be made sharp, i.e. equal to $\frac{1}{c}$?*[3]

I also answer a variant of the following question of Schlenk. For every bounded subset S of $\mathbb{R}^m$ we define the *bounded hull of S* to be the union of S and all bounded connected components of $\mathbb{R}^m \setminus S$.

Question 3. [Schlenk, [5], Question 2.2] *Let $n \geq 2$, φ be a symplectic embedding of B_1^{2n} into $B_1^2 \times \mathbb{R}^{2n-2}$, and $a < \pi$. Does there exist $z \in \mathbb{R}^{2n-2}$ such that the bounded hull of the closure of the section $\iota_z^{-1}\big(\varphi\big(B_1^{2n}\big)\big)$ has area at least a?*

The main result of this article is the following.

Theorem 4. *For every $n \geq 2$ and $c \in [1,\infty)$, there exists a symplectic embedding $\varphi : (0,1)^{2n} \to (0,1)^{2n-1} \times (0,c)$, such that for every $z \in \mathbb{R}^{2n-2}$ the following holds:*

(i) The section $\iota_z^{-1}\big(\varphi\big((0,1)^{2n}\big)\big)$ has area equal to $\frac{1}{c}$ or is empty.
(ii) Its complement in $\mathbb{R}^2$ is path-connected.

This theorem answers Questions 1 and 2 affirmatively. It also provides a negative answer to Schlenk's Question 3 with the word "closure" dropped. It even implies that there exists a symplectic embedding for which the bounded hull of each section has arbitrarily small area:

Corollary 5. *For every $n \geq 2$ and $a > 0$, there exists a symplectic embedding $\psi : B_1^{2n} \to B_1^2 \times \mathbb{R}^{2n-2}$, such that the bounded hull of each section of $\psi\big(B_1^{2n}\big)$ has area at most a.*

(For a proof see p. 9.) This corollary is optimal in the sense that its statement becomes false if we replace B_1^{2n} and B_1^2 by the *closed* balls $\overline{B}_1^{2n}$ and $\overline{B}_1^2$. Even the following is true:

Proposition 6. [F. Lalonde, D. McDuff] *Let $n \in \mathbb{N}$ and $\varphi : \overline{B}_1^{2n} \to \overline{B}_1^2 \times \mathbb{R}^{2n-2}$ be a symplectic embedding.*[4] *Then, there exists $z \in \mathbb{R}^{2n-2}$, such that the section $\iota_z^{-1}\big(\varphi\big(\overline{B}_1^{2n}\big)\big)$ contains the circle of radius 1 around 0.*

[3] There is always a section of area at least $\frac{1}{c}$, by Fubini's theorem. Hence, $a = \frac{1}{c}$ is the minimal possible bound.

[4] We do not impose any restrictions on how φ maps the boundary of the ball.

In particular, the bounded hull of this section equals $\overline{B}^2_1$, which has area π.

Proof of Proposition 6. This follows from [3, Lemma 1.2] and from Gromov's non-squeezing theorem. □

Remark. Let φ be as in the statement of Theorem 4. Then each section of the image of φ equals its own bounded hull. Hence, φ is a *sharp* counterexample to a variant of Question 3 concerning embeddings of cubes.

In the case $n = 2$, the idea of proof of Theorem 4 is to consider the linear symplectic map $\Psi : (q, p) \mapsto (Q, P)$ induced by the Lagrangian shear $p \mapsto P := \big(p_1, cp_1 + p_2\big)$. The P_2-sections of the image of the square $(0, 1)^2$ under this shear have length at most $\frac{1}{c}$. Hence, the area of each section of $\Psi\big((0, 1)^4\big)$ is at most $\frac{1}{c}$. To make the image of Ψ fit in the polydisc $(0, 1)^3 \times (0, c)$, we wrap its upper part (in P_2-direction) back to the lower part, by passing to the quotient $\mathbb{R}/c\mathbb{Z}$. We also wrap the Q^1-coordinate. See Fig. 1.

Finally, we compose the resulting map with the product of two area preserving embeddings of finite cylinders into rectangles. This yields a symplectic embedding with the desired properties.

Remark. (method of proof, related result, terminology).

- This construction is similar to L. Traynor's symplectic wrapping construction, which she used e.g. to show that certain polydiscs embed into certain cubes, see [7] and [6, Chap. 7]. One difference is that I wrap coordinates of mixed type (Q and P), whereas Traynor wraps coordinates of pure type.
- Schlenk proved a nonsharp result regarding the areas of the bounded hulls of the sections. More precisely, his folding method [6, Sect. 8.3] can be used to prove that for every $n \geq 2$, positive integer k, and $\ell \in (0, 1)$ there exists a symplectic embedding $\varphi : (0, \ell)^{2n} \to (0, 1)^{2n-1} \times (0, k)$, such that the bounded hull of every section of $\varphi\big((0, \ell)^{2n}\big)$ has area at most $\frac{1}{k}$. Theorem 4 improves this in the following ways:
  - It treats the critical case $\ell = 1$.
  - It makes the area estimate sharp.
  - It holds for any real number $c \geq 1$, not only for an integer $c = k$.
  - The proof of Theorem 4 is easier than the folding method.
- In [5] and [6, p. 226], Schlenk calls the bounded hull of the closure of a set its "simply connected hull". The simply connected hull of a simply connected compact subset S of $\mathbb{R}^m$ need not be equal to S. In the case $m \geq 3$, an example is given by the sphere $S := S^{m-1}$, and in the case $m = 2$ by the Warsaw circle. This set is produced by closing up the topologist's sine curve with an arc. For this reason, I prefer the terminology "bounded hull". Since this notion is only defined for *bounded* subsets of $\mathbb{R}^m$, no confusion should arise from the fact that the bounded hull of a bounded set S can differ from S.
- For more information about related work, see [6].

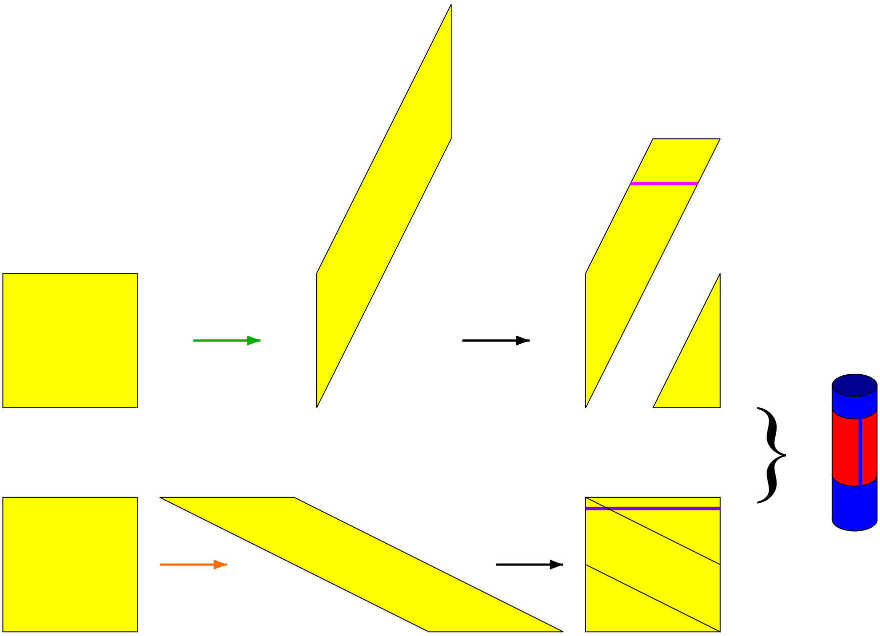


Figure 1. The green arrow depicts the Lagrangian shear $p \mapsto P := \big(p_1, cp_1 + p_2\big)$, and the orange arrow the induced shear in the q-plane. The black arrows depict the wrapping maps. The magenta line segment is a $\overline{P}_2$-section of the image of the square under the composed map in the p-plane, where $\overline{P}_2 \in \mathbb{R}/(c\mathbb{Z})$. The violet set in the lower part of the picture depicts a $\overline{Q}^2$-section of the image of the open square under the composed map in the q-plane, where $\overline{Q}^2 \in \mathbb{R}/\mathbb{Z}$. The bracket } indicates that the product of these two sets is given by the red ribbon on the blue cylinder. The image of this ribbon under some area-preserving map is a section of the image of the desired symplectic embedding φ. It has area equal to $\frac{1}{c}$. In the picture c equals 2. If c is not an integer then the horizontal edge of the upper wrapped triangle in the lower part of the picture has length less than 1

2. Proofs of Theorem 4 and of Corollary 5

In the proofs of Theorem 4 and Corollary 5, we will use the following lemma.

Lemma 7. [squaring the disc and the cylinder] *We denote* $r := \pi^{-\frac{1}{2}}$.

(i) There exists a homeomorphism

$$\kappa : \overline{B}_r^2 \to [0,1]^2,$$

that restricts to a (smooth) symplectomorphism between the interiors.

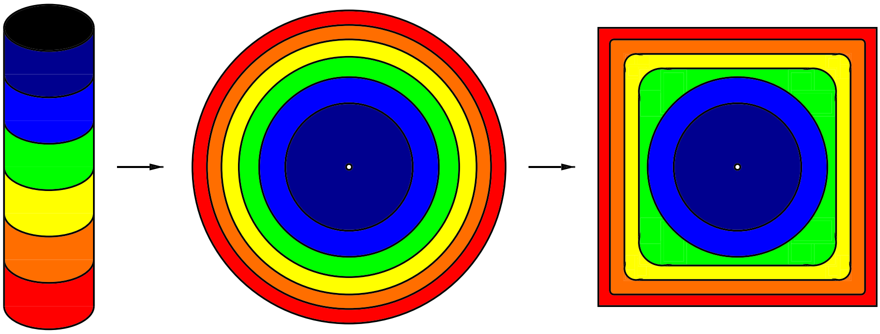


Figure 2. The two arrows depict area-preserving smooth embeddings whose composition is an area-preserving embedding of the open cylinder into the open square. The idea of proof of Lemma 7 is to choose such maps in such a way that they continuously extend to the closed cylinder and the closed disc, respectively

(ii) For every $y_0 \in (0,1)^2$*, there exists continuous map*

$$\lambda : (\mathbb{R}/\mathbb{Z}) \times [0,1] \to [0,1]^2$$

that maps $(\mathbb{R}/\mathbb{Z}) \times \{1\}$ *to* y_0*, and restricts to a homeomorphism from* $(\mathbb{R}/\mathbb{Z}) \times [0,1)$ *to* $[0,1]^2 \setminus \{y_0\}$ *and to a symplectomorphism from* $(\mathbb{R}/\mathbb{Z}) \times (0,1)$ *to* $(0,1)^2 \setminus \{y_0\}$.

The idea of proof of this lemma is explained by Fig. 2.

In the proof of Lemma 7, we will use the following.

Remark 8. [straightening corners] We denote by Σ the square $[0,1]^2$ without the corners. Let $r > 0$ and S be a subset of the circle of radius r consisting of four points. There exists a homeomorphism $\theta : [0,1]^2 \to \overline{B}_r^2$ that restricts to a diffeomorphism from Σ onto $\overline{B}_r^2 \setminus S$, such that $(\theta|\Sigma)_*\omega_0$ extends to a nonvanishing smooth 2-form on $\overline{B}_r^2$.

To see this, observe that the map

$$\widetilde{\theta} : [0,\infty)^2 \to \mathbb{R} \times [0,\infty), \quad \widetilde{\theta}(z) := \frac{z^2}{|z|},$$

is a homeomorphism that restricts to a diffeomorphism from $[0,\infty)^2 \setminus \{0\}$ onto $\big(\mathbb{R} \times [0,\infty)\big) \setminus \{0\}$, that satisfies

$$\big(\widetilde{\theta}\big|[0,\infty)^2 \setminus \{0\}\big)_* \omega_0 = \frac{\omega_0}{2}.$$

The desired map θ can be constructed from four copies of $\widetilde{\theta}$ (one for each corner), using charts for $\overline{B}_r^2$ and a cut off argument.

Proposition 9. [Banyaga's Moser stability with boundary] *Let M be a compact connected oriented smooth manifold, and Ω_0, Ω_1 volume forms on M satisfying*

$$\int_M \Omega_0 = \int_M \Omega_1.$$

Then there exists a diffeomorphism φ of M satisfying

$$\varphi_* \Omega_0 = \Omega_1, \quad \varphi|\partial M = \mathrm{id}.$$

Proof. See [1, Théorème, p. 127]. □

Proof of Lemma 7. To prove (i), we define $r := \pi^{-\frac{1}{2}}$ and choose a map θ as in Remark 8. We define

$$M := \overline{B}^2_r, \quad \Omega_0 := \theta_* \omega_0, \quad \Omega_1 := \omega_0.$$

We have

$$\int_M \Omega_0 = \int_\Sigma \omega_0 = 1 = \int_M \Omega_1.$$

Hence, the hypotheses of Proposition 9 are satisfied. We choose a diffeomorphism φ as in the statement of this proposition. The map

$$\kappa := (\varphi \circ \theta)^{-1} : \overline{B}^2_r \to [0,1]^2$$

has the required properties.

We prove (ii). There exists a symplectomorphism

$$\chi : (\mathbb{R}/\mathbb{Z}) \times [0,1) \to \overline{B}^2_r \setminus \{0\}.$$

For example, consider $y : \mathbb{R}/\mathbb{Z} \to \mathbb{C} = \mathbb{R}^2$, $y(\bar{q}) := e^{2\pi i q}$, where $q \in \bar{q}$ is an arbitrary representative, and define

$$\chi(\bar{q}, p) := r\sqrt{1-p}\, y(\bar{q}).$$

We choose a symplectomorphism ξ of $[0,1]^2$ that equals the identity in a neighbourhood of the boundary and maps $\kappa(0)$ to y_0. We obtain such a map as the Hamiltonian flow of a suitable function on $(0,1)^2$ with compact support. The map

$$\lambda := \begin{cases} \xi \circ \kappa \circ \chi & \text{on } (\mathbb{R}/\mathbb{Z}) \times [0,1), \\ y_0 & \text{on } (\mathbb{R}/\mathbb{Z}) \times \{1\} \end{cases}$$

has the required properties. This proves (ii) and completes the proof of Lemma 7. □

Remark. The proof of part (i) of Lemma 7 is based on Proposition 9. The proof of that result in turn uses Moser isotopy and a lemma that roughly states that a primitive of an exact top degree form can be chosen in such a way that it vanishes on the boundary of the manifold. An alternative approach for proving Lemma 7(i) is based on the proof of [6, Lemma 3.1.5]. That lemma states the following.

Lemma 10. *Let U and V be bounded and simply connected domains in $\mathbb{R}^2$ of equal area and let $\mathcal{L}_U$ and $\mathcal{L}_V$ be admissible*[5] *families of loops in U and V, respectively. Then there is a symplectomorphism between U and V mapping loops to loops.*

We define $U := B_r^2$ and $V := (0,1)^2$. The idea of the alternative approach to Lemma 7(i) is to choose admissible families of loops in such a way that the symplectomorphism constructed in the proof of Lemma 10 extends continuously and injectively to the closure of $\overline{U}$. (Neither condition is automatically satisfied.) The extension will then have the desired properties.

Proof of Theorem 4. Consider first the case $n = 2$. We denote by

$$\pi : \mathbb{R}^4 \to (\mathbb{R}/\mathbb{Z}) \times \mathbb{R} \times \mathbb{R} \times (\mathbb{R}/c\mathbb{Z})$$

the canonical projection, and equip $(\mathbb{R}/\mathbb{Z}) \times \mathbb{R} \times \mathbb{R} \times (\mathbb{R}/c\mathbb{Z})$ with the symplectic form induced by ω_0 and π. We denote $y_0 := z_0 := \left(\frac{1}{2}, \frac{1}{2}\right)$. We choose a map λ as in Lemma 7(ii). It follows from the same lemma that there exists a symplectomorphism

$$\lambda' : (0,1) \times (\mathbb{R}/c\mathbb{Z}) \to \big((0,1) \times (0,c)\big) \setminus \{z_0\}.$$

We define

$$\Psi : \mathbb{R}^4 \to \mathbb{R}^4, \quad \Psi\big(q^1, p_1, q^2, p_2\big) := \big(q^1 - cq^2, p_1, q^2, cp_1 + p_2\big),$$
$$\varphi := (\lambda \times \lambda') \circ \pi \circ \Psi\big|(0,1)^4.$$

The map φ is well defined, since $\pi \circ \Psi$ maps $(0,1)^4$ to the product of the domains of λ and λ'. The map φ is a symplectic immersion, as it is the composition of three symplectic immersions. A straight-forward argument shows that $\pi \circ \Psi\big|(0,1)^4$ is injective. Since $\lambda|(\mathbb{R}/\mathbb{Z}) \times (0,1)$ and λ' are injective, it follows that the same holds for φ. Hence, φ is a symplectic embedding of $(0,1)^4$ into $(0,1)^3 \times (0,c)$.

We verify that the map φ has properties (i) and (ii) stated in Theorem 4. We fix a point $(Q^2, \overline{P}_2) \in (0,1) \times (\mathbb{R}/c\mathbb{Z})$. We have

$$\begin{aligned}
U_{Q^2,\overline{P}_2} &:= \Big\{(\overline{Q}^1, P_1) \in (\mathbb{R}/\mathbb{Z}) \times (0,1) \,\Big|\, \big(\overline{Q}^1, P_1, Q^2, \overline{P}_2\big) \in \pi \circ \Psi\big((0,1)^4\big)\Big\} \\
&= V_{Q^2} \times W_{\overline{P}_2}, && (1) \\
V_{Q^2} &:= \big\{q^1 - cQ^2 + \mathbb{Z} \,\big|\, q^1 \in (0,1)\big\} = (\mathbb{R}/\mathbb{Z}) \setminus \{-cQ^2 + \mathbb{Z}\}, && (2) \\
W_{\overline{P}_2} &:= \Big\{P_1 \in (0,1) \,\Big|\, \exists p_2 \in (0,1) : \, cP_1 + p_2 + c\mathbb{Z} = \overline{P}_2\Big\} \\
&= (0,1) \cap \bigcup_{p_2 \in (0,1)} \frac{\overline{P}_2 - p_2}{c}, && (3)
\end{aligned}$$

where $\frac{\overline{P}_2 - p_2}{c} \in \mathbb{R}/\mathbb{Z}$. The set $W_{\overline{P}_2}$ is an open subinterval of $(0,1)$ or the union of two such subintervals. It has length $\frac{1}{c}$. Using (1) and (2), it follows that $U_{Q^2,\overline{P}_2}$ has area equal to $\frac{1}{c}$. Figure 3 depicts the set $U_{Q^2,\overline{P}_2}$ in the case that $W_{\overline{P}_2}$ is connected.

[5] See [6, Definition 3.1.4].

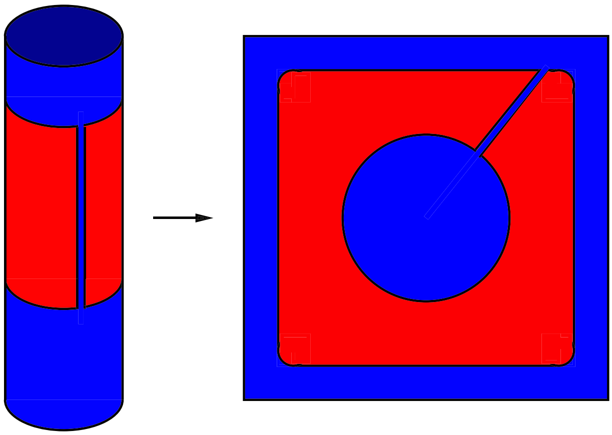


Figure 3. The arrow depicts the area-preserving map $\lambda : (\mathbb{R}/\mathbb{Z}) \times (0,1) \to (0,1)^2$. (Compare to Fig. 2.) It sends the upper part of the cylinder close to the centre of the disc. The red ribbon on the cylinder is $U_{Q^2,\overline{P}_2}$, the section of the image of φ. The point $\overline{P}_2$ determines the height of the upper boundary of the red ribbon, and therefore the radius of the circle inside the square. The point Q^2 determines the position of the blue slit. Because of this slit, the blue set on the right is path-connected. This is the complement of the image of the section under the map λ

Let now $z \in \big((0,1) \times (0,c)\big) \setminus \{z_0\}$. We denote $(Q^2, \overline{P}_2) := \lambda'^{-1}(z)$. We have

$$\lambda^{-1}\left(\iota_z^{-1}\big(\varphi\big((0,1)^4\big)\big)\right) = U_{Q^2,\overline{P}_2}. \tag{4}$$

Since λ is area-preserving, it follows that the section $\iota_z^{-1}\big(\varphi\big((0,1)^4\big)\big)$ has area equal to $\frac{1}{c}$. For $z = z_0$ or z outside of $(0,1) \times (0,c)$, the section is empty. This proves (i).

To prove property (ii), consider the continuous path

$$y : [0,1] \to [0,1]^2, \quad y(t) := \lambda\big(-cQ^2 + \mathbb{Z}, t\big).$$

The point $y(0)$ lies on the boundary of the square $[0,1]^2$. It follows from (2) that the path y lies in the complement of $\iota_z^{-1}\big(\varphi\big((0,1)^4\big)\big)$ in $\mathbb{R}^2$. Every point outside $(0,1)^2$ can be connected to $y(0)$ through a continuous path outside of $(0,1)^2$. Every point in the complement of $\iota_z^{-1}\big(\varphi\big((0,1)^4\big)\big)$ in $(0,1)^2$ can be connected to a point on the path y through a path in this complement. This follows from (4) and the facts $U_{Q^2,\overline{P}_2} = V_{Q^2} \times W_{\overline{P}_2}$, $V_{Q^2} = (\mathbb{R}/\mathbb{Z}) \setminus \{-cQ^2 + \mathbb{Z}\}$. See again Fig. 3. This proves (ii).

Hence, φ has the desired properties. This proves Theorem 4 in the case $n = 2$. For $n \geq 3$ we take the product of φ with the identity map. $\square$

In the proof of Corollary 5, we will use the following.

Remark 11. (monotonicity) The bounded hull is monotone in the sense that if $A \subseteq B \subseteq \mathbb{R}^m$ are bounded sets then the bounded hull of A is contained in the bounded hull of B.

Proof of Corollary 5. We define $r := \pi^{-\frac{1}{2}}$. By a rescaling argument it suffices to show that for every $a \in (0,1]$ there exists a symplectic embedding $\psi : B_r^{2n} \to B_r^2 \times \mathbb{R}^{2n-2}$, such that the bounded hull of each section of $\psi(B_r^{2n})$ has area at most a. To prove this statement, we choose φ is as in the conclusion of Theorem 4 with $c := \frac{1}{a}$. We choose a map κ as in Lemma 7(i). The map

$$\psi := (\kappa^{-1} \times \mathrm{id}) \circ \varphi \circ \big(\kappa \times \cdots \times \kappa\big) : B_r^{2n} \to B_r^2 \times \mathbb{R}^{2n-2}$$

is a symplectic embedding. Let $z \in \mathbb{R}^{2n-2}$. Property (ii) in Theorem 4 implies that the complement of $V := \kappa^{-1}\Big(\iota_z^{-1}\big(\varphi\big((0,1)^{2n}\big)\big)\Big)$ in $\mathbb{R}^2$ is path-connected. Hence, V equals its bounded hull. The section $\iota_z^{-1}\big(\psi(B_r^{2n})\big)$ is contained in V. Using Remark 11, it follows that the bounded hull of this section is also contained in V. Using Theorem 4(i) and that κ is area-preserving, it follows that the bounded hull of $\iota_z^{-1}\big(\psi(B_r^{2n})\big)$ has area at most $\frac{1}{c} = a$. Hence, ψ has the desired properties. This proves Corollary 5. □

Acknowledgements

I would like to thank Felix Schlenk for an interesting discussion and for proof-reading the first version of this article. I also thank the anonymous referee for useful comments.

References

[1] Banyaga, Augustin: Formes-volume sur les variétés à bord. Enseignement Math **20**(2), 127–131 (1974). (**MR 0358649**)

[2] Gromov, Mikhail: Pseudo holomorphic curves in symplectic manifolds. Invent. Math **82**(2), 307–347 (1985). (**MR 809718**)

[3] Lalonde, François, McDuff, Dusa: Local non-squeezing theorems and stability. Geom. Funct. Anal. **5**(2), 364–386 (1995). (**MR 1334871**)

[4] McDuff, Dusa.: *Fibrations in symplectic topology*, Proceedings of the International Congress of Mathematicians, Vol. I (Berlin, 1998), no. Extra Vol. I, pp. 339–357.(1998) MR 1648038

[5] Schlenk, Felix.: *On a question of Dusa McDuff*, Int. Math. Res. Not. no. 2, 77–107. (2003) MR 1936580

[6] Schlenk, Felix: Embedding problems in symplectic geometry, De Gruyter Expositions in Mathematics, vol. 40, p. 2147307. Walter de Gruyter GmbH & Co. KG, Berlin (2005)

[7] Traynor, Lisa: Symplectic packing constructions. J. Differential Geom **42**(2), 411–429 (1995). (**MR 1366550**)

Fabian Ziltener
Department of Mathematics
Universiteit Utrecht
3584 CD Utrecht
The Netherlands
e-mail: `F.Ziltener@uu.nl`

Accepted: April 20, 2021.

Correction to: Symplectic Geometry: A Festschrift in Honour of Claude Viterbo's 60th Birthday

Helmut Hofer, Alberto Abbondandolo, Urs Frauenfelder, and Felix Schlenk

Correction to:
H. Hofer et al. (eds.), *Symplectic Geometry*,
https://doi.org/10.1007/978-3-031-19111-4

This book was inadvertently published with volume editor names instead of the names of the chapter authors in the online version of the chapters. This has been corrected throughout the book.

The updated version of the book can be found at
https://doi.org/10.1007/978-3-031-19111-4

H. Hofer et al. (eds.), *Symplectic Geometry*, https://doi.org/10.1007/978-3-031-19111-4_36