Panos M. Pardalos
Themistocles M. Rassias *Editors*

Essays in Mathematics and its Applications

In Honor of Stephen Smale's 80th Birthday

Springer

Essays in Mathematics and its Applications

Panos M. Pardalos • Themistocles M. Rassias
Editors

Essays in Mathematics and its Applications

In Honor of Stephen Smale's 80th Birthday

Springer

Editors
Panos M. Pardalos
Department of Industrial
and Systems Engineering
University of Florida
Gainesville, Florida, USA

Themistocles M. Rassias
Department of Mathematics
National Technical University of Athens
Athens, Greece

ISBN 978-3-642-28820-3 ISBN 978-3-642-28821-0 (eBook)
DOI 10.1007/978-3-642-28821-0
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012944569

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This volume is dedicated to Professor Stephen Smale on the occasion of his 80th birthday. Besides his startling result of the proof of the Poincaré conjecture for all dimensions greater than or equal to five in 1960, Professor Smale's groundbreaking contributions in various fields of mathematics have marked the second part of the twentieth century and beyond.

Stephen Smale has done pioneering work in differential topology, global analysis, dynamical systems, nonlinear functional analysis, numerical analysis, theory of computation, and machine learning as well as applications in the physical and biological sciences and economics. In sum, he has manifestly broken the barriers among the different fields of mathematics and dispelled some remaining prejudices. He is indeed a universal mathematician.

Smale has been honored with several prizes and honorary degrees, including, among others, the Fields Medal (1966), the Veblen Prize (1966), the National Medal of Science (1996), and the Wolf Prize (2006/2007).

Besides mathematics, Smale has been a keen learner and collector of rare minerals. Last but not least, Smale is a humanist and has been an active advocate of freedom and equality of rights for all people in the USA and worldwide. He has been politically involved in a number of such movements in the past and is still active.

We wish to express our gratitude to the many distinguished professors who accepted the opportunity to contribute to this publication.

Gainesville — Panos M. Pardalos
Athens — Themistocles M. Rassias

Contents

Transitivity and Topological Mixing for C^1 Diffeomorphisms

Flavio Abdenur and Sylvain Crovisier

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract We prove that, on connected compact manifolds, both C^1-generic conservative diffeomorphisms and C^1-generic transitive diffeomorphisms are topologically mixing. This is obtained through a description of the periods of a homoclinic class and by a control of the period of the periodic points given by the closing lemma.

1 Introduction

In his seminal dissertation about differentiable dynamical systems [22], Smale described the recurrence of hyperbolic diffeomorphisms:

Theorem 1 (Smale's spectral decomposition theorem). *Consider a diffeomorphism f of a compact manifold. If the non-wandering set $\Omega(f)$ is hyperbolic and contains a dense set of periodic points then it decomposes uniquely as the finite union $\Omega(f) = \Omega_1 \cup \cdots \cup \Omega_s$ of disjoint, closed, invariant subsets on each of which f is topologically transitive.*

Recall that the restriction of f to an invariant compact set Λ is *topologically transitive* if there exists a dense forward orbit, or equivalently, if for any non-empty

F. Abdenur (✉)
Ventor Investimentos, Rio de Janeiro, Brazil
e-mail: flavio.abdenur@gmail.com

S. Crovisier
Département de Mathématiques, UMR 8628, Université Paris-Sud 11, 91405 Orsay, Cedex, France
e-mail: Sylvain.Crovisier@math.u-psud.fr

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_1, © Springer-Verlag Berlin Heidelberg 2012

open sets U, V of Λ, there exists $n \geq 1$ such that $f^n(U) \cap V \neq \emptyset$. Later on, Bowen noticed [6] that each piece Ω_i admits a further decomposition $\Omega_i = X_{i,1} \cup \cdots \cup X_{i,\ell_i}$ into disjoint closed subsets on each of which $g = f^{\ell_i}$ is *topologically mixing*: for any non-empty open sets U, V of $X_{i,j}$, there exists $n_0 \geq 1$ such that $f^n(U) \cap V \neq \emptyset$ for any $n \geq n_0$.

Let $\text{Diff}^1(M)$ denote the space of C^1-diffeomorphisms of a connected compact boundaryless manifold M endowed with the C^1-topology. Our goal is to study the recurrence of its generic non-hyperbolic elements. A robust obstruction to the transitivity is the existence of a trapping region, i.e. a non-empty open set $U \neq M$ such that $f(\overline{U}) \subset U$. When this obstruction does not occur, it follows from [3] that the generic dynamics is transitive on the whole manifold. More precisely, in this case M is a homoclinic class (see the Sect. 2 below) implying that an iterate $g = f^n$ of f is topologically mixing. Our goal here is to show that this is also the case for the first iterate f:

Theorem 2. *There exists a dense G_δ subset $\mathcal{G} \subset \text{Diff}^1(M)$ such that any transitive diffeomorphism $f \in \mathcal{G}$ is topologically mixing.*

A similar statement was obtained in [1] for flows but the case of diffeomorphisms is more difficult: the proof requires closing and connecting lemmas in order to build segments of orbit which visit successively two given regions U, V. For technical reasons, the obtained orbits may be shorter than what is expected (see [9]) so that the intersections between $f^n(U)$ and V could occur only at some particular times n, breaking down the topological mixing. The main point of the present paper is thus a closing lemma with control of the connecting time (Sect. 3).

Many examples of non-hyperbolic robustly transitive diffeomorphisms have been constructed, see for instance [4] and [19]. Theorem 2 trivially implies that these dynamics become topologically mixing *modulo an arbitrarily small C^1-perturbation*. In some particular cases, there is a stronger result: among robustly transitive partially hyperbolic diffeomorphisms with one-dimensional center bundle, the set topologically mixing dynamics contains an open and dense subset, see [11] and [5, Corollary 3].

As far as we know, all of the known examples of robustly transitive diffeomorphisms are topologically mixing. This raises the following questions:

Questions.

1. Is *every* robustly transitive diffeomorphism topologically mixing?
2. Failing that, is topological mixing at least a C^1-open-and-dense condition within the space of all robustly transitive diffeomorphisms?

When M is connected and ω is a volume or a symplectic form, we denote by $\text{Diff}^1_\omega(M)$ the space of the C^1-diffeomorphisms which preserve ω, endowed with the C^1-topology. The results stated before still hold in the conservative setting and moreover any C^1-generic diffeomorphism is transitive [2, 3]. One thus gets:

Theorem 3. *Any diffeomorphism in a dense G_δ subset $\mathcal{G}_\omega \subset \text{Diff}^1_\omega(M)$ is topologically mixing.*

We in fact obtain a version of the previous statement for *locally maximal* sets, i.e. invariant compact sets $\Lambda \subset M$ having a neighborhood U such that $\Lambda = \cap_{i\in\mathbb{Z}} f^i(U)$. Conley has proved [7] for homeomorphisms of Λ that the non-existence of a trapping region is equivalent to *chain-transitivity*: for any $\varepsilon > 0$, there exists a ε-dense periodic sequence $(x_0, \dots, x_n = x_0)$ in Λ which is a ε-pseudo orbit, that is satisfies $d(f(x_i), x_{i+1}) < \varepsilon$ for any $0 \le i < n$. As a consequence of [3], for C^1-generic diffeomorphisms, any maximal invariant set which is chain-transitive is also transitive. Generalizing Bowen's result for hyperbolic diffeomorphisms, we prove:

Theorem 4. *There exists a dense G_δ subset $\mathcal{G} \subset \mathrm{Diff}^1(M)$ (or $\mathcal{G}_\omega \subset \mathrm{Diff}^1_\omega(M)$) of diffeomorphisms f such that any chain-transitive locally maximal set Λ decomposes uniquely as the finite union $\Lambda = \Lambda_1 \cup \dots \cup \Lambda_\ell$, of disjoint compact sets on each of which f^ℓ is topologically mixing.*

Moreover, for $1 \le i \le \ell$, any hyperbolic periodic $p, q \in \Lambda_i$ with same stable dimension satisfy:

- *ℓ is the smallest positive integer such that $W^u(f^\ell(p)) \cap W^s(p) \cap \Lambda \neq \emptyset$,*
- *Λ_i coincides with the closure of $W^u(p) \cap W^s(q) \cap \Lambda$.*

Clearly, Theorems 2 and 3 follow from Theorem 4.

2 The Period of a Homoclinic Class

Let f be a C^1-diffeomorphism and O be a hyperbolic periodic orbit. We denote by $W \pitchfork W'$ the set of transversal intersection points between two submanifolds $W, W' \subset M$.

2.1 Homoclinic Class

The *homoclinic class* $H(O)$ of O is the closure of the set of transverse intersection points between the stable and unstable manifolds $W^s(O)$ and $W^u(O)$. We refer to [13] for its basic properties, which we now recall:

- Two hyperbolic periodic orbit O_1, O_2 are *homoclinically related* if $W^s(O_1)$ intersects transversally $W^u(O_2)$ and $W^u(O_2)$ intersects transversally $W^s(O_1)$. This defines an equivalence relation on the set of hyperbolic periodic orbits.
- $H(O)$ is the closure of the union of the periodic orbits homoclinically related to O.
- If O, O' are homoclinically related, $H(O)$ coincides with the closure of the set of transversal intersections between $W^u(O)$ and $W^s(O')$.
- A homoclinic class is a transitive invariant set.

2.2 Period of a Homoclinic Class

The *period* $\ell(O) \geq 1$ of the homoclinic class $H(O)$ of O is the greatest common divisor of the periods of the hyperbolic periodic points homoclinically related to O. The group $\ell(O).\mathbb{Z}$ is called the *set of periods* of $H(O)$. We have the following characterization:*for $p \in O$, and $n \in \mathbb{Z}$, the manifolds $W^u(f^n(p))$ and $W^s(p)$ have a transversal intersection if and only if $n \in \ell(O).\mathbb{Z}$.* More generally:

Proposition 1. *Consider a hyperbolic periodic point q whose orbit is homoclinically related to O and such that $W^u(p) \pitchfork W^s(q) \neq \emptyset$. Then $W^u(f^n(q)) \pitchfork W^s(p) \neq \emptyset$ if and only if $n \in \ell(O).\mathbb{Z}$. In particular $W^u(q)$ intersects transversally $W^s(p)$.*

This proposition is a consequence of Smale's theorem on transversal homoclinic points [21] and of Palis' inclination lemma [15] (or λ-lemma).

Theorem 5 (Smale's homoclinic theorem). *Consider a local diffeomorphism f, a hyperbolic fixed point p and a transverse homoclinic intersection $x \in W^s(p) \pitchfork W^u(p)$. Then, in any neighborhood of $\{p\} \cup \{f^k(x)\}_{k \in \mathbb{Z}}$, there exists, for some iterate f^n, a hyperbolic set K containing p and x.*

Lemma 1 (Palis' inclination lemma). *Let p be a hyperbolic fixed point and $N \subset M$ be a submanifold which intersects $W^s(p)$ transversally. Then for any compact disc $D \subset W^u(p)$ there exists a sequence (D_k) of discs of N and an increasing sequence (n_k) of positive integers such that $f^{n_k}(D_k)$ converges to D in the C^1-topology.*

Proof (Proof of Proposition 1). Let p, q be two hyperbolic periodic points whose orbits are homoclinically related and assume that $W^u(p) \pitchfork W^s(q) \neq \emptyset$. Let $G_{p,q}$ be the set of integers n such that $W^u(f^n(q)) \pitchfork W^s(p) \neq \emptyset$.

The set $G_{p,q}$ is invariant by addition. Indeed if $n \in G_{p,q}$, then $W^u(f^n(q)) \pitchfork W^s(p)$ and $W^u(f^n(p)) \pitchfork W^s(f^n(q))$ are non-empty. The inclination lemma implies that $W^s(p)$ accumulates on $W^s(f^n(q))$ so that it transversally intersects $W^u(f^n(p))$. If moreover $m \in G_{p,q}$, we have $W^u(f^{n+m}(q)) \pitchfork W^s(f^n(p)) \neq \emptyset$, so that $W^s(p)$ intersects transversally $W^u(f^{n+m}(q))$ and $n + m \in G_{p,q}$.

The set $G_{p,q}$ is invariant by subtraction by the period r of p. Hence, for $n \in G_{p,q}$, the opposite $-n = (r-1).n - r.n$ also belongs to $G_{p,q}$. So $G_{p,q}$ coincides with $G_{q,p}$ and is a group.

If q' is another hyperbolic periodic point whose orbit is homoclinically related to those of p, q and satisfies $W^u(q') \pitchfork W^s(p) \neq \emptyset$, then $G_{p,q} = G_{p,q'}$. Indeed the stable and the unstable manifolds of q, q' intersect transversally, and the unstable manifolds of $f^n(q)$ and $f^n(q')$ intersect the same stable manifolds. Consequently the group $G = G_{p,q}$ contains all the periods of the hyperbolic periodic orbits homoclinically related to the orbit O of p. In particular, G contains $\ell(O).\mathbb{Z}$.

Conversely, let us consider $n \in G$ and an intersection point $x \in W^u(f^n(p)) \pitchfork W^s(p)$. One defines a local diffeomorphism g which coincides with f^r in a (fixed) neighborhood of p and which sends an iterate $x^u = f^{-n-k_u.r}(x) \in W^u(p)$ onto an iterate $x^s = f^{k_s.r}(x) \in W^s(p)$. Since by Smale's homoclinic theorem the orbits of x^s, x^u, p for g are contained in a hyperbolic set, one can shadow a pseudo-orbit $p, g^{-m}(x^s), g^{-m+1}(x^s), \dots, g^{m-1}(x^s), p$ by a hyperbolic periodic orbit that is homoclinically related to p. By construction this orbit is contained in a hyperbolic periodic orbit O' of f that is homoclinically related to O and whose period has the form $n + k.r$ for some $k \in \mathbb{Z}$, where r is the period of p. This implies that n belongs to $\ell(O).\mathbb{Z}$, so that $G = \ell(O).\mathbb{Z}$.

2.3 Pointwise Homoclinic Class

If p is a point of the hyperbolic periodic orbit O, its *pointwise homoclinic class* $h(p)$ is the closure of the set of transverse intersection points between the manifolds $W^s(p)$ and $W^u(p)$: this set is in general **not** invariant by f.

Lemma 2. *If the orbit of a hyperbolic periodic point q is homoclinically related to O and $W^u(p), W^s(q)$ have a transverse intersection point, then $h(p)$ coincides with the closure of the set of transversal intersections between $W^u(p)$ and $W^s(q)$. In particular $h(p) = h(q)$.*

Proof. By Proposition 1 $W^u(q)$ and $W^s(p)$ have a transverse intersection point. If n, m are the periods of p and q, then for f^{nm} the points p, q are fixed, homoclinically related and their homoclinic class coincide with $h(p), h(q)$ and with the set of transversal intersections between $W^u(p)$ and $W^s(q)$.

The following proposition decomposes the homoclinic classes in the form $\Lambda_1 \cup \dots \cup \Lambda_\ell$ such that f^ℓ is topologically mixing on each piece Λ_i. However, a priori the pieces are not disjoint.

Proposition 2. *Let $p \in O$ and $\ell = \ell(O)$ be the period of the homoclinic class. Then:*

- *$H(O)$ is the union of the iterates $f^k(h(p))$;*
- *$h(p)$ is invariant by f^ℓ;*
- *The restriction of f^ℓ to $h(p)$ is topologically mixing;*
- *If $f^j(h(p)) \cap f^k(h(p))$ has non-empty interior in $H(O)$, then $f^j(h(p)) = f^k(h(p))$.*

Proof. Let m, n, k be three integers. We claim that the closure of $W^u(f^k(p)) \pitchfork W^s(f^m(p))$ is either empty or coincides with $f^{m+n\ell}(h(p))$. Indeed the first set coincides with the image by $f^{m+n.\ell}$ of the closure of $W^u(f^{k-m-n.\ell}(p)) \pitchfork W^s(f^{-n.\ell}(p))$. If this set is non-empty, one deduces that $k - m - n.\ell$ belongs to $\ell.\mathbb{Z}$, hence $W^u(f^{k-m-n.\ell}(p))$ and $W^u(p)$ accumulate on each other. Similarly, $W^s(f^{-n.\ell}(p))$ and $W^s(p)$ accumulate on each other. Consequently the closure of $W^u(f^{k-m-n.\ell}(p)) \pitchfork W^s(f^{-n.\ell}(p))$ coincides with $h(p)$, proving the claim.

The claim immediately implies that $H(O)$ coincides with the union of the iterates of $h(p)$ and that $f^\ell(h(p))$ coincides with $h(p)$. Hence the two first items hold.

Let $U, V \subset M$ be two open sets which intersect $h(p)$. We have to show that for any large n, the intersection $f^{n.\ell}(U) \cap V$ intersects $h(p)$. We first introduce two points $x \in U \cap (W^u(p) \pitchfork W^s(p))$ and $y \in V \cap (W^u(p) \pitchfork W^s(p))$. Let us consider a disc $D \subset W^u(p) \cap U$ containing x. The inclination lemma shows that for n large $f^{n.\ell}(D)$ accumulates on any disc of $W^u(p)$, and hence on the local unstable manifold of y. As a consequence, for n large $f^{n.\ell}$ intersects transversally in V the local stable manifold of y, which proves the third item in the statement.

Let A_k denote the interior of $f^k(h(p))$ in $H(O)$: it is non-empty and dense in $f^k(h(p))$. The open and dense subset $A_0 \cup \cdots \cup A_{\ell-1}$ of $H(O)$ is the disjoint union of elements of the form $A_{k_1} \cap A_{k_2} \cap \cdots \cap A_{k_s}$. By construction this partition is invariant by f. Since the restriction of f^ℓ to each set A_k is topologically mixing, one deduces that A_k is not subdivided by the partition. This means that either $A_j = A_k$ or $A_j \cap A_k = \emptyset$. In the latter case one gets $f^j(h(p)) = f^k(h(p))$ proving the last item.

2.4 *Perturbation of the Period*

For any diffeomorphism g close to f, one can consider the hyperbolic continuation O_g of O. By the implicit function theorem a given transverse intersection between $W^s(O)$ and $W^u(O)$ will persist, and hence the period of the homoclinic class of O_g depends upper-semi-continously with g. The following perturbation lemma provides a mechanism for the non-continuity of the period.

Proposition 3. *Let us consider $f \in \mathrm{Diff}^r(M)$, for some $r \geq 1$, and two hyperbolic periodic orbits O, O' having a* cycle*: $W^u(O) \cap W^s(O') \neq \emptyset$ and $W^u(O') \cap W^s(O) \neq \emptyset$.*

If the period of the orbit O' does not belong to the set of periods $\ell(O).\mathbb{Z}$ of the class $H(O)$, then there exists a diffeomorphism g that is arbitrarily C^r-close to f such that $\ell(O_g) < \ell(O)$.

Proof. Let us assume that the stable dimension of O_1 is smaller than or equal to that of O_2. Let us choose $p \in O$ and $q \in O'$ so that $W^s(p)$ and $W^u(q)$ have an intersection point x and for some $n \in \mathbb{Z}$ the manifolds $W^u(f^n(p))$ and $W^s(q)$ have an intersection point y. One can perturb f in an arbitrarily small neighborhood of y so that the intersection becomes quasi-transversal (i.e. $T_y W^u(O) + T_y W^s(O') = T_y M$), hence robust, and we have not modified the orbits of p, q, x.

The inclination lemma ensures that $W^u(f^n(p))$ accumulates on $W^u(q)$. This shows that if one fixes a small neighborhood U of x, there exist $x_s \in W^s(p)$ and $x_u \in W^u(f^n(p))$ arbitrarily close to x whose respective future and past semiorbits avoid U. By a small C^r-perturbation it is thus possible to create a transverse intersection between $W^s(p)$ and $W^u(f^n(p))$ so that after the perturbation n belongs to the set of periods of the homoclinic class of O.

If $n \notin \ell(O).\mathbb{Z}$ we are done. Otherwise, if r is the period of O' then $W^u(f^{n+r}(p)) \cap W^s(q) \neq \emptyset$. One can thus repeat the same construction replacing n by $n + r \notin \ell(O)$ and obtain the conclusion.

2.5 *Relative Homoclinic Classes*

If O is contained in an open set U, one defines the *relative homoclinic class* $H(O, U)$ of O in U as the closure of $W^s(O) \pitchfork W^u(O) \cap (\bigcap_{n\in\mathbb{Z}} f^n(U))$. It is a transitive invariant compact set contained in $\overline{U}$.

All the results stated in the previous sections remain valid if one considers hyperbolic periodic orbits and transverse homoclinic/heteroclinic orbits in U. For instance, two hyperbolic periodic orbits $O_1, O_2 \subset U$ are *homoclinically related in* U if both $W^s(O_1) \pitchfork W^u(O_2)$ and $W^s(O_2) \pitchfork W^u(O_1)$ meet $\bigcap_{n\in\mathbb{Z}} f^n(U)$. The homoclinic class $H(O, U)$ coincides with the closure of the set of hyperbolic periodic points whose orbit is homoclinically related to O in U.

The relative pointwise homoclinic class $h(p, U)$ is the intersection $h(p) \cap H(O, U)$.

3 A Closing Lemma with Time Control

Pugh's closing Lemma [16] allows one to turn any non-wandering point into a periodic point via a small C^1-perturbation of the dynamics. The proof selects a segment of orbit of the original diffeomorphism which will be closed, so that it is difficult to control the period of the obtained orbit. In order to control the period of the closed orbit we propose here a different argument which uses several orbit segments of the original dynamics, as in the proof of Hayashi's connecting lemma. A technical condition appears on the periodic points.

Definition 1. A periodic point x is *non-resonant* if, for the tangent map $D_x f^r$ at the period, the eigenvalues having modulus equal to one are simple (i.e. their characteristic spaces are one-dimensional) and do not satisfy relations of the form

$$\lambda_1^{k_1} \lambda_2^{k_2} \dots \lambda_s^{k_s} = 1,$$

where the numbers $\lambda_1, \overline{\lambda_1}, \dots, \lambda_s, \overline{\lambda_s}$ are distinct and the k_i are positive integers.

This is obviously satisfied by hyperbolic periodic points. Also this condition is generic in $\mathrm{Diff}^1(M)$ and in $\mathrm{Diff}^1_\omega(M)$, see [12, 18, 20]. The statement of the closing lemma with time control is the following.

Theorem 6 (Closing lemma with time control). *Let f be a C^1-diffeomorphism, $\ell \geq 2$ be an integer, x be either a non-periodic point or a non-resonant periodic* *point. Assume that each neighborhood V of x intersects some iterate $f^n(V)$ such that n is not a multiple of ℓ. Then, for diffeomorphisms g arbitrarily C^1-close to f, x is periodic and its period is not a multiple of ℓ.*

If moreover there exists an open set U such that each small neighborhood V of x has a forward iterate $f^n(V)$ which intersects V and such that $f(V), \dots, f^{n-1}(V)$ are contained in U, then the orbit of x under g can be chosen in U. If f belongs to $\mathrm{Diff}^1_\omega(M)$, so does g.

3.1 Pugh's Algebraic Lemma and Tiled Perturbation Domains

The main connexion results for the C^1-topology [2, 3, 8, 10, 16, 17] are obtained by using the two following tools. The first one allows to perform independent elementary perturbations. They are usually obtained through Pugh's "algebraic" lemma (the name refers to the proof which only involves sequences of linear maps) and with combinatorial arguments.

Lemma 3 (Elementary perturbation lemma). *For any neighborhood $\mathcal{V}$ of the identity in $\mathrm{Diff}^1(M)$ (or in $\mathrm{Diff}^1_\omega(M)$), there exists $\theta \in (0,1)$ and $\delta > 0$ such that for any finite collection of disjoint balls $B_i = B(x_i, r_i)$, with $r_i < \delta$, and for any collection of points $y_i \in B(x_i, \theta.r_i)$, there exists a diffeomorphism $h \in \mathcal{V}$ supported on the union of the B_i which satisfies $h(x_i) = y_i$ for each i.*

Let d be the dimension of M. A *cube C* of $\mathbb{R}^d$ is the image of the standard cube $[-1,1]^d$ by a translation and an homothety. For $\lambda > 0$ we denote $\lambda.C$ the cube having the same barycenter and whose edges have a length equal to λ times those of C. A cube C of a chart $\varphi\colon V \to \mathbb{R}^d$ of M is the preimage by φ of a cube C' of $\mathbb{R}^d$. The cube $\lambda.C$ is the preimage $\varphi^{-1}(\lambda.C')$.

Lemma 4 (Pugh's algebraic lemma). *For any $f \in \mathrm{Diff}^1(M)$ and any $\eta \in (0,1)$, there exists $N \geq 1$ and a covering of M by charts $\varphi\colon V \to \mathbb{R}^d$ whose cubes C have the following property.*

For any $a, b \in C$, there is a connecting sequence *$(a = a_0, a_1, \dots, a_N = b)$ such that for each $0 \leq k \leq N-1$ the point a_k belongs to $f^k(5/4.C)$ and the distance $d(a_k, f^{-1}(a_{k+1}))$ is smaller than η times the distance between $f^k(5/4.C)$ and the complement of $f^k(3/2.C)$.*

In the following, one will fix a C^1-diffeomorphism f, a neighborhood $\mathcal{U} \subset \mathrm{Diff}^1(M)$, and:

- Some constants θ, δ provided by the elementary perturbation lemma and associated to the neighborhood $\mathcal{V} = \{h = f^{-1} \circ g, g \in \mathcal{U}\}$ of the identity;
- An integer $N \geq 1$ and a finite collection of charts $\{\varphi_s\colon V_s \to \mathbb{R}^d\}_{s \in S}$ given by Pugh's algebraic lemma and associated to the constant $\eta = (\theta/4)^{4^d}$.

Definition 2. The collection of charts $\{\varphi_s\}_{s\in S}$ is a *tiled perturbation domain* if we have:

- The $f^k(V_s)$, with $s \in S$, $0 \le k < N-1$, have diameter $< \delta$ and are pairwise disjoint;
- Each set V_k is tiled, i.e. is the union of cubes (the *tiles*) with pairwise disjoint interior satisfying: each tile C intersects (is *adjacent to*) at most 4^d other tiles, each of them having a diameter which differs from the diameter of C by a factor in $[1/2, 2]$.

Any point distinct from its $N-1$-first iterates belongs to a tiled domain, see Fig. 1 in [3]. Note also that if the interior of $3/2.C$ and $3/2.C'$ intersect, then the tiles C, C' are adjacent.

3.2 The Orbit Selection

The perturbation domain is used to connect together a collection of segments of orbits of f.

Definition 3. A *pseudo-orbit with jumps in the perturbation domain* is a sequence (y_i) such that for each i, either $f(y_i) = y_{i+1}$ or the points y_i, $f^{-1}(y_{i+1})$ are contained in a same set V_s.

We are interested by the following additional properties:

1. When the y_i, $f^{-1}(y_{i+1})$ belong to V_s, there is a connecting sequence $(a_{i,0},\dots, a_{i,N})$ with $a_{i,0} = y_i$ and $a_{i,N} = f^{N-1}(y_{i+1})$ such that for each $0 \le k < N$, the ball $B_{i,k} = B(a_{i,k}, r_{i,k})$ with $r_{i,k} = \theta^{-1}.d(a_{i,k}, f^{-1}(a_{i,k+1}))$ is contained in $f^k(V_s)$.
2. The balls $B_{i,k}$ are pairwise disjoint.

In order to control the periodic pseudo-orbits, we also introduce an integer $\ell \ge 1$.

3. The length of the periodic pseudo-orbit $(y_1,\dots,y_n = y_0)$ is not a multiple of ℓ.

When a pseudo-orbit $(y_1,\dots,y_n = y_0)$ satisfies conditions (1), (2) and $f(y_0) \neq y_1$, one can apply the elementary perturbation lemma and build a diffeomorphism $g \in \mathcal{U}$ by perturbing f in the union of the balls $B_{i,k}$ such that the point y_0 belongs to a periodic orbit of length n.

These conditions can be obtained by the following proposition.

Proposition 4. *Let (y_i) be a periodic pseudo-orbit with jumps in the perturbation domain such that when y_i, $f^{-1}(y_{i+1})$ differ, they are contained in a same tile of the perturbation domain.*

Then there exists another periodic pseudo-orbit with jumps in the perturbation domain which satisfies (1) and (2). Moreover if the first pseudo-orbit satisfies (3), then so does the second one.

Proof. Note that by an arbitrarily small modification of the initial pseudo-orbit, the points of the pseudo-orbit do not belong to the boundaries of the tiles. In this way, to each point of the pseudo-orbit which belong to the perturbation domain is associated a unique tile.

3.2.1 The Shortcut Process

The new orbit is obtained from the first one by performing successive shortcuts: if $(y_1, \dots, y_n)$ is a first periodic pseudo-orbit with jumps in the perturbation domain and if y_i, y_j for some $i < j$ belong to a same set V_k, then $(y_1, \dots, y_i, y_{j-1}, \dots, y_n)$ and $(y_{i+1}, \dots, y_j)$ are two new periodic pseudo-orbits with jumps in the perturbation domain. In the process, we keep one of them and continue with further shortcuts. Note that if the initial orbit satisfies (3), i.e. if n is not a multiple of ℓ, then the periods of the two new orbits cannot be both multiple of ℓ: we can thus choose a new orbit which still satisfies (3).

3.2.2 Primary Shortcuts Avoiding Accumulations in Tiles

In a first step, we perform shortcuts so that the new periodic pseudo-orbit still has jumps in the tiles of the perturbation domain, but intersect each tile at most once: we perform a shortcut each time we have a pair y_i, y_j in a same tile of the perturbation domain.

3.2.3 Construction of Connecting Sequences

We then consider each jump of the obtained periodic pseudo-orbit $(y_1, \dots, y_n)$ at the end of the first step: these are the indices i such that y_i is different from $f^{-1}(y_{i+1})$. By definition the two points belong to a same tile C_i of a domain V_s. One can thus use the property given by Pugh's algebraic lemma and build a connecting sequence $(a_{i,0}, \dots, a_{i,N})$ with $a_{i,0} = y_i$, $x_{i,N} = f^{N-1}(y_{i+1})$, such that for each $0 \leq k \leq N-1$, the distance $d_{i,k} = d(a_{i,k}, f^{-1}(a_{i,k+1}))$ is smaller than η times the distance between $f^k(5/4.C_i)$ and the complement of $f^k(3/2.C_i)$. We then set $r_{i,k} = \theta^{-1}.d_{i,k}$ and introduce the ball $B_{i,k} = B(a_{i,k}, r_{i,k}) \subset f^k(V_s)$. By construction the condition (1) is satisfied, but the different balls $B_{i,k}$ may have nonempty intersection when i varies.

3.2.4 Secondary Shortcuts Avoiding Ball Intersections

Let us now consider the case where two balls $B_{i,k}, B_{j,k'}$ intersect. Note that this has to occur in some domain $f^k(V_s)$ for a given $s \in S$, hence we have $k = k'$. We then perform the shortcut associated to the pair y_i, y_j. Let us assume for instance that one keeps the orbit $(y_1, \dots, y_i, y_{j+1}, \dots, y_n)$ (the other case is similar). As a new connecting sequence between y_i and $f^{N-1}(y_j)$ one introduces

$$(a'_{i,0}, \dots, a'_{i,N}) = (a_{i,0}, \dots, a_{i,k}, a_{j,k+1}, \dots a_{j,N}).$$

In this way all the balls associated to the new sequence but one coincide with balls of the former sequences. Only the new ball $B'_{i,k}$ is different: it has the same center as as $B_{i,k}$ but a larger radius $r'_{i,k}$. Since the distance between $x_{i,k}$ and $f^{-1}(x_{j,k+1})$ is smaller than $2(r_{i,k} + r_{j,k})$, we have

$$r'_{i,k} \leq 2\theta^{-1}.(r_{i,k} + r_{j,k}). \tag{1}$$

3.2.5 The Process Stops

Since the initial length of the pseudo-orbit is finite, the process necessarily stops in finite time. We have however to explain why along the secondary shortcut procedure each ball $B_{i,k}$ does not increase too much and does not leave the sets $f^k(V_s)$.

By construction it is centered at a point $a_{i,k}$ associated to a tile C_i. Let us assume that the radius $r_{i,j}$ is a priori bounded by the distance between $f^k(5/4.C_i)$ and the complement of $f^k(3/2.C_i)$. Since $a_{i,k} \in 5/4.C_i$, the ball $B_{i,k}$ is contained in $3/2.C_i$ and can only intersect the cubes $3/2.C$ such that C and C_i are adjacent tiles. If $B_{i,k}$ intersects $B_{j,k}$, the point $a_{j,k+1}$ is thus associated to a tile adjacent to C_i.

Provided the a priori bound is preserved, the balls centered at $a_{i,k}$ during the process can thus intersect successively at most 4^d other balls coming from adjacent tiles. The diameter of the tiles adjacent to C_i is at most twice the diameter of C_i, hence from (1) after 4^d shortcuts, the diameter of the ball centered at $a_{i,k}$ is bounded from above by $(\theta/4)^{4^d}\eta$ times the distance between $f^k(5/4.C_i)$ and the complement of $f^k(3/2.C_i)$. From our choice of η this gives the a priori estimate.

When the process stops, all the balls are disjoint, hence properties (1) and (2) are satisfied. As we already explained, property (3) is preserved.

3.3 *Proof of Theorem 6*

Let us introduce as before an integer $N \geq 1$ and a chart $\varphi\colon V \to M$ of a neighborhood V of x, given by Pugh's algebraic lemma.

3.3.1 The Non-periodic Case

Let us first assume that x is non-periodic. If V is taken small enough, it is disjoint from its $N - 1$ first iterates. It can also be tiled, so that it defines a perturbation domain and x belongs to the interior of some tile C.

By assumption, there exists $z \in C$ and an iterate $f^n(z) \in C$ with $n \geq 1$ which is not a multiple of ℓ: the sequence $(z, f(z), \dots f^{n-1}(z))$ thus defines a periodic pseudo-orbit which satisfies the property (3). Applying Proposition 4, there exists a pseudo-orbit with jumps in the perturbation domain which satisfies all the properties (1), (2) and (3).

One deduces that there exists a diffeomorphism g in the neighborhood $\mathcal{U}$ of f having a periodic point in V (close to x) whose period is not a multiple of ℓ. By a new perturbation (a conjugacy), one can ensure that this periodic point coincides with x, as required. The proof is the same in $\mathrm{Diff}^1_\omega(M)$. When one gives an open set U containing $\{x, f(x), \dots, f^{N-1}(x)\}$ and $\{z, f(z), \dots f^{n-1}(z)\}$, it also contains the obtained periodic orbit.

3.3.2 The Periodic Case

When x is periodic, it cannot belong to a tiled domain disjoint from a large number of iterates. However from [2, Proposition 4.2], since x is non-resonant, the orbit O of x satisfies the following property (see [2, Definition 3.10]).

Definition 4. A periodic orbit O is *circumventable for* (φ, N) if there exists

- Some arbitrarily small neighborhoods $W^- \subset W^+$ of O,
- An open subset $V' \subset V$ which is a tiled domain of the chart φ,
- Some families of compact sets $\mathcal{D}^-, \mathcal{D}^+$ contained in the interior of the tiles of V',

such that

- Any finite segment of orbit which connects W^- to $M \setminus W^+$ (resp. which connects $M \setminus W^+$ to W^-) has a point in a compact set of $\mathcal{D}^-$ (resp. of $\mathcal{D}^+$),
- For any compact sets $D^+ \in \mathcal{D}^+$, $D^- \in \mathcal{D}^-$, there exists a pseudo-orbit with jumps in the perturbation domain V' which connects D^+ to D^- and is contained in W^+.

Note that one can assume that the period r of p is a multiple of ℓ since otherwise the conclusion of Theorem 6 already holds. One can consider as before $z \in V \cap W^-$ and an iterate $f^n(z) \in V \cap W^-$ with $n \geq 1$ which is not a multiple of ℓ. Let $f^{k^-}(z)$, $f^{k^+}(z)$ be the first and the last iterates $f^k(z)$ of z with $0 \leq k \leq n$ which belong to $V \setminus W^+$. The integers k^- and $n - k^+$ are multiples of r, and hence of ℓ. As a consequence $k^+ - k^-$ is not a multiple of ℓ. By Definition 4, there exist also some iterates z^-, $f^{m_1}(z^-)$ of z which belong respectively to some compact sets $D^- \in \mathcal{D}^-$ and $D^+ \in \mathcal{D}^+$ respectively and such that $m_1 \geq 1$ is not a multiple of ℓ. There also exist a pseudo-orbit $(y_0, \dots, y_{m_2})$ contained in W^+, with jumps in the tiles of V' and such that $y_0 \in D^+$ and $y_{m_2} \in D^-$. In particular, m_2 is a multiple of r, and hence of ℓ. One deduces that the pseudo-orbit $(f(z^-), \dots, f^{m_1}(z^-), y_1, \dots, y_{m_2})$ has jumps in the tiles of the domain V' and its length $m_2 + m_1$ is not a multiple of ℓ.

Applying Proposition 4, there exists a pseudo-orbit with jumps in the perturbation domain which satisfies properties (1), (2) and (3). One concludes as in the non-periodic case.

4 Consequences

We now give the proof of the Theorem 4, which implies Theorems 2 and 3. It combines the classical generic properties and a standard Baire argument.

4.1 The Non-conservative Case

There exists a dense G_δ subset $\mathcal{G} \subset \mathrm{Diff}^1(M)$ of diffeomorphisms f which satisfy:

1. *All the periodic points are hyperbolic.*
2. *Any intersection x between the stable $W^s(O)$ and the unstable manifolds $W^u(O')$ of two hyperbolic periodic orbit is transverse, i.e. $T_xM = T_xW^s(O) + T_xW^u(O')$.*

   These two items together form the Kupka-Smale property, see [12, 20].
3. *Any locally maximal chain-transitive set is a relative homoclinic class.*
4. *If two hyperbolic periodic orbits O, O' are contained in a same chain-transitive set Λ, then by an arbitrarily small C^1-perturbation there exists a cycle between O and O' which is contained in an arbitrarily small neighborhood of Λ.*
5. *Any two hyperbolic periodic orbit with the same stable dimension, contained in a same chain-transitive set Λ, are homoclinically related in any neighborhood of Λ.*

   The three last items are direct consequences of the connecting lemma for pseudo-orbits [3] (see also [8, Theorem 6] for a local version).
6. For any $\ell \geq 1$ and any open set U, let $K_{\ell,U}(f)$ denote the closure of the set of periodic points whose period is not a multiple of ℓ and whose orbit is contained in U.

   If g is C^1-close to f, then $K_{\ell,U}(g)$ is contained in a small neighborhood of $K_{\ell,U}(f)$.

   Proof. When all the periodic orbits of f are hyperbolic (or more generally have no eigenvalue equal to 1), f is a lower-semi-continuity point of the map ${}_{\ell,U}: g \mapsto K_{\ell,U}(g)$ for the Hausdorff topology. One deduces from Baire's theorem that $K_{\ell,U}$ is continuous in restriction to a dense G_δ subset $\mathcal{G}_0 \subset \mathrm{Diff}^1(M)$. If $f \in \mathcal{G}_0$ does not satisfy the item 6, then there exists a point $x \notin K_{\ell,U}(f)$ which is arbitrarily close to a periodic point p of a diffeomorphism g close to f and whose period is not a multiple of ℓ. By a small perturbation, one can assume that the periodic point p has no eigenvalue equal to 1, and hence one can replace g by any diffeomorphism close: taking $g \in \mathcal{G}_0$, one contradicts the continuity of $K_{\ell,U}$ on $\mathcal{G}_0$. This proves the property.

7. *For any hyperbolic periodic orbit O, any neighborhood U of O and any diffeomorphism g C^1-close to f, the relative homoclinic class $H(O_g, U)$ of g has the same periods as $H(O, U)$.*

   Indeed we noticed in Sect. 2.4 that the period map $g \mapsto \ell(O_g)$ is upper-semi-continuous, and hence is locally constant on an open and dense subset of $\mathrm{Diff}^1(M)$.

We now fix $f \in \mathcal{G}$ and a locally maximal chain-transitive set $\Lambda = \cap_{i\in\mathbb{Z}} f^i(U)$ in an open set U. By item 3, Λ is a relative homoclinic class $H(O, U)$. Then by Proposition 2, the set Λ admits an invariant decomposition into compact sets

$$\Lambda = \Lambda_1 \cup \cdots \cup \Lambda_\ell,$$

such that for each i, the restriction of f^ℓ to Λ_i is topologically mixing, Λ_i coincides with the pointwise relative homoclinic class of a point of O in U, and ℓ is the period of the relative homoclinic class.

Let us assume by contradiction that Λ_1 and Λ_i intersect for some $1 < i \le \ell$ at a point x. If x is periodic, then it is hyperbolic by item 1. Since f^ℓ is transitive in Λ_1, one deduces that for any neighborhood V of x there exists $k \ge 0$ and a segment of orbit $y, f(y), \dots, f^{k\ell+j}(y)$ in U with endpoints in V. One can thus apply Theorem 6 and by a C^1-perturbation build a periodic point arbitrarily close to x, whose period is not a multiple of ℓ and whose orbit is contained in U. From the item 6, this shows that $K_{\ell,U}(f)$ contains x. Since Λ is the locally maximal invariant set in U, this shows that it contains a periodic orbit O' whose period is not a multiple of ℓ and which is hyperbolic by the item 1. From the item 4, one can create by an arbitrarily small perturbation a cycle between O and O'. From item 7 and Proposition 3 the period of O' is contained in the set of periods $\ell(O).\mathbb{Z}$, a contradiction. The sets Λ_i are thus pairwise disjoint.

The uniqueness of the decomposition is easy: considering any small open set V intersecting $H(O, U)$, then a large iterate $f^n(V)$ meets $V \cap H(O, U)$ if and only if n is a multiple of ℓ. Moreover the closure of $\bigcup_{k\ge k_0} f^{k\ell}(V \cap H(O, U))$, for k_0 large, coincides with one of the sets Λ_i. We have thus obtained the main conclusion of Theorem 4.

Let us consider two hyperbolic periodic points $p, q \in \Lambda_1$ having the same stable dimension. By item 5, their orbits are homoclinically related in U. The previous discussion shows that $\Lambda_1 = h(p, U) = h(q, U)$ and ℓ is the minimal positive integer such that $(W^u(f^\ell(p)) \pitchfork W^s(p)) \cap \Lambda$ is non-empty. By item 2, the intersections between $W^u(f^\ell(p))$ and $W^s(p)$ are all transverse, giving the first item of the Theorem 4.

There exists a transverse intersection point in Λ between $W^u(p)$ and an iterate $W^s(f^k(q))$. Using the fact that the decomposition of the theorem is an invariant partition into disjoint compact sets, one deduces that $f^k(q)$ belongs to Λ_1 and that k is a multiple of ℓ. By Proposition 1, this implies that $(W^u(p) \pitchfork W^s(q)) \cap \Lambda$ is non-empty. Lemma 2 and item 2 now show that $\Lambda_1 = h(p)$ is the closure of $(W^u(p) \pitchfork W^s(q)) \cap \Lambda = W^u(p) \cap W^s(q) \cap \Lambda$, proving the second item of the theorem.

4.2 The Conservative Case

When $\dim(M) \geq 3$ and ω is a volume form, the previous proof goes through. In the other cases ω is a symplectic form and the item 1 may fail. However the same proof can be done replacing the item 1 by the properties 1' and 1" below.

Any diffeomorphism in a dense G_δ subset $\mathcal{G}_\omega \subset \text{Diff}^1_\omega(M)$ satisfies the items 2–7 and moreover:

1' *All the periodic points are non-resonant.*

1" *Any neighborhood of a periodic orbit O contains a hyperbolic periodic orbit O'. Consider $\ell \geq 1$. If the period of O is not a multiple of ℓ, then the same holds for O'.*

Proof (Proof). By [18], there exists a dense G_δ subset $\mathcal{G}'_\omega$ of diffeomorphisms satisfying the item 1'.

One may then use similar arguments as in [14, Proposition 3.1]. Consider any non-hyperbolic periodic point x of a diffeomorphism f, with some period r. Using generating functions, it is possible to build a diffeomorphism $\tilde{f} \in \text{Diff}^1_\omega(M)$ that is C^1-close to f such that the dynamics of $\tilde{f}^r$ in a neighborhood of x is conjugated to a non-hyperbolic linear symplectic map A which is diagonalizable over $\mathbb{C}$.

Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of A with modulus one. One can assume moreover that they have the form $e^{2i\pi p_k/q_k}$ where $\ell \wedge q_k = 1$. One deduces that x is the limit of periodic points y whose minimal period is $r.L$ where L is the least common multiple of the q_k. The tangent map at y at the period coincides with the identity on its central part. Consequently, one can by a small perturbation turn y to a hyperbolic periodic point. This shows that a diffeomorphism arbitrarily C^1-close to f in $\text{Diff}^1_\omega(M)$ has a hyperbolic periodic orbit contained in an arbitrarily small neighborhood of the orbit of x and having a period which is not a multiple of ℓ.

We end with a Baire argument. For $n, \ell \geq 1$, let us denote by $D_{n,\ell} \subset \mathcal{G}'_\omega$ the subset of diffeomorphisms whose periodic orbits of period less than n which are not a multiple of ℓ are $1/n$-approximated by hyperbolic periodic orbits whose period is not a multiple of ℓ. Since the periodic points of period less than n are finite and vary continuously with the diffeomorphism, this set is open; it is dense by the first part of the proof. We then set $\mathcal{G}_\omega = \bigcap_{n,\ell} D_{n,\ell}$.

References

1. F. Abdenur, A. Avila, J. Bochi, Robust transitivity and topological mixing for C^1-flows. Proc. Am. Math. Soc. **132**, 699–705 (2004)
2. M.-C. Arnaud, C. Bonatti, S. Crovisier, Dynamiques symplectiques génériques. Ergod. Theory Dyn. Syst. **25**, 1401–1436 (2005)
3. C. Bonatti, S. Crovisier, Récurrence et généricité. Invent. Math. **158**, 33–104 (2004)
4. C. Bonatti, L.J. Díaz, Persistent nonhyperbolic transitive diffeomorphisms. Ann. Math. **143**, 357–396 (1996)

5. C. Bonatti, L. J. Díaz, R. Ures, Minimality of strong stable and unstable foliations for partially hyperbolic diffeomorphisms. J. Inst. Math. Jussieu **4**, 513–541 (2002)
6. R. Bowen, Periodic points and measures for Axiom A diffeomorphisms. Trans. Am. Math. Soc. **154**, 377–397 (1971)
7. C. Conley, *Isolated Invariant Sets and the Morse Index*. CBMS Regional Conference Series in Mathematics, vol. 38 (American Mathematical Society, Providence, 1978)
8. S. Crovisier, Periodic orbits and chain-transitive sets of C^1-diffeomorphisms. Publ. Math. Inst. Hautes Études Sci. **104**, 87–141 (2006)
9. S. Crovisier, Perturbations of C^1-diffeomorphisms and dynamics of generic conservative diffeomorphisms of surface. Panor. Synth. **21**, 1–33 (2006). ArXiv:1011.4692
10. S. Hayashi, Connecting invariant manifolds and the solution of the C^1-stability and Ω-stability conjectures for flows. Ann. Math. **145**, 81–137 (1997); **150**, 353–356 (1999)
11. F.R. Hertz, M.A.R. Hertz, R. Ures, Some results on the integrability of the center bundle for partially hyperbolic diffeomorphisms. Fields Inst. Commun. **51**, 103–109 (2007)
12. I. Kupka, Contributions à la théorie des champs génériques. Contrib. Differ. Equ. **2**, 457–484 (1963)
13. S. Newhouse, Hyperbolic limit sets. Trans. Am. Math. Soc. **167**, 125–150 (1972)
14. S. Newhouse, Quasi-elliptic periodic points in conservative dynamical systems. Am. J. Math. **99**, 1061–1087 (1977)
15. J. Palis, On Morse-Smale dynamical systems. Topology **8**, 385–404 (1968)
16. C. Pugh, The closing lemma. Am. J. Math. **89**, 956–1009 (1967)
17. C. Pugh, C. Robinson, The C^1 closing lemma, including Hamiltonians. Ergod. Theory Dyn. Syst. **3**, 261–313 (1983)
18. C. Robinson, Generic properties of conservative systems, I and II. Am. J. Math. **92**, 562–603, 897–906 (1970).
19. M. Shub, Topological transitive diffeomorphism on T^4. Lect. Notes Math. **206**, 39–40 (1971)
20. S. Smale, Stable manifolds for differential equations and diffeomorphisms. Ann. Scuola Norm. Sup. Pisa **17**, 97–116 (1963)
21. S. Smale, Diffeomorphisms with many periodic points, in *Differential and Combinatorial Topology* (Princeton University Press, Princeton, 1965), pp. 63–80
22. S. Smale, Differentiable dynamical systems. Bull. Am. Math. Soc. **73**, 747–817 (1967)

Recent Results on the Size of Critical Sets

Dorin Andrica and Cornel Pintea

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract In the first part of this survey we review some special cases of $\varphi_{\mathcal{F}}$-category of a pair (M, N) of manifolds such as φ-category, Morse-Smale characteristic, and Morse-Smale characteristic for circular functions. Section 2 presents examples of pairs with finite φ, and Sect. 3 provides lower estimates for the size of the critical sets in terms of topological dimension. We employ the cardinality when the manifolds admit maps with finitely many critical points and the topological dimension when no such maps exist.

1 Introduction

Let M^m, N^n be smooth manifolds and $\mathcal{F} \subseteq C^\infty(M, N)$ be a family of smooth mappings. The *$\varphi_{\mathcal{F}}$-category* of pair (M, N) is defined by

$$\varphi_{\mathcal{F}}(M, N) = \min\{\mu(f) : \ f \in \mathcal{F}\}\ , \tag{1}$$

D. Andrica (✉)
King Saud University, College of Science, Riyadh, Saudi Arabia

Department of Mathematics and Computer Science, Babeş-Bolyai University, Str. Mihail Kogalniceănu nr. 1 400084, Cluj-Napoca, Romania
e-mail: dandrica@math.ubbcluj.ro;dandrica@math.ksu.sa

C. Pintea
Department of Mathematics and Computer Science, Babeş-Bolyai University, Str. Mihail Kogalniceănu nr. 1 400084, Cluj-Napoca, Romania
e-mail: cpintea@math.ubbcluj.ro

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_2, © Springer-Verlag Berlin Heidelberg 2012

where $\mu(f)$ stands for the cardinality of the critical set $C(f)$ of $f : M \to N$. It is clear that $0 \leq \varphi_{\mathcal{F}}(M,N) \leq +\infty$ and we have $\varphi_{\mathcal{F}}(M,N) = 0$ if and only if family $\mathcal{F}$ contains immersions (if $m < n$), submersions (if $m > n$) or local diffeomorphisms (if $m = n$).

In the following we shall point out some important particular cases as they are presented in the book [1, pp. 145–147].

1. Let us consider $\mathcal{F} = C^{\infty}(M,N)$. Then $\varphi_{\mathcal{F}}(M,N)$ represents the φ-category of pair (M,N) and it is simply denoted by $\varphi(M,N)$. Remark that $\varphi(M,N)$ point out a class of mappings $f \in C^{\infty}(M,N)$ having a minimal critical set $C(f)$. In a series of papers Church, and Church and Timourian [10–12] using results from dimension theory, studied the mappings $f \in C^{\infty}(M,N)$ with small critical set. Some properties of the invariant $\varphi(M,N)$ are given in Sect. 4.4 of the book [1].
2. Consider the case when $N = \mathbb{R}$, the real line, and the family $\mathcal{F}$ is given by $\mathcal{F}(M) = C^{\infty}(M,\mathbb{R})$, the algebra of all smooth real functions defined on M. In this situation $\varphi_{\mathcal{F}}(M,\mathbb{R})$ represents the φ-category of M (or the functional category) and it is denoted by $\varphi(M)$. The invariant $\varphi(M)$ was first investigated by Takens. The effective computation of $\varphi(M)$ is a difficult problem (see also Sect. 4.3 of the book [1]).

   It is interesting to remark that we have not an example of closed manifold M^m such that $cat(M) < \varphi(M)$ and also the equality $cat(M) = \varphi(M)$ is proved only for some isolated classes of manifolds. Note that $cat(N)$ stands for the Lusternik-Schnirelmann category of M [1]. To understand the difficulty of the problem if $cat(M) = \varphi(M)$ for every closed manifold let us look only to the following particular situation: $cat(M) = \varphi(M) = 2$. From $cat(M) = 2$ one obtains that M is a homotopic sphere. Taking into account the well-known Reeb's result, from the equality $\varphi(M) = 2$ it follows that M is a topological m-sphere. Therefore, the equality $cat(M) = \varphi(M) = 2$ is equivalent to the Poincaré conjecture. According to the fact that Poincaré conjecture was proved to be true, it follows that for any closed manifold with $cat(M) = 2$ we have $\varphi(M) = 2$.
3. Let us consider $N = \mathbb{R}$, $\mathcal{F} = \mathcal{F}_m(M) \subset C^{\infty}(M,\mathbb{R})$, the set of all Morse functions defined on M. In this case one obtains $\varphi_{\mathcal{F}}(M,\mathbb{R}) = \gamma(M)$, the *Morse-Smale characteristic* of manifold M, an important invariant of M intensively studied by many authors. We mention here only the papers [21,22] and Sect. 4.1–4.2 of [1]. An important situation when the Morse-Smale characteristic can be computed in terms of the homology of M is given by the situation when the manifold M is simply-connected of dimension > 5. This property was proved in the celebrated paper of Smale [26]. Efforts have been made to generalize Smale's result to the case that manifold M is not simply-connected. For example, Sharko [25] proved that still it is possible to compute the Morse-Smale characteristic of the manifolds with infinite cyclic fundamental group. But a complete answer for general M is not known.
4. Using the notations above, consider $N = S^1$ and the family $\mathcal{F} = \mathcal{F}_m(M,S^1) \subseteq C^{\infty}(M,S^1)$, given by the set of all circle-valued Morse functions defined on M.

The systematic study of circle-valued Morse functions was initiated by Novikov in 1980. The motivation came from a problem in hydrodynamics, where the application of the variational approach led to a multi-valued Lagrangian. In this case we denote $\varphi_{\mathcal{F}}(M, S^1)$ by $\gamma_{S^1}(M)$, and we call it the *Morse-Smale characteristic* of manifold M for circle-valued Morse functions $M \to S^1$. So, in this situation we have

$$\gamma_{S^1}(M) = min\{\mu(f) : f \in \mathcal{F}_m(M, S^1)\}. \tag{2}$$

The invariant $\gamma_{S^1}(M)$ is quite different than $\gamma(M)$, and it is better reflected by the structure of the fundamental group $\pi_1(M)$. The computation of $\gamma_{S^1}(M)$ represents an interesting problem which is related to the topology and the geometry of the closed one-forms on manifold M.

5. Let G be a compact Lie group which act on the manifolds M^m, N^n. Recall that the mapping $f : M \to N$ is said to be G-equivariant if $f(gp) = gf(p))$ for all $p \in M$ and all $g \in G$. Observe that the critical set of such a map is invariant under the action of G on M, as the rank of equivariant maps is actually constant on orbits within the source manifold M under the action of G on M. In other words, the critical sets of G-equivariant maps are unions of orbits, usually called *critical orbits* of f, inside M under the action of G on M. This is the reason to work, within this context, with the cardinality of the set of critical orbits $C(f)/G$ rather than the cardinality of the critical set itself. Therefore one defines the *G-equivariant φ-category* of the pair (M, N) as

$$\varphi_G(M, N) := \min\{\text{card}(C(f)/G) : f \in C^\infty_G(M, N)\},$$

where $\mathcal{F} := C^\infty_G(M, N)$ stands for the family of all smooth equivariant mappings from M to N. In the case of the trivial action of G on the target manifold N, note that the G-equivariant maps are actually the invariant ones and $N/G \equiv N$ in this case. In the case of free actions of G on both manifolds one could try to compare $\varphi(M/G, N/G)$ with $\varphi_G(M, N)$ as well as $\varphi(M/G, N)$ with $\varphi_G(M, N)$ when G acts freely on M and trivially on N.

Let $\mathcal{F} \subseteq C^\infty(M, N)$ be a family of smooth mappings $M \to N$. Taking into account the definition of $\varphi_{\mathcal{F}}(M, N)$, one obtains $\varphi(M, N) \leq \varphi_{\mathcal{F}}(M, N)$, where $\varphi(M, N)$ is the φ-category of (M, N) related to the family $C^\infty(M, N)$ (see the particular case (1)). In some papers we gave sufficient conditions expressed in terms of some topological invariants of manifolds M and N in order to have $\varphi(M, N) = +\infty$ (see Sects. 4.4 and 4.5 of [1]). These conditions imply that $\varphi_{\mathcal{F}}(M, N) = +\infty$, for any family $\mathcal{F} \subseteq C^\infty(M, N)$.

In Sect. 2 we review our the last results involving the φ-category of a pair of surfaces and spheres, obtained in the paper [2, 4]. One of the main purposes of this section is to present two constructions for a mapping $M \to S^2$ having three critical points, where M is an orientable surface which is not diffeomorphic to the 2-dimensional sphere S^2. The last subsection contains some additional results to the paper [3]. Also, we present some recent results on φ-category of some pair of manifolds which admit maps with finitely many critical points. Pairs of compact surfaces as well as connected sums of products of spheres are investigated from this point of view and it reflects the contents of the works [2–4, 14]. In Sect. 3 we review recent theorems on maps with high dimensional critical sets. In the zero codimension case, the main class of such maps consists in those of zero degree, while the higher codimension case is represented by partial results inspired by the zero codimension one. This section reflects the contents of the works [19, 20].

2 Functions with Finitely Many Critical Points

2.1 The φ-Category of a Pair of Surfaces and Spheres

Denote by $[[r]]$ the smallest integer greater than or equal to r and by $[x]$ the largest number smaller or equal to x. Write, when $\chi(N) < 0$:

$$|\chi(M)| = d|\chi(N)| + v, \text{where } d, v \in \mathbb{Z}_+,\ 0 \leq v < |\chi(N)|$$

so that

$$d = \left[\frac{\chi(M)}{\chi(N)}\right]$$

The following result was proved in paper [2].

Theorem 1. *Let M and N be connected closed orientable surfaces.*

1. *If $\chi(M) > \chi(N)$ then $\varphi(M, N) = \infty$.*
2. *If $M \neq S^2$ then $\varphi(M, S^2) = 3$.*
3. *If N is the torus $S^1 \times S^1$ then*

$$\varphi(M, S^1 \times S^1) = \begin{cases} 1, & \text{if } \chi(M) < 0 \\ 0, & \text{if } M = S^1 \times S^1 \\ \infty, & \text{if } M = S^2 \end{cases}$$

4. *If $\chi(N) < 0$, then*

$$\varphi(M, N) = \begin{cases} \left[\left[\dfrac{v}{d-1}\right]\right], & \text{if } d \geq 2 \\ \infty, & \text{otherwise} \end{cases}$$

The situation when at least one of surfaces M and N is not orientable was studied in the paper [4]. The main result proved in [4] is the following:

Theorem 2. *1. Suppose that M and N are nonorientable.*

a. If $N = \mathbb{RP}^2$ then

$$\varphi(M, \mathbb{RP}^2) = \begin{cases} 0, \text{ if } M = \mathbb{RP}^2 \\ 2, \text{ otherwise} \end{cases}$$

b. When N is the Klein bottle we have

$$\varphi(M, \mathbb{RP}^2\#\mathbb{RP}^2) = \begin{cases} 1, & \text{if } \chi(M) \equiv 0 \pmod 2 \\ 0, & \text{if } M = \mathbb{RP}^2\#\mathbb{RP}^2 \\ \infty, & \text{if } \chi(M) \not\equiv 0 \pmod 2 \end{cases}$$

c. Assume from now on that $\chi(N) < 0$, so N is not $\mathbb{RP}^2$ nor the Klein bottle.

i. If $\chi(M) \geq 2\chi(N)$ then

$$\varphi(M, N) = \begin{cases} 0, & \text{if } \chi(M) = 2\chi(N) \\ \infty, & \text{otherwise} \end{cases}$$

ii. If $\chi(M) < 2\chi(N)$.

(A) Assume that $\chi(N) \equiv 0 \pmod 2$. Then

$$\varphi(M, N) = \begin{cases} \left[\left[\dfrac{v}{d-1}\right]\right], & \text{if } \chi(M) \equiv \chi(N) \equiv 0 \pmod 2 \\ \infty, & \text{if } \chi(M) \equiv 1 \pmod 2, \chi(N) \equiv 0 \pmod 2 \end{cases}$$

(B) Suppose that $\chi(N) \equiv 1 \pmod 2$. Then

$$\varphi(M, N) = \begin{cases} \left[\left[\dfrac{v}{d-1}\right]\right], & \text{if } d \equiv \chi(M) \pmod 2 \\ \max\left(\left[\left[\dfrac{v}{d-1}\right]\right], & \text{if } d \not\equiv \chi(M) \pmod 2, \ d \geq 3 \\ \quad \left[\left[\dfrac{v+|\chi(N)|}{d}\right]\right]\right), & \\ \infty, & \text{if } d \not\equiv \chi(M) \pmod 2, \ d = 2 \end{cases}$$

d. Suppose that M is nonorientable and N is orientable. Then $\varphi(M, N) = \infty$.

e. M is orientable and N is not orientable.

i. If $\chi(N) < 0$

$$\varphi(M,N) = \begin{cases} \max\left(\left[\left[\frac{v}{d-1}\right]\right], \left[\left[\frac{v+|\chi(N)|}{d}\right]\right]\right), & \text{if } d \text{ is odd } d \geq 5 \\ \left[\left[\frac{v}{d-1}\right]\right], & \text{if } d \text{ is even } d \geq 4 \\ 0, & \text{if } M = \widehat{N} \\ \infty, & \text{if } M \neq \widehat{N}, d \leq 3 \end{cases}$$

ii. If $N = \mathbb{RP}^2$ *then*

$$\varphi(M, \mathbb{RP}^2) = \begin{cases} 2, \text{ if } \chi(M) \leq 0 \\ 0, \text{ if } M = S^2 \end{cases}$$

iii. If $N = \mathbb{RP}^2 \# \mathbb{RP}^2$ *then*

$$\varphi(M, \mathbb{RP}^2 \# \mathbb{RP}^2) = \begin{cases} 1, & \text{if } \chi(M) < 0 \\ 0, & \text{if } M = S^1 \times S^1 \\ \infty, & \text{if } M = S^2 \end{cases}$$

Remark 1. Computations were previously done for orientable surfaces in [2]. In the paper [18] it was proved that $\varphi(Y, X)$ is infinite when $\chi(Y) > \chi(X)$. In the later case it was actually proved in the paper [19] that $\dim(C(f)) \geq 1$ for every differentiable map $f : X \longrightarrow Y$ and $\dim(B(f)) = 1$ whenever $R(f) \neq \emptyset$ (See also Example 2(1) of this work).

Moreover, recent results of Bogatyi, Kudryatseva, and Zieschang [7,8] show that the minimal number of critical points can be achieved by using maps $f : Y \to X$ which are primitive branched coverings, i.e. mappings inducing surjective maps at the level of fundamental groups.

The computation of φ-category for any pair of spheres is a very difficult problem. It is clear that we have $\varphi(S^m, S^n) = 0$ if $m \leq n$. Also, it is not difficult to show that $\varphi(S^m, S^1) = 2$ if $m \geq 2$. Some partial results when $m > n \geq 2$ were obtained in the paper [2] and they are contained in:

Theorem 3. *1. The values of* $m > n \geq 2$ *for which* $\varphi(S^m, S^n) = 0$ *are exactly those arising in the Hopf fibrations, that is,* $n \in \{2, 4, 8\}$ *and* $m = 2n - 1$.

2. One has $\varphi(S^4, S^3) = \varphi(S^8, S^5) = \varphi(S^{16}, S^9) = 2$.

3. If $m \leq 2n - 3$, *then* $\varphi(S^m, S^n) = \infty$.

4. If $\varphi(S^{2n-2}, S^n)$ *is finite, then* $n \in \{2, 3, 5, 9\}$.

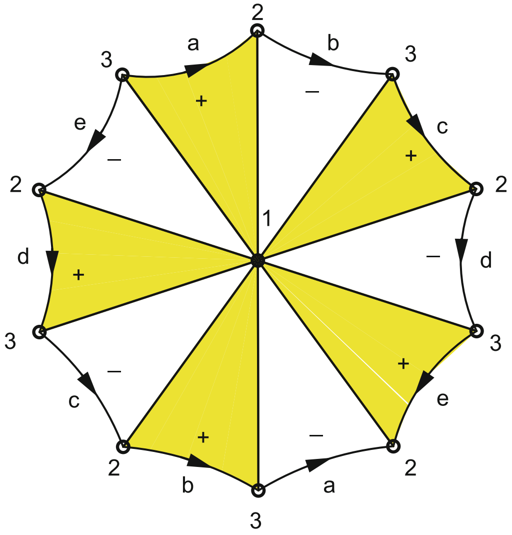
2.2 *The Construction of a Mapping with Three Critical Points from M to S^2*

In [2] we treated separately the case $N = S^2$, by making use of Belyi maps (see [24]), but the proof was rather sketchy. Morris Hirsch asked us for more details and later gave us the following simple proof using triangulations.

The algebraic topology arguments in [2] show that $\varphi(\Sigma, S^2) \geq 3$ and it suffices to see that there exists a Belyi map having precisely three critical points, namely one critical point above each critical value.

There exist triangulations of the surface M with any number of vertices $s \geq 1$, in particular for $s = 3$. In fact, we fix the vertices and then add, inductively, a number of disjoint arcs joining the vertices such that no two arcs be homotopically equivalent, by a homotopy keeping the endpoints fixed. Consider the maximal collection of pairwise nonhomotopic such arcs. Moreover the complementary regions are triangles, since otherwise we can add more arcs, contradicting the maximality. We obtained, therefore, a triangulation of M having s vertices, $2s - 2\chi(M)$ triangles and $3s - 3\chi(M)$ edges. Although each cell has its vertices among the three vertices of the triangulation, they are not necessarily distinct.

However, we need a special triangulation, namely one in which all triangles have the same (distinct) three vertices. Given $n \geq 1$ we consider the regular polygon in the hyperbolic plane (respectively the Euclidean plane, when $n = 1$) with $2(2n+1)$ vertices and angles $\dfrac{2\pi}{2n+1}$. We identify the opposite edges by means of isometries reversing the orientation. There are then two orbits of the vertices, and around each vertex the total angle is 2π. We obtain then a closed hyperbolic surface.



Let us subdivide it into equal triangles with a common vertex at the center of the polygon. This triangulation has three vertices, $2(2n+1)$ triangles and $3(2n+1)$ edges and thus the surface has genus n. Label the central vertex by 1 and the two other vertices by 2 and 3. Then each triangle of the triangulation has the vertices 1–3. The hyperbolic surface is oriented and thus each triangle inherits an orientation.

We say that a triangle is positive if the cyclic order of the labels of its vertices is 1–3 and negative otherwise. Observe that adjacent triangles have distinct signs since the order of 2, 3 is reversed. Consider next the triangulation of S^2 consisting of two triangles whose boundaries are identified. Now, map the triangulation of M onto that of the sphere S^2 by mapping each triangle of M to one or the other triangles of the sphere according to the sign. This yields a map $M \to S^2$ which is ramified only at the three vertices.

There is also a beautiful example due to John Hubbard of a Belyi function with only three critical points, obtained by considering $\Sigma \subset \mathbb{CP}^2$ to be the projective algebraic curve determined by the (non-homogeneous) equation

$$y^{2g+1} = x^2 - 1.$$

The projection onto the first coordinate is a holomorphic map $\Sigma \to \mathbb{CP}^1$, which is ramified over $1, -1$ and ∞ and has three critical points. By the Riemann–Hurwitz formula we have $\chi(\Sigma) = 2 - 2g$.

One can seek for the topological classification of smooth functions $f : M \to S^2$ with three critical points, i.e. up to the action by left multiplication by diffeomorphisms of M. The pull-back by f of the triangulation of S^2 consisting of two triangles (with vertices at critical values) is a special triangulation of M, in which triangles can be given signs, according to the triangle covered on the sphere. If we label the vertices by 1–3 then the sign of a triangle corresponds to the cyclic order of the boundary labels. Further, if we fix a vertex, say the one labeled 1, and look at the edges incident to it, then their endpoints are labeled only by 2 and 3; moreover, consecutive edges correspond to different labels, and thus the cyclic order of these labels is an alternate sequence $2, 3, 2, \dots, 3$. The union of these triangles is a fundamental polygon P for the surface M. Thus P has $2(2n + 1)$ edges, where n is the genus of M. In particular P is the polygon drawn above.

Furthermore, one obtains M by identifying the edges of P by means of an involution j. The involution h should satisfy the following conditions:

1. j reverses the orientation of the edges inherited from the circle, such that the quotient is orientable;
2. j preserves the labels;
3. The orbit of a vertex by the permutation group generated by j is the set of all vertices with the same label; this means that there are precisely two vertices in the quotient M.
4. Adjacent edges are not identified by j.

Thus, up to a homeomorphism of M the special triangulations of M correspond to polygons with an involution j as above. By direct inspection it follows that there are not any other involutions j except the standard one from above, when the genus is at most 3.

2.3 Further Examples of Maps with Finitely Many Critical Points

Recall from the paper [2] that $\varphi(S^{2n}, S^{n+1}) = 2$, if $n = 2, 4$ or 8. This equality is realized by taking suspensions of both spaces in the Hopf fibration $h : S^{2n-1} \to S^n$, where $n = 2, 4$ or 8, and then smoothing the new map at both ends. The extension $H : S^{2n} \to S^{n+1}$ has precisely two critical points. This is also the basic example of a Montgomery-Samelson fibration with finitely many critical points, as considered in [5]. Our aim in this section is to define fiber sums of Hopf fibrations leading to other examples of pairs of manifolds with finite φ and it reflects the content of the work [14]. A characterization of the closed $2k$-manifolds admitting smooth maps onto $(k + 1)$-manifolds, $k \in \{2, 4\}$, with finitely many critical points is provided by Funar in [13], where he also provides an upper bound for the φ-category of such pairs.

Identify S^{n+1} (and respectively S^{2n}) with the suspension of S^n (respectively S^{2n-1}) and thus equip it with the coordinates (x, t), where $|x|^2 + t^2 = 1$, and $t \in [-1, 1]$. We call the coordinate t the height of the respective point. The suspension H is then given by:

$$H(x, t) = \left(\psi(|x|)h\left(\frac{x}{|x|}\right), t\right),$$

where $\psi : [0, 1] \longrightarrow [0, 1]$ is a smooth increasing function infinitely flat at 0 such that $\psi(0) = 0$ and $\psi(1) = 1$.

Pick-up a number of points $x_1, x_2, \dots, x_k \in S^{n+1}$ and their small enough disk neighborhoods $x_i \in D_i \subset S^{n+1}$, such that:

1. The projections of D_i on the height coordinate axis are disjoint;
2. The D_i's do not contain the two poles, i.e. their projections on the height axis are included in the open interval $(-1, 1)$.

Let A_k be the manifold with boundary obtained by deleting from S^{n+1} of the interiors of the disks D_i, $1 \leq i \leq k$. Let also B_k denote the preimage $H^{-1}(A_k) \subset S^{2n}$ by the suspended Hopf map. Since H restricts to trivial fibration over the disks D_i it follows that B_k is a manifold, each one of its boundary component being diffeomorphic to $S^{n-1} \times S^n$. Moreover, the boundary components are endowed with a natural trivialization induced from D_i.

Let now consider a finite connected graph Γ. To each vertex v of valence k we associate a block $(B_v, A_v, H|_{B_v})$ which will be denoted $(B_k, A_k, H|_{B_k})$ when we want to emphasize the dependence on the number of boundary components. Each boundary component of A_v or B_v corresponds to an edge incident to the vertex v. We define the fiber sum along Γ as the following triple $(B_\Gamma, A_\Gamma, H_\Gamma)$, where:

1. A_Γ is the result of gluing the manifolds with boundary A_v, associated to the vertices v of Γ, by identifying for each edge e joining the vertices v and w (which might coincide) the pair of boundary components in A_v and A_w corresponding to the edge e. The identification is made by using an orientation-reversing diffeomorphism of the boundary spheres.

2. B_Γ is the result of gluing the manifolds with boundary B_v, associated to the vertices v of Γ, by identifying for each edge e joining the vertices v and w (which might coincide), the boundary components in B_v and B_w corresponding to the pair of boundary components in A_Γ associated to the edge e. Gluings in B_Γ are realized by some orientation-reversing diffeomorphisms with respect the product structure over boundaries of A_v and A_w. We choose the identification diffeomorphism $\nu : \partial B_v \longrightarrow \partial B_w$ to be one from the construction of the double of B_v.
3. As the boundary components are identified, the natural trivializations of the boundary components of B_v agree in pairs. Thus the maps H_v induce a well-defined map $H_\Gamma : B_\Gamma \to A_\Gamma$.

Proposition 1. *The map $H_\Gamma : B_\Gamma \to A_\Gamma$ has $2m$ critical points, where m is the number of vertices of Γ.*

Proposition 2. *If Γ has e edges and c cycles, i.e. $e - c + 1$ vertices, then B_Γ is diffeomorphic to $\Sigma^{2n}\#_e S^n \times S^n \#_c S^1 \times S^{2n-1}$ (where Σ^{2n} is a homotopy sphere which is trivial when $n = 2$), while A_Γ is diffeomorphic to $\#_c S^1 \times S^n$. For $c = 0$, the notation $\#_c S^1 \times S^n$ stands for actually S^{n+1}.*

Corollary 1. *Let $n \in \{2, 4, 8\}$, $e \geq c \geq 0$ with $c \neq 1$ and Σ^{2n} be a homotopy $2n$-sphere. If $n = 2$, assume further that $\Sigma^4 \setminus \mathrm{int}(D^4)$ embeds smoothly in S^4. Then*

$$\varphi(\Sigma^{2n}\, \#_e S^n \times S^n \#_c S^1 \times S^{2n-1}, \#_c S^1 \times S^n) \leq 2e - 2c + 2.$$

The opposite inequality works for every dimension and is based on some elementary algebraic topology arguments, such as:

Proposition 3. *Let M^{2n} and N^{n+1} be closed manifolds and $n \geq 2$. Assume that $\pi_1(M) \cong \pi_1(N)$ is a free group $\mathbb{F}(c)$ with c generators, $c \neq 1$ (with $\mathbb{F}(0) = 0$), and that $\pi_j(M) = \pi_j(N) = 0$, for $2 \leq j \leq n-1$ and $H_{n-1}(M) = 0$. Then $\varphi(M, N) \geq \beta_n(M) - 2c + 2$, where β_k denotes the k-th Betti number.*

Corollary 2. *For any dimension $n \geq 2$ and $e, c \in \mathbb{Z}_+$, with $c \neq 1$ we have*

$$\varphi(\Sigma^{2n}\#_e S^n \times S^n \#_c S^1 \times S^{2n-1}, \#_c S^1 \times S^n) \geq 2e - 2c + 2$$

Here $\#_c S^1 \times S^n = S^{n+1}$ when $c = 0$ and $\#_e S^n \times S^n \#_c S^1 \times S^{2n-1} = S^{2n}$ when $e = c = 0$ and Σ^{2n} denotes a homotopy $2n$-sphere, for $n \neq 2$.

As a final conclusion of this section we have the following result:

Theorem 4. *Let $n \in \{2, 4, 8\}$, $e \geq c \geq 0$ with $c \neq 1$ and Σ^{2n} be a homotopy $2n$-sphere. If $n = 2$, assume further that $\Sigma^4 \setminus \mathrm{int}(D^4)$ embeds smoothly in S^4. Then*

$$\varphi(\Sigma^{2n}\#_e S^n \times S^n \#_c S^1 \times S^{2n-1}, \#_c S^1 \times S^n) = 2e - 2c + 2$$

Here $\#_c S^1 \times S^n = S^{n+1}$ when $c = 0$, and $\#_e S^n \times S^n \#_c S^1 \times S^{2n-1} = S^{2n}$ when $e = c = 0$.

3 Mappings with Large Critical Sets: Mappings of Zero Degree

While the evaluation tool for the size of critical sets within the previous sections is the *cardinality*, the maps of zero degree between compact oriented manifolds have all infinitely many critical points, which makes the *cardinality* unsuitable to evaluate the size of critical sets of such maps. The zero degree maps have actually high dimensional critical sets and make the *topological dimension* a more appropriate evaluation tool to measure their size. On the other hand, the remarkable Sard theorem [23] ensures us that the set of critical values have zero measure and indirectly points out that the measure cannot distinguish the sets of critical values of *different* maps. Note that an infinite dimensional version of the Sard theorem was proved by Smale [26]. For example the set of critical values of a constant map as well as the set of critical values of the projection $p : S^n \longrightarrow \mathbb{R}^n$, $p(x_1, \ldots, x_{n+1}) = (x_1, \ldots, x_n)$ have both zero measure, yet one of them is a zero dimensional manifold while the other one is the $(n-1)$-dimensional sphere. Consequently the *topological dimension* may play some role in the evaluation process, both for the size of the critical sets and the size of the sets of critical values, the series of works by Church and Timourian from the mid-1960s to mid-1970s being good arguments in this respect. We only mention here three of them, namely [10–12].

3.1 Topological Approach

In this section we also employ the topological dimension to provide some examples of maps with large critical sets.

Theorem 5. *([16]) Every compact connected manifold M^m is a Cantor manifold, that is no subset of M of dimension $\leq m-2$ separates M.*

The next two propositions provide us with an explicit representation of the set of critical values as well as with information on its size, in terms of dimension, for some differentiable maps [9].

Proposition 4. *If M^m, N^n are differential manifolds with M compact, N connected and $m \geq n$, then for all $f \in C^\infty(M, N)$, one has $B(f) = \partial\big(f(R(f))\big) \cup \partial Im\, f \cup A(f)$, where $R(f)$ stands for the set $M \setminus C(f)$ of regular points of f and $A(f)$ for the set $B(f) \cap f(R(f))$.*

Proposition 5. *Let M^m, N^n be smooth manifolds such that $m \geq n \geq 2$. If N is additionally connected and $f : M \to N$ is a closed non-surjective C^1 mapping such that $R(f) \neq \emptyset$, then* $\dim[B(f)] = n-1$.

The next results of this subsection are based on the paper [19].

Theorem 6. *Let M^n, N^n, $(n \geq 2)$ be smooth manifolds such that M is compact. If M, N are orientable and $f : M \to N$ has zero degree, then either $C(f) = M$ or the set $R(f) = M \backslash C(f)$ is not connected. In any case,* $\dim[C(f)] \geq n - 1$.

Proof. We first recall that $\operatorname{sign}(df)_x$, $x \in R(f)$, is defined to be $+1$ or -1 as $(df)_x$ preserves or reverses the orientation and observe that the function $R(f) \longrightarrow \mathbb{Z}$ is locally constant, i.e. it is actually constant on each component of $R(f)$. Recall also that the *degree* $\deg(f)$ of f is defined to be

$$\sum_{x \in f^{-1}(y)} \operatorname{sign}(df)_x, \tag{3}$$

where $y \in \operatorname{Im}(f)$ is a regular value of f, as the sum (3) is independent of $y \in N \setminus B(f)$ [17, p. 28]. On the other hand the equalities

$$0 = \deg(f) = \sum_{x \in f^{-1}(y)} \operatorname{sign}(df)_x,$$

show that the sign map $\operatorname{sgn}(df)_{(\cdot)}$ takes both values ±1. Consequently, $R(f) = M \backslash C(f)$ is not connected, which shows, by using Theorem 5, that $\dim[C(f)] \geq n - 1$. □

Corollary 3. *If M^n, N^n, $(n \geq 2)$ are smooth manifolds with M is compact and N orientable, then* $\dim[C(f)] \geq n - 1$ *for all $f \in C^1(M, N)$ in each of the following situations:*

1. *N is not compact.*
2. *N is compact and M nonorientable.*

Theorem 7. *If M^m, N^n are compact connected smooth manifolds and $f : M \longrightarrow N$ is a C^1-differentiable map such that $[\pi_1(N) : \operatorname{Im}(f_*)]$ is infinite, then the following statements hold:*

1. $\dim[B(f)] = n - 1$, *whenever $m \geq n$ and $R(f) \neq \emptyset$.*
2. $\deg(f) = 0$ *whenever $m = n$ and M, N are orientable.*

We are next interested in pairs (M^n, N^n) of connected orientable manifolds with the property that $\deg(f) = 0$ for all $f \in C^1(M, N)$. As we have already seen in Theorem 7(2), this is the case when $\varphi_{alg}(\pi_1(M), \pi_1(N)) = \infty$. Note that $\varphi_{alg}(G, H)$ stands for the *algebraic φ-category* of the pair (G, H) of groups, i.e.

$$\varphi_{alg}(G, H) := \min\{[H : \operatorname{Im}(f)] \mid f \in Hom(G, H)\}.$$

If $[H : \operatorname{Im}(f)]$ is infinite for all $f \in Hom(G, H)$, then the notation $\varphi_{alg}(G, H) = \infty$ is used. Recall that if G, H are finitely generated abelian groups such that the inequality $\operatorname{rank}(G/t(G)) < \operatorname{rank}(H/t(H))$ holds, then $\varphi_{alg}\big(G, H\big) = \infty$ [18].

Theorem 8. *If M^n, N^n $(n \geq 2)$ are compact connected manifolds, then* $\dim\left[C(f)\right] \geq n-1$*, for all $f \in C^1(M,N)$, in each of the following situations:*

1. *$\varphi_{alg}(\pi_1(M), \pi_1(N)) = \infty$ and N is orientable.*
2. *$\pi_1(M)$ is finite and $\pi_1(N)$ is infinite.*

Proposition 6. *For every groups G, H, the following inequalities hold:*

1. *$\varphi_{alg}(G,H) \geq \varphi_{alg}\left(\frac{G}{[G,G]}, \frac{H}{[H,H]}\right)$.*
2. *$\varphi_{alg}(G,H) \geq \varphi_{alg}\left(\frac{G}{t(G)}, \frac{H}{t(H)}\right)$.*

Corollary 4. *Let X, Y be pathwise connected spaces and $\beta_1(X), \beta_1(Y)$ their first Betti numbers.Then :*

1. *The inequality $\varphi_{alg}\big(\pi_1(X), \pi_1(Y)\big) \geq \varphi_{alg}\big(H_1(X), H_1(Y)\big)$ holds.*
2. *If X, Y are compact ENR spaces such that $\beta_1(X) < \beta_1(Y)$, then we have $\varphi_{alg}\big(\pi_1(X), \pi_1(Y)\big) = \infty$.*

If M is a differentiable manifold, we denote by $\#_r M$ the connected sum $M\#M\#\cdots\#M$ of r copies of M. The connected sum $\#_g T^2$ of g copies of the torus T^2 is also denoted by T_g and the connected sum $\#_g\mathbb{RP}^2$ of g copies of the projective plane $\mathbb{RP}^2$ by P_g.

Example 1. If M^m, N^n $(m,n \geq 2)$ are closed differential manifolds, then in the following cases one has $\varphi_{alg}\big(\pi_1(\#_r M), \pi_1(\#_s N)\big) = \infty$:

1. M, N orientable and $r\beta_1(M) < s\beta_1(N)$.
2. M, N nonorientable and $r(\beta_1(M)+1) < s(\beta_1(N)+1)$.
3. M nonorientable, N orientable and $r(\beta_1(M)+1) < s\beta_1(N)+1$.
4. M orientable, N nonorientable and $r\beta_1(M)+1 < s(\beta_1(N)+1)$.

Indeed, $\beta_1(\#_k X) = k\beta_1(X)$ if X is orientable and $\beta_1(\#_k X) = k-1+k\beta_1(X)$ if X is not orientable [15, p. 258], as the connected sum of nonorientable manifolds is nonorientable [27, p. 92]. In particular one gets:

1. If $g < g'$, then $\varphi_{alg}\big(\pi_1(T_g), \pi_1(T_{g'})\big) = \infty$.
2. If $g < g'$, then $\varphi_{alg}\big(\pi_1(P_g), \pi_1(P_{g'})\big) = \infty$.
3. If $g < 2g'+1$, then $\varphi_{alg}\big(\pi_1(P_g), \pi_1(T_{g'})\big) = \infty$.
4. If $2g < g'-1$, then $\varphi_{alg}\big(\pi_1(T_g), \pi_1(P_{g'})\big) = \infty$.

Example 2. Let us consider two closed differential manifolds M^m, N^n $(m,n \geq 2)$ and $f : \#_r M \longrightarrow \#_s N$ be a C^1 map.

1. If $m = n$, then $\dim(C(f)) \geq n-1$ in each of the following cases:
   a. M, N orientable and $r\beta_1(M) < s\beta_1(N)$.
   b. M nonorientable, N orientable and $r(\beta_1(M)+1) < s\beta_1(N)+1$.

2. If $m \geq n$ and $R(f) \neq \emptyset$, then $\dim(B(f)) = n - 1$, in each of the following cases:

   a. M, N orientable and $r\beta_1(M) < s\beta_1(N)$.
   b. M, N nonorientable and $r(\beta_1(M) + 1) < s(\beta_1(N) + 1)$.
   c. M nonorientable, N orientable and $r(\beta_1(M) + 1) < s\beta_1(N) + 1$.
   d. M orientable, N nonorientable and $r\beta_1(M) + 1 < s(\beta_1(N) + 1)$.

In particular one gets:

1. $\dim[C(f)] \geq 1$ for every C^1 map $f : M \to N$ and $\dim[B(f)] = 1$ whenever $R(f) \neq \emptyset$ in each of the following situations:

   a. $M = T_g$, $N = T_{g'}$ and $g < g'$.
   b. $M = P_g$, $N = T_{g'}$ and $g < 2g' + 1$.

2. $\dim[B(f)] = 1$ for every C^1 map $f : M \to N$ with $R(f) \neq \emptyset$ in each the following situations:

   a. $M = P^2$, $N = P_g$ and $g \geq 2$.
   b. $M = P_g$, $N = P_{g'}$ and $g < g'$.
   c. $M = T_g$, $N = P_{g'}$ and $2g < g' - 1$.

3. If $n > k \geq 1$ and r is an arbitrary positive integer, then $\dim[C(f)] \geq n - 1$, for every C^1 map $f : \#_r(T^k \times S^{n-k}) \to \#_r T^n$ and $\dim[B(f)] = n - 1$ whenever $R(f) \neq \emptyset$.
4. If r, s are arbitrary positive integers, then the inequality $\dim[C(f)] \geq 2$ holds, for every C^1 map $f : \#_r\mathbb{R}P^3 \to \#_s(S^1 \times \mathbb{R}P^2)$ and $\dim[B(f)] = 2$ whenever $R(f) \neq \emptyset$.
5. If r, s are arbitrary positive integers, then the inequality $\dim[C(f)] \geq 3$ holds, for every C^1 map $f : \#_r\mathbb{C}P^2 \to \#_s(T^2 \times \mathbb{R}P^2)$ and $\dim[B(f)] = 3$ whenever $R(f) \neq \emptyset$.

Corollary 5. *If M^m, N^n are compact connected manifolds such that $m \geq n \geq 2$ and $\varphi_{alg}(\pi_1(M), \pi_1(N)) = \infty$, then no submanifold of M of dimension less than or equal to $n - 2$ is the critical set of any differentiable mapping $f : M \longrightarrow N$.*

The next result shows that the critical sets of some maps, possibly with nonzero degree, are still large.

Theorem 9. *If $M^n, N^n, n \geq 3$ are compact orientable smooth manifolds and $f : M \longrightarrow N$ is a smooth map such that $|\deg(f)| < \varphi_{alg}\big(\pi_1(M), \pi_1(N)\big)$, then one gets* $\dim[C(f)] \geq n - 2$.

The next result provides estimates on the algebraic φ-category of a pair of finite groups in terms of their orders and the orders' prime decomposition structures. This might be useful to find examples of manifolds, with finite fundamental groups, satisfying the requirements of Theorem 9. Such an example appears in the paper [19].

Theorem 10. *If* G, H *are finite Abelian groups with* $\gcd\big(o(G), o(H)\big) = p_1^{r_1} \cdot \ldots \cdot p_k^{r_k}$*, then*

$$\frac{o(H)}{\gcd\big(o(G), o(H)\big)} \leq \varphi_{alg}(G, H) \leq \frac{o(H)}{p_1^{\gamma_1} \cdots p_k^{\gamma_k}},$$

where $\gamma_i = \min(\alpha_1, \beta_1) + \cdots + \min(\alpha_z, \beta_z), z = \min(x, y)$*, and* $S_i = \mathbb{Z}_{p_i^{\alpha_1}} \times \mathbb{Z}_{p_i^{\alpha_2}} \times \cdots \times \mathbb{Z}_{p_i^{\alpha_x}}$, $\Sigma_i = \mathbb{Z}_{p_i^{\beta_1}} \times \mathbb{Z}_{p_i^{\beta_2}} \times \cdots \times \mathbb{Z}_{p_i^{\beta_y}}$ *are the* p_i*-Sylow subgroups of* G *and* H *respectively, and the exponents* $\alpha_1, \alpha_2, \ldots, \alpha_x, \beta_1, \beta_2, \ldots, \beta_y$ *satisfy* $\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_x, \beta_1 \geq \beta_2 \geq \cdots \geq \beta_y$.

3.2 Geometrical Approach

Another approach to provide different examples of maps with high dimensional critical sets is provided by Pintea [20] and uses top volume forms of the target oriented manifolds as the key tools.

Let M be an m-dimensional manifold and let η be a k-form on M. We define the *vanishing set of* η, by the collection of points $z \in M$ at which η is (identically) zero, that is the set

$$V(\eta) := \{z \in M : \eta_z(v_1, \ldots, v_k) = 0 \text{ for all } v_i \in T_z(M)\}.$$

Note that if $\eta = \sum\limits_{1 \leq j_1 < \cdots < j_k \leq m} a_{j_1 \ldots j_k} dx_{j_1} \wedge \cdots \wedge dx_{j_k}$ is the local representation of η with respect to some local chart (U, ψ), then

$$V(\eta) \cap U = \{z \in M : a_{j_1 \ldots j_k}(z) = 0, \text{ for all } 1 \leq j_1 < \cdots < j_k \leq m\}.$$

For more details on vanishing sets and their properties we refer to [6].

Theorem 11. *Let* M^m, N^n *be differential manifolds such* $m \geq n$ *and* N *is orientable. If* ω_N *is a volume form on* N *and* $f : M \to N$ *is a differentiable mapping, then* $V(f^*\omega_N) = C(f)$.

Remark 2. 1. If N is a compact connected orientable n-dimensional manifold, ω_N is a volume form on N and $\varphi : N \to \mathbb{R}$ is a differentiable function, then $\int_N \varphi\omega_N = 0$ implies that either $\varphi \equiv 0$, or $N \backslash \varphi^{-1}(0)$ is not connected, namely $\varphi^{-1}(0) = V(\varphi\omega_N)$ separates N. Consequently, the equality $\int_N \theta = 0$, for some differential form $\theta \in \Omega^n(N)$, implies that either $\theta = 0$, or $V(\theta)$ separates N. In any case $\dim(V(\theta)) \geq n - 1$.

2. By using the above item (1), one can provide an alternative proof of Theorem 6. Indeed, if ω_N is a volume form on N, observe that $\deg(f) = 0$ if and only if

$\int_M f^*\omega_{_N} = 0$, which shows that either $C(f) = V(f^*\omega_{_N}) = M$ or $C(f) = V(f^*\omega_{_N})$ separates M. In any case, the conclusion $\dim[C(f)] \geq n-1$ follows immediately. Next, if $f^*\omega_{_N} = d\alpha$, for some $\alpha \in \Omega^{n-1}(M)$, then obviously $\deg(f) = 0$.

Corollary 6. *Let M^n, N^n $(n \geq 2)$ be compact connected orientable manifolds, let $\omega_{_N}$ be a volume form on N and let $f : M \to N$ be a differentiable mapping. If $f^*\omega_{_N}$ is exact, then* $\dim[C(f)] \geq n-1$. *In particular, if f is homotopic to a map $g : M \longrightarrow N$ having just critical points, then* $\dim[C(f)] \geq n-1$.

Proposition 7. *Let M^m, N^n, P^{m-n}, $(m > n \geq 2)$ be compact orientable manifolds such that one of the de Rham cohomology groups $H^n_{dR}(M)$ or $H^{m-n}_{dR}(M)$ is trivial. Then, every smooth map $f : M \longrightarrow N \times P$ has zero degree and* $\dim(C(f)) \geq m-1$ *therefore.*

Proof. Let $\omega_{_N}$ be a volume form on N and $\omega_{_P}$ be a volume form on P. Observe that $f = (\pi_N \circ f, \pi_P \circ f)$, where $\pi_N : N \times P \longrightarrow N$ and $\pi_P : N \times P \longrightarrow P$ are the projections. Recall that $N \times P$ is orientable and $\pi_N^*\omega_N \wedge \pi_P^*\omega_P$ is a volume form on $N \times P$, which shows that

$$\deg(f) = \frac{\displaystyle\int_M f^*\left(\pi_N^*\omega_N \wedge \pi_P^*\omega_P\right)}{\displaystyle\int_{N\times P} \pi_N^*\omega_N \wedge \pi_P^*\omega_P}.$$

and, by means of Theorem 11,

$$C(f) = V\left(f^*\left(\pi_N^*\omega_N \wedge \pi_P^*\omega_P\right)\right) = V\left((\pi_N \circ f)^*\omega_N \wedge (\pi_P \circ f)^*\omega_P\right).$$

Assume that $H^n_{dR}(M)$ is trivial, namely $(\pi_N \circ f)^*\omega_N$ is exact, i.e. $(\pi_N \circ f)^*\omega_N = d\alpha$, for some $\alpha \in \Omega^{n-1}(M)$. Taking into account that $(\pi_P \circ f)^*\omega_P$ is closed, one gets successively

$$\begin{aligned}\int_M f^*\left(\pi_N^*\omega_N \wedge \pi_P^*\omega_P\right) &= \int_M (\pi_N \circ f)^*\omega_N \wedge (\pi_P \circ f)^*\omega_P \\ &= \int_M d\alpha \wedge (\pi_P \circ f)^*\omega_P \\ &= \int_M d\left(\alpha \wedge (\pi_P \circ f)^*\omega_P\right) \pm \int_M \alpha \wedge d(\pi_P \circ f)^*\omega_P \\ &= \int_M d\left(\alpha \wedge (\pi_P \circ f)^*\omega_P\right) = 0.\end{aligned}$$

At this point we may choose to use Theorem 6 or Remark 2 in order to prove the statement. The case $H_{dR}^{m-n}(M) = \{0\}$ can be treated similarly. □

Example 3. If $m, n \geq 2$, then every differentiable map $f : S^{m+n} \to S^m \times S^n$ has zero degree, hence $\dim[C(f)] \geq m + n - 1$.

Recall that two submanifolds $N_1,\ N_2$ of a given finite dimensional manifold M are said to *intersect transversally* at $p \in N_1 \cap N_2$, written $N_1 \pitchfork_p N_2$, if $T_p(M) = T_p(N_1) + T_p(N_2)$. If $N_1,\ N_2$ do not intersect transversally at $p \in N_1 \cap N_2$, we use the notation $N_1 \not\pitchfork_p N_2$.

Proposition 8. *Let M^m, N^n, P^p be manifolds such that $m \geq n+p$. If N and P are orientable and ω_N, ω_P are volume forms on N and P respectively, then the inclusion $\mathcal{V}_f\big(g^*\omega_P\big) \subseteq C(g) \cup U(f, g)$ holds, for every differentiable maps $f : M \longrightarrow N$, $g : M \longrightarrow P$, where*

$$U(f, g) := \big\{x \in R(f) \cap R(g) | f^{-1}\big(f(x)\big) \not\pitchfork_x g^{-1}\big(g(x)\big)\big\}.$$

In what follows we are going to provide an approach for the higher codimension case ($\dim M =: m > n := \dim N$), in which, the role of the form $f^*\omega_N$ will be played by forms of type $f^*\omega_N \wedge \theta$, where $\theta \in \Omega^{m-n}(M)$ are closed. If $f : M^m \to N^n, (m > n)$ is a differentiable mapping and $\omega \in \Omega^{m-n}(M)$, consider the set $R(f) := M \backslash C(f)$ of regular points of f and

$$\mathcal{V}_f(\omega) := \big\{p \in R(f) \,|\, \omega_p(u_1, \ldots, u_{m-n}) = 0,\ \text{for all } u_1, \ldots, u_{m-n} \in \ker(df)_p\big\}.$$

Observe that $\mathcal{V}_f(\omega) = \{p \in R(f) : (i^*_{f(p)}\omega)_p = 0\}$, where $i_{f(p)}$ stands for the inclusion $f^{-1}(f(p)) \backslash C(f) \hookrightarrow R(f)$ of the fiber of $f\big|_{R(f)}$ passing through p. In other words

$$\mathcal{V}_f(\omega) = \bigcup_{p \in R(f)} V\big(i^*_{f(p)}\omega\big).$$

Theorem 12. *Let $M^m, N^n, m \geq n$, be compact oriented manifolds and let $f : M \to N$ be a differentiable mapping. If $\theta \in \Omega^{m-n}(M)$, then $V(f^*\omega_N \wedge \theta) = C(f) \cup \mathcal{V}_f(\theta)$.*

Theorem 13. *Let M^m, N^n, $(m > n \geq 2)$ be compact orientable manifolds and let ω_N be a volume form on N. If $f : M \to N$ is a differentiable map such that $f^*\omega_N$ is an exact form, then, for all $\theta \in Z^{m-n}(M)$, the inequality* $\dim C(f) \geq m - \gamma_f - 1$ *holds, where* $\gamma_f := \min\{\dim \mathcal{V}_f(\theta) \,|\, \theta \in Z^{m-n}(M)\}$.

Proof. Since $f^*\omega_N$ is exact, it follows that $f^*\omega_N = d\alpha$ for some $\alpha \in \Omega^{n-1}(M)$. This means that

$$\int_M f^*\omega_N \wedge \theta = \int_M d\alpha \wedge \theta = \int_M d(\alpha \wedge \theta) + (-1)^n \int_M \alpha \wedge d\theta = 0,$$

for each $\theta \in Z^{m-n}(M)$. Therefore $\dim C(f) + \dim \mathcal{V}_f(\theta) \geq \dim V(f^*\omega_N \wedge \theta) \geq m - 1$ for all $\theta \in Z^{m-n}(M)$ for any closed differential form $\theta \in Z^{m-n}(M)$. By considering the minimum with respect to the closed forms $\theta \in Z^{m-n}(M)$, one gets $\dim C(f) \geq m - \gamma_f - 1$. □

Acknowledgements The first author would like to express all his thanks to Professor Themistocles M. Rassias for the strong encouraging to prepare this survey.

The first author is partially supported by the King Saud University D.S.F.P. Program

References

1. D. Andrica, *Critical Point Theory and Some Applications* (Cluj University Press, Cluj-Napoca, 2005)
2. D. Andrica, L. Funar, On smooth maps with finitely many critical points. J. Lond. Math. Soc. **69**(2), 783–800 (2004)
3. D. Andrica, L. Funar, On smooth maps with finitely many critical points: addendum. J. Lond. Math. Soc. **73**(2), 231–236 (2006)
4. D. Andrica, L. Funar, E.A. Kudryavtseva, The minimal number of critical points of maps between surfaces. Russ. J. Math. Phys. **16**(3), 363–370 (2009)
5. P.L. Antonelli, Differentiable Montgomery-Samelson fiberings withfinite singular sets. Can. J. Math. **21**, 1489–1495 (1969)
6. Z.M. Balogh, C. Pintea, H. Rohner, Size of tangencies to non-involutive distributions. Preprint (2010)
7. S.A. Bogatyi, D.L. Gonçalves, E.A. Kudryavtseva, H. Zieschang, Realization of primitive branched coverings over closed surfaces following the Hurwitz approach. Cent. Eur. J. Math. **1**, 184–197 (2003)
8. S.A. Bogatyi, D.L. Gonçalves, E.A. Kudryavtseva, H. Zieschang, Realization of primitive branched coverings over closed surfaces, in *Advances in Topological Quantum Field Theory* ed. by JM Bryden. NATO Science Series II, Mathematics Physics and Chemistry, vol. 179 (Kluwer, Dordrecht, 2004), pp. 297–316
9. G. Cicortaş, C. Pintea, G. Ţopan, Isomorphic homotopy groups of certain regular sets and their images. Topol. Appl. **157**, 635–642 (2010)
10. P.T. Church, Factorization of differentiable maps with branch set dimension at most n-3. Trans. Am. Math. Soc. **115**, 370–387 (1965)
11. P.T. Church, Differentiable monotone maps on manifolds. Trans. Am. Math. Soc. **128**(2), 185–205 (1967)
12. P.T. Church, J.G. Timourian, Differentiable maps with 0-dimensional critical set, I. Pac. J. Math. **41**(3), 615–629 (1972)
13. L. Funar, Global classification of isolated singularities in dimensions $(4, 3)$ and $(8, 5)$, arXiv:0911.3517v2 [math.GT] 16 Sep 2010
14. L. Funar, C. Pintea, P. Zhang, Examples of smooth maps with finitely many critical points in dimensions $(4, 3)$, $(8, 5)$ and $(16, 9)$. Proc. Am. Math. Soc. **138**(1), 355–365 (2010)
15. A. Hatcher, *Algebraic Topology* (Cambridge University Press, Cambridge/New York, 2002)
16. W. Hurewicz, H. Wallman, *Dimension Theory* (Princeton University Press, Princeton, 1948)
17. J.W. Milnor, *Topology from the Differentiable Viewpoint* (University Press of Virginia, Charlottesville, 1965)
18. C. Pintea, Continuous mappings with an infinite number of topologically critical points. Ann. Polonici Math. **67**, 87–93 (1997)
19. C. Pintea, The size of some critical sets by means of dimension and algebraic φ-category. Topol. Method Nonlinear Anal. **35**, 395–407 (2010)

20. C. Pintea, Smooth mappings with higher dimensional critical sets. Can. Math. Bull. **53**, 542–549 (2010)
21. G.M. Rassias, On the non-gegenerate critical points of differentiable functions. Tamkang J. Math. **10**, 67–73 (1979)
22. G.M. Rassias, On the Morse-Smale characteristic of a differentiable manifold. Bull. Aust. Math. Soc. **20**(2), 281–283 (1979)
23. A. Sard, The measure of the critical values of differentiable maps. Bull. Am. Math. Soc. **48**, 883–890 (1942)
24. G.B. Shabat, V.A. Voevodsky, Drawing curves over number fields, in *Grothendieck Festschrift, vol. 3*, ed. by P. Cartier. Progress in Mathematics, vol. 88 (Birkhäuser, Boston/Berlin, 1990), pp. 199–227
25. V.V. Sharko, *Functions on Manifolds: Algebraic and Topological Aspects*. Translations of Mathematical Monographs, vol. 131 (American Mathematical Society, Providence, 1993)
26. S. Smale, An infinite dimensional version of Sard's theorem. Am. J. Math. **87**(4), 861–866 (1965)
27. T. Tao, *Poincaré's Legacies, Part II* (American Mathematical Society, Providence, 2009)

The Fox–Li Operator as a Test and a Spur for Wiener–Hopf Theory

Albrecht Böttcher, Sergei Grudsky, and Arieh Iserles

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract The paper is a concise survey of some rigorous results on the Fox–Li operator. This operator may be interpreted as a large truncation of a Wiener–Hopf operator with an oscillating symbol. Employing theorems from Wiener–Hopf theory one can therefore derive remarkable properties of the Fox–Li operator in a fairly comfortable way, but it turns out that Wiener–Hopf theory is unequal to the task of answering the crucial questions on the Fox–Li operator.

1 Masers, Lasers and the Fox–Li Operator

The story begun 50 years ago. Fox and Li [13] considered the repeated reflection of an electromagnetic wave of wave length λ between two plane-parallel rectangular mirrors. By a tensor product phenomenon, it suffices to suppose that the mirrors are infinite strips of height $2a$ with distance b between them. A distribution $u(x)$, $x \in (-a, a)$, of the field on one mirror goes over into the distribution given by

$$(Au)(x) = \frac{\mathrm{e}^{\mathrm{i}\pi/4}}{2\sqrt{\lambda}} \int_{-a}^{a} \kappa(x-y)u(y)\mathrm{d}y, \quad x \in (-a, a), \tag{1}$$

A. Böttcher
Faculty of Mathematics, TU Chemnitz, 09107 Chemnitz, Germany

S. Grudsky
Department of Mathematics, CINVESTAV , Mexico-City, Mexico

A. Iserles (✉)
Department of Applied Mathematics and Theoretical Physics Centre for Mathematical Sciences
University of Cambridge, Wilberforce Rd, Cambridge CB3 0WA, United Kingdom,
e-mail: A.Iserles@damtp.cam.ac.uk

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_3, © Springer-Verlag Berlin Heidelberg 2012

on the other mirror. Here κ is the function

$$\kappa(t) = \frac{\mathrm{e}^{-\mathrm{i}k\sqrt{t^2+b^2}}}{(t^2+b^2)^{1/4}}\left(1+\frac{b}{\sqrt{t^2+b^2}}\right) \tag{2}$$

where $k = 1/\lambda$ denotes the wave number. What Fox and Li were interested in were the eigenvalues and eigenfunctions of the operator A: if $Au = \mu u$, then the distribution $u(x)$ will after n reflections be transformed into $\mu^n u(x)$. The number $1 - |\mu|^2$ is the energy loss of the mode u at one step. This setup is called a maser in the case of microwaves ($\lambda \approx 1$ cm) and a laser when working with light waves, in the range $\lambda \approx 5 \cdot 10^{-5}$ cm.

Let us consider the integral operator A given by (1) on $L^2(-a, a)$. Being compact, it has at most countably many eigenvalues with the origin as the only possible cluster point. Cochran [11] and Hochstadt [16] provided a rigorous argument which proves that A has at least one eigenvalue. However, there is no theorem that would imply more or anything else of interest about the operator A. Well, A has a difference kernel and hence one would expect that for large a the eigenvalues of A somehow mimic the values of the Fourier transform of κ,

$$\hat{\kappa}(\xi) := \int_{-\infty}^{\infty} \kappa(t)\mathrm{e}^{\mathrm{i}\xi t}\mathrm{d}t, \quad \xi \in \mathbb{R}.$$

The function $\hat{\kappa}(\xi)$ is even, exponentially decaying as $|\xi| \to \infty$, and in $L^1(\mathbb{R})$. Had it been in $C(\mathbb{R})$, we would have had a theorem implying that the eigenvalues of A cluster along the range $\hat{\kappa}(\mathbb{R})$ as $a \to \infty$. However, $\hat{\kappa}(\xi)$ behaves like

$$\sqrt{\frac{\pi}{2b|\xi - k|}}\,[1 + \mathrm{i}\,\mathrm{sign}(\xi - k)]$$

as $\xi \to k$ and hence it is not even in $L^\infty(\mathbb{R})$. In addition we should mention that the case $a \gg b$ is not the really interesting case in physics. One is therefore left with tackling the eigenvalue problem for A numerically, the big problem in this connection being that the kernel κ is highly oscillating: note that $k \approx 20{,}000\,\mathrm{cm}^{-1}$ for light waves.

Fox and Li found an ingenious way out. The physically relevant case is the one where $a \ll b$. They wrote

$$\exp(-\mathrm{i}k\sqrt{t^2+b^2}) = \exp\left(-\mathrm{i}kb\left(1+\frac{t^2}{2b^2}+O\left(\frac{t^4}{b^4}\right)\right)\right),$$

and since $|t| < a$, one may ignore the O term if $kba^4/b^4 \ll 1$, that is, if $a^4 \ll \lambda b^3$. As $\lambda \ll b$, this assumption automatically implies that $a \ll b$, and therefore $(t^2 + b^2)^{1/4}$ and $b/\sqrt{t^2+b^2}$ may be replaced by $\sqrt{b}$ and 1, respectively. In summary, the operator A may be approximated by the operator

$$(A_1u)(x) = \frac{e^{i\pi/4}e^{-ikb}}{\sqrt{\lambda b}} \int_{-a}^{a} e^{-i(k/2b)(x-y)^2} u(y)\,dy, \quad x \in (-a, a).$$

The change of variables $x \to ax$, $y \to ay$ yields the operator

$$(A_2u)(x) = \frac{a e^{i\pi/4}e^{-ikb}}{\sqrt{\lambda b}} \int_{-1}^{1} e^{-i(ka^2/2b)(x-y)^2} u(y)\,dy, \quad x \in (-1, 1), \qquad (3)$$

and abbreviating $\omega := ka^2/(2b) = a^2/(2\lambda b)$ and $\sqrt{i} := e^{i\pi/4}$ we arrive at the equality $A_2 = \sqrt{2\pi}e^{-ikb}\mathcal{F}_\omega^*$ with $\mathcal{F}_\omega^*$ and $\mathcal{F}_\omega$ defined on $L^2(-1,1)$ by

$$(\mathcal{F}_\omega^* u)(x) = \sqrt{\frac{\omega i}{\pi}} \int_{-1}^{1} e^{-i\omega(x-y)^2} u(y)\,dy, \quad (\mathcal{F}_\omega u)(x) = \sqrt{\frac{\omega}{\pi i}} \int_{-1}^{1} e^{i\omega(x-y)^2} u(y)\,dy.$$

Note that $\mathcal{F}_\omega^*$ is really the adjoint of $\mathcal{F}_\omega$. The operator $\mathcal{F}_\omega$ is now called the Fox–Li operator, and the eigenvalues and eigenfunctions of this operator are what one wants to know.

After the change of variables $x \to x/\sqrt{\omega} - 1$, $y \to y/\sqrt{\omega} - 1$ the operator $\mathcal{F}_\omega$ becomes the operator given by

$$(F_\omega u)(x) = \frac{1}{\sqrt{\pi i}} \int_0^{2\sqrt{\omega}} e^{i(x-y)^2} u(y)\,dy, \quad x \in (0, 2\sqrt{\omega}), \qquad (4)$$

on $L^2(0, 2\sqrt{\omega})$, and since $\omega = a^2/(2\lambda b)$ may also be assumed to be very large, F_ω is a very large truncation of a Wiener–Hopf operator.

In summary, the Fox–Li operator is a reasonable approximation to the original physical problem and at the same time a large truncated Wiener–Hopf operator whenever $\lambda^2 b^2 \ll a^4 \ll \lambda b^3$. Using the dimensionless parameters $\hat{a} := ka$ and $\hat{b} := kb$, these inequalities read $\hat{b}^{1/2} \ll \hat{a} \ll \hat{b}^{3/4}$, and ω becomes $\hat{a}^2/(2\hat{b})$. Fox and Li themselves showed that already the moderate choice $\hat{a} = 25$, $\hat{b} = 100$ leads to acceptable numerical results.

2 Wiener–Hopf Operators

An integral operator on $L^2(0,\infty)$ of the form

$$(Wu)(x) = \int_0^\infty \varrho(x-y)u(y)\,dy, \quad x \in (0,\infty),$$

is called a *Wiener–Hopf operator*. Such an operator is bounded on $L^2(0,\infty)$ if and only if the Fourier transform $a := \hat{\varrho}$, taken in the distributional sense, is a function in $L^\infty(\mathbb{R})$. The function ϱ is uniquely determined by its Fourier transform

a, henceforth we denote the operator W by $W(a)$. The function a is usually referred to as the *symbol* of $W(a)$. Note that $W(a)$ is the compression to $L^2(0,\infty)$ of the operator which acts on $L^2(\mathbb{R})$ by the following rule: take the Fourier transform, multiply the result by a, and then take the inverse Fourier transform.

For $\tau \in (0,\infty)$, the truncated Wiener–Hopf operator $W_\tau(a)$ is defined on $L^2(0,\tau)$ by

$$(W_\tau u)(x) = \int_0^\tau \varrho(x-y)u(y)\mathrm{d}y, \quad x \in (0,\tau). \tag{5}$$

The Fourier transform of $\varrho(t) = \mathrm{e}^{\mathrm{i}t^2}$ is $\hat{\varrho}(\xi) = \sqrt{\pi\mathrm{i}}\,\mathrm{e}^{-\mathrm{i}\xi^2/4}$. Thus, letting $\sigma(\xi) = \mathrm{e}^{-\mathrm{i}\xi^2/4}$, we see that the Fox–Li operator F_ω given by (4) is nothing but $W_{2\sqrt{\omega}}(\sigma)$, and the problem is to find the eigenvalues and eigenfunctions of $W_\tau(\sigma)$ as $\tau = 2\sqrt{\omega} \to \infty$.

The spectral theory of Wiener–Hopf operators is well developed, one could say that Wiener–Hopf operators and their discrete analogues, Toeplitz operators, are the best understood nontrivial classes of non-selfadjoint operators. We refer to [4] for a presentation of the matter. However, as already said, no result of this theory is immediately applicable to provide any deeper insight into the spectrum $\operatorname{sp} W_\tau(\sigma)$ of $W_\tau(\sigma)$. The best that is available to date is the following result.

Theorem 1. *We have* $\operatorname{sp} W(\sigma) = \overline{\mathbb{D}}$ *and* $\operatorname{sp} W_\tau(\sigma) \subset \overline{\mathbb{D}}$ *for every* $\tau > 0$, *where* $\overline{\mathbb{D}}$ *is the closed unit disc in the complex plane.*

This was established in [7]. The nontrivial part of the theorem is that $\operatorname{sp} W(\sigma)$ is all of $\overline{\mathbb{D}}$. In [7] it is actually shown that $\operatorname{sp} W_\tau(\sigma)$ is contained in the open unit disc $\mathbb{D}$ and that each point $\lambda \in \overline{\mathbb{D}}$ belongs to the essential spectrum of $W(\sigma)$, which means that $W(\sigma) - \lambda I$ is not even invertible modulo compact operators.

3 Eigenvalues

The physicists's intuition, like in Vainshtein's paper [23], and numerical computations, made by Cochran and Hinds [12] for probably the first time, indicate that the eigenvalues of $W_\tau(\sigma)$ lie along a spiral commencing at 1 and rotating clockwise to the origin: cf. Fig. 1. To date, no person alive has been able to prove this, even less so to derive rigourously the shape of the spiral. The following result gives an idea of what one is already proud of.

Theorem 2. *The operator* $W_\tau(\sigma)$ *is a trace class operator with at least one eigenvalue for every* $\tau > 0$, *and with the possible exception of at most countably many* $\tau \in (0,\infty)$, *the operator* $W_\tau(\sigma)$ *has a countable number of eigenvalues.*

This was proved in [11, 16, 19]. The approach of [11, 16] is based on proving that $\det(I - zW_\tau(\sigma))$ is a nonconstant entire function of z. This function has infinitely many discrete zeros of finite multiplicity unless it reduces to a polynomial, which

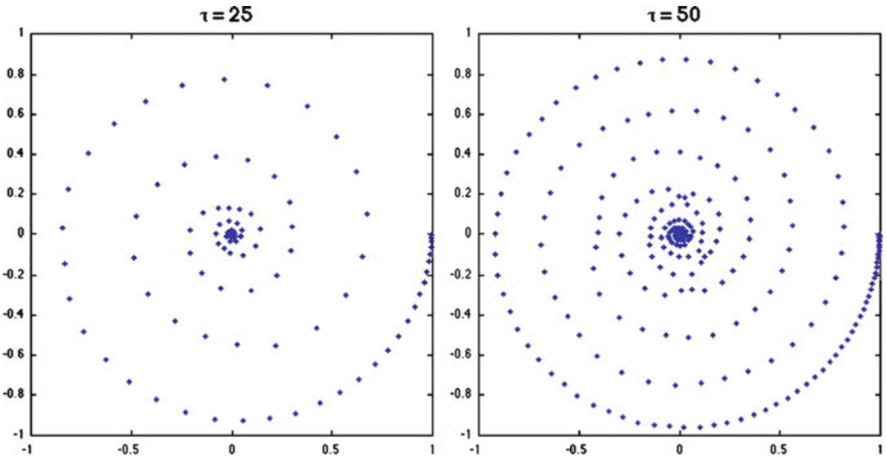


Fig. 1 The eigenvalues of $W_\tau(\sigma)$, with $\tau = 2\sqrt{\omega} = 25, 50$ and σ given by (4)

is shown to happen for at most countably many values of τ. Combining Theorem 1 with the observation that $W_\tau(a)$ is of trace class, one can say even a little more. Namely, let $\{\mu_n(W_\tau(\sigma))\}_{n=1}^N$ denote the eigenvalues of $W_\tau(\sigma)$ counted with their algebraic multiplicities. Then

$$\sum_{n=1}^{N} \mu_n(W_\tau(\sigma)) = \operatorname{tr} W_\tau(\sigma) = \frac{1}{\sqrt{\pi \mathrm{i}}} \int_0^\tau \mathrm{e}^{\mathrm{i}\cdot 0^2} \mathrm{d}x = \frac{\tau}{\sqrt{\pi \mathrm{i}}},$$

and since $|\mu_n(W_\tau(\sigma))| \le 1$ for all n, it follows that

$$\frac{\tau}{\sqrt{\pi}} = \left| \sum_{n=1}^{N} \mu_n(W_\tau(\sigma)) \right| \le \sum_{n=1}^{N} |\mu_n(W_\tau(\sigma))| \le N,$$

which reveals that $W_\tau(\sigma)$ has at least $\tau/\sqrt{\pi}$ eigenvalues.

Vainshtein [23] even raised a conjecture on the shape of the spiral.[1] It says that its parametric representation is $\mu = \exp(-\alpha(\tau)x^\nu - \mathrm{i}\beta(\tau)x^\nu)$, $x \in (0, \infty)$, with

$$\nu = 2, \quad \alpha(\tau) \approx \frac{\zeta(1/2)\pi^{3/2}}{8\sqrt{2}\,\tau^3}, \quad \beta(\tau) \approx \frac{\pi^2}{4\tau^2}, \tag{6}$$

where $\zeta(1/2)$ is Riemann's zeta function at the point $1/2$, and that $x = n$ gives approximately μ_n. We will return to this conjecture below.

[1] According to [12], this conjecture comes from "using a distinctly physical approach based on wave-guide theory", but we admit that we have not been able to follow the argument of [23]. Moreover, numerical computations do not support the conjecture.

Theorems of the type of Szegő's limit theorem [14] give asymptotic expansions for the trace $\operatorname{tr}\varphi(W_\tau(a)) = \sum_n \varphi(\mu_n(W_\tau(a)))$, where $\varphi : \mathbb{C} \to \mathbb{C}$ belongs to a certain class of so-called test functions. The following first-order result for the case $\varphi(z) = z^j$ was proved in [7].

Theorem 3. *For each fixed natural number j,*

$$\operatorname{tr} W_\tau^j(\sigma) = \frac{\tau}{\sqrt{\pi \mathrm{i} j}} + o(\tau) \quad \textit{as} \quad \tau \to \infty.$$

The operator $W_\tau^j(\sigma)$ is the integral operator on $L^2(0,\tau)$ with the kernel

$$m_j(x,y) := \frac{1}{(\pi \mathrm{i})^{j/2}} \int_0^\tau \dots \int_0^\tau \exp\left(\mathrm{i} \sum_{n=1}^{j} (x_n - x_{n+1})^2 \right) \mathrm{d}x_2 \dots \mathrm{d}x_j,$$

where $x_1 = x$ and $x_{j+1} = y$. The trace of $W_\tau^j(\sigma)$ is $\int_0^\tau m_j(x,x)\mathrm{d}x$, and in [7] we proved that the leading term of the asymptotics of this multivariate oscillatory integral is $\tau/\sqrt{\pi \mathrm{i} j}$. We have not been able to determine the second term of the asymptotic expansion for general j.

Results like Theorem 3 can be used to test conjectures on the asymptotic eigenvalue distribution. Suppose we are given a family $\{b_\tau\}_{\tau>0}$ of functions $b_\tau{:}(0,\infty) \to \mathbb{C}$ and we want to know whether it might be true that the eigenvalues of $W_\tau(\sigma)$ are asymptotically distributed like samples of $b_\tau(x)$ at $x = n$. We have

$$\operatorname{tr} W_\tau^j(\sigma) = \sum_n \mu_n^j(W_\tau(\sigma)), \qquad \int_0^\infty b_\tau^j(x)\mathrm{d}x \approx \sum_n b_\tau^j(n),$$

and this is the motivation for saying that the eigenvalues of $W_\tau(\sigma)$ are asymptotically distributed as the values of b_τ (in a very weak sense) if, for each natural number $j \geq 1$,

$$\operatorname{tr} W_\tau^j(\sigma) = \int_0^\infty b_\tau^j(x)\mathrm{d}x + o(\tau) \quad \text{as} \quad \tau \to \infty.$$

Using Theorem 3 we showed the following theorem in [7], which justifies at least a few pieces of Vainshtein's conjecture.

Theorem 4. *Let $b_\tau(x) = \exp(-\alpha(\tau)x^\nu - \mathrm{i}\beta(\tau)x^\nu)$ with positive real numbers $\alpha(\tau)$, $\beta(\tau)$, ν. Then the eigenvalues of $W_\tau(\sigma)$ are asymptotically distributed as the values of b_τ if and only if*

$$\nu = 2, \quad \alpha(\tau) = o\left(\frac{1}{\tau^2}\right), \quad \beta(\tau) = \frac{\pi^2}{4\tau^2} + o\left(\frac{1}{\tau^2}\right).$$

4 Singular Values

The singular values of $W_\tau(\sigma)$ are the positive square roots of the eigenvalues of $W_\tau(\sigma)W_\tau^*(\sigma)$. Since $W_\tau^*(\sigma) = W_\tau(\overline{\sigma})$, we have

$$(W_\tau(\sigma)W_\tau^*(\sigma)u)(x) = \frac{1}{\pi}\int_0^\tau \left(\int_0^\tau \mathrm{e}^{\mathrm{i}(x-t)^2}\mathrm{e}^{-\mathrm{i}(t-y)^2}\mathrm{d}t\right)u(y)\mathrm{d}y, \quad x \in (0,\tau),$$

and hence $W_\tau(\sigma)W_\tau^*(\sigma) = V^*C_1V$ where V is the unitary operator given by $(Vu)(x) = \mathrm{e}^{\mathrm{i}x(\tau-x)}u(x)$ and C_1 is defined by

$$(C_1u)(x) = \frac{1}{\pi}\int_0^\tau \frac{\sin(\tau(x-y))}{x-y}u(y)\mathrm{d}y, \quad x \in (0,\tau).$$

The change of variables $x \to x/\tau$, $y \to y/\tau$ shows that C_1 may be replaced by

$$(C_2u)(x) = \frac{1}{\pi}\int_0^{\tau^2} \frac{\sin(x-y)}{x-y}u(y)\mathrm{d}y, \quad x \in (0,\tau^2).$$

The Fourier transform of $\sin t/(\pi t)$ is $\chi_{(-1,1)}$, the characteristic function of the interval $(-1,1)$. Consequently, the singular values of $W_\tau(\sigma)$ are the square roots of the eigenvalues of the operator $C_2 = W_{\tau^2}(\chi_{(-1,1)})$. This observation was probably first made in [6].

We are thus led to Wiener–Hopf with real-valued symbols. So, let us suppose that $a \in L^\infty(\mathbb{R})$ is real-valued. Then the operators $W(a)$ and $W_\tau(a)$ are selfadjoint. Hartman and Wintner [15] showed that $\mathrm{sp}\, W(a)$ equals the convex hull of the essential range of a. In [5] it was proved that $\mathrm{sp}\, W_\tau(a) \subset \mathrm{sp}\, W(a)$ for all $\tau > 0$ and that $\mathrm{sp}\, W_\tau(a)$ converges to $\mathrm{sp}\, W(a)$ in the Hausdorff metric. Using these general results and taking into account that $\|W_\tau(a)\| < \|a\|_\infty$ unless a is a constant, we arrive at the following.

Theorem 5. *The set of the singular values of $W_\tau(\sigma)$ is contained in $[0,1)$ for every $\tau > 0$ and converges to the segment $[0,1]$ in the Hausdorff metric as $\tau \to \infty$.*

Szegő's limit theorem gives the first term of the asymptotics of the trace of $\varphi(W_\tau(a))$ for arbitrary real-valued $a \in L^\infty(\mathbb{R})$ and the first two terms of the asymptotics if, in addition, a is smooth enough; see [4, 14]. Hence, for $a = \chi_{(-1,1)}$ we cannot derive a second order result in this way. Fortunately, the case where $a = \gamma\chi_{(\alpha,\beta)}$ was studied in detail by Landau and Widom [18].[2] They proved that if

[2] The reader might enjoy knowing the following, which is cited from [1]: "Harold Widom grew up in Brooklyn, New York. He went to Stuyvesant High School where he was captain of the math team. Coincidentally, the captain of the rival team at the Bronx High School of Science was Henry Landau ...".

$\alpha < \beta$ and $\gamma > 0$ are real numbers, then

$$\operatorname{tr} \varphi(W_\tau(\gamma \chi_{(\alpha,\beta)})) = \tau \, \frac{\varphi(\gamma)(\beta - \alpha)}{2\pi} + \frac{\log \tau}{\pi^2} \int_0^\gamma \frac{\gamma \varphi(x) - x\varphi(\gamma)}{x(\gamma - x)} \, \mathrm{d}x + O(1)$$

for every $\varphi \in C^\infty(\mathbb{R})$ satisfying $\varphi(0) = 0$. This was conjectured by Slepian [20]. A second proof of this result is in [24]. In [6] we applied this formula to $W_{\tau^2}(\chi_{(-1,1)})$ in order to get the following result on the finer distribution of the singular values of $W_\tau(\sigma)$.

Theorem 6. *Denote by $N(x, y)$ the number of singular values of $W_\tau(\sigma)$, counted with their multiplicities, which lie in the interval $(\sqrt{x}, \sqrt{y})$. Then for each δ in $(0, 1/2)$,*

$$N(1-\delta, 1) = \frac{\tau^2}{\pi} - \frac{2 \log \tau}{\pi^2} \log \frac{1-\delta}{\delta} + o(\log \tau),$$

$$N(\delta, 1-\delta) = \frac{4 \log \tau}{\pi^2} \log \frac{1-\delta}{\delta} + o(\log \tau),$$

$$N(0, \delta) = \infty.$$

Thus, although, by Theorem 5, the singular values fill $[0, 1]$ densely as τ goes to ∞, the overwhelming majority of them are concentrated extremely close to the endpoints of the segment.

5 Complex Wave Numbers

Let us assume that the wave number k lies in the lower complex half-plane, $k = k_0 - \mathrm{i}\varepsilon$ with $k_0 = 1/\lambda$ and $\varepsilon > 0$. This assumption may not be of great interest in maser and laser theory, but it might be satisfied in problems of acoustics and, more importantly, it makes the problem nicely accessible to Wiener–Hopf theory.

Replacing k by $k_0 - \mathrm{i}\varepsilon$ in (2) and proceeding as in Sect. 1, the operator (3) now becomes

$$(A_{2,\varepsilon} u)(x) = \frac{a \mathrm{e}^{\mathrm{i}\pi/4} \mathrm{e}^{-\mathrm{i}kb}}{\sqrt{\lambda b}} \int_{-1}^{1} \mathrm{e}^{-\mathrm{i}(k_0 a^2/2b)(x-y)^2} \mathrm{e}^{-(\varepsilon a^2/2b)(x-y)^2} u(y) \mathrm{d}y, \qquad (7)$$

and letting $\omega = k_0 a^2/(2b)$ and $\tau = 2\sqrt{\omega}$, we get the operator

$$(F_{\omega,\varepsilon} u)(x) = \frac{1}{\sqrt{\pi \mathrm{i}}} \int_0^\tau \mathrm{e}^{\mathrm{i}(x-y)^2} \mathrm{e}^{-(\varepsilon/k_0)(x-y)^2} u(y) \mathrm{d}y, \quad x \in (0, \tau) \qquad (8)$$

in place of the operator (4). Here τ is a large number. The spectrum of (7) is what we are looking for, and this spectrum is $\sqrt{2\pi}\,\mathrm{e}^{-\mathrm{i}k_0 b}\mathrm{e}^{-\varepsilon b}$ times the complex conjugates of the points in the spectrum of $F_{\omega,\varepsilon}$. The Fourier transform of $(1/\sqrt{\pi \mathrm{i}})\mathrm{e}^{\mathrm{i}t^2}\mathrm{e}^{-(\varepsilon/k_0)t^2}$ is

$$\sigma_{\varepsilon/k_0}(\xi) = \frac{1}{\sqrt{1+\mathrm{i}\varepsilon/k_0}} \exp\left(-\frac{(\varepsilon/k_0)\xi^2}{4(1+\varepsilon^2/k_0^2)}\right) \exp\left(-\mathrm{i}\frac{\xi^2}{4(1+\varepsilon^2/k_0^2)}\right)$$

and hence we may write $F_{\omega,\varepsilon} = W_\tau(\sigma_{\varepsilon/k_0})$. Obviously, for $\varepsilon = 0$, the symbol σ_{ε/k_0} coincides with σ. The function σ is in $L^\infty(\mathbb{R})$ but not in $L^1(\mathbb{R})$, neither it is continuous on the one-point compactification $\dot{\mathbb{R}}$ of $\mathbb{R}$, which causes a great deal of problems in employing Wiener–Hopf theory. In contrast to this, σ_{ε/k_0} is in $L^1(\mathbb{R}) \cap C(\dot{\mathbb{R}})$, which facilitates matters significantly.

The kernels of the operators (4) and (8) are complex-symmetric, which implies that the symbol, i.e. the Fourier transform of the kernel function, is even. Note that if a is even, $a(\xi) = a(-\xi)$ for $\xi \in \mathbb{R}$, then we may think of the essential range $\mathcal{R}(a)$ of a as a curve which is traced out by $a(\xi)$ from $a(\infty)$ to $a(0)$ as ξ moves from $-\infty$ to 0 and then backwards from $a(0)$ to $a(\infty)$ as ξ moves further from 0 to $+\infty$. Complex-symmetric Toeplitz matrices and Wiener–Hopf operators with complex-symmetric kernels have certain peculiarities. The following was established in [6] and is the continuous analogue of results by Tilli [21] and Widom [25]. Namely, let $a \in L^1(\mathbb{R}) \cap C(\dot{\mathbb{R}})$, suppose a is even, and assume also that the essential range $\mathcal{R}(a)$ of a does not contain interior points. The last assumption is always satisfied if a has some minimal smoothness. Then the spectrum of $W_\tau(a)$ converges to $\mathcal{R}(a)$ in the Hausdorff metric. Secondly, if $\varphi : \mathbb{C} \to \mathbb{C}$ is any continuous function such that $\varphi(z)/z$ converges to a finite limit as $z \to 0$, then

$$\sum_n \varphi(\mu_n(W_\tau(a))) = \frac{\tau}{2\pi}\int_{-\infty}^{\infty} \varphi(a(\xi))\mathrm{d}\xi + o(\tau).$$

Applying these two general results to $a = \sigma_{\varepsilon/k_0}$, we obtain the following two theorems from [6].

Theorem 7. *As $\tau \to \infty$, the spectrum of $W_\tau(\sigma_{\varepsilon/k_0})$ converges in the Hausdorff metric to the logarithmic spiral*

$$\mathcal{R}(\sigma_{\varepsilon/k_0}) = \left\{ z \in \mathbb{C} : z = \frac{1}{\sqrt{1+\mathrm{i}\varepsilon/k_0}}\mathrm{e}^{-(\mathrm{i}+\varepsilon/k_0)\theta} \;\textit{for some}\; \theta \in [0,\infty] \right\}.$$

Theorem 8. *The number of eigenvalues of $W_\tau(\sigma_{\varepsilon/k_0})$ which lie close to the piece of the logarithmic spiral of the previous theorem given by $\theta \in (0,\theta_0)$ is*

$$\frac{2\tau}{\pi}\sqrt{(1+\varepsilon^2/k^2)\theta_0} + o(\tau).$$

Note that we are not able to prove something like these two theorems for $W_\tau(\sigma)$ because σ is neither in $L^1(\mathbb{R})$ nor in $C(\dot{\mathbb{R}})$.

6 Pseudospectrum

Fix $\varepsilon > 0$. The ε-pseudospectrum $\mathrm{sp}_\varepsilon B$ of a bounded linear operator B on some complex Hilbert space is the set of all $\mu \in \mathbb{C}$ for which $\|(B - \mu I)^{-1}\| \geq 1/\varepsilon$. The spectrum of B is considered to be a subset of $\mathrm{sp}_\varepsilon B$. If B is a normal operator, then $\mathrm{sp}_\varepsilon B$ is simply the closed ε-neigbourhood of $\mathrm{sp}\, B$. However, for non-normal operators this is in general no longer the case, and for such operators the pseudospectrum is in many instances of even greater use than the spectrum [22]. The notion of the psedospectrum was independently invented several times [22], and one of these inventions was made by Landau [17] when studying the Fox–Li operator. We first state a simple result from [7].

Theorem 9. *Given $\varepsilon > 0$, there is a $\tau_0 > 0$ such that $\mathrm{sp}_\varepsilon W_\tau(\sigma) \supset \overline{\mathbb{D}}$ for $\tau > \tau_0$.*

This theorem may be restated as follows. Given $\varepsilon > 0$ and $\mu \in \mathbb{D}$, there is a $\tau_0 > 0$ such that for every $\tau > \tau_0$ we can find $u_\tau \in L^2(0, \tau)$ satisfying $\|u_\tau\| = 1$ and $\|W_\tau(\sigma)u_\tau - \mu u_\tau\| \leq \varepsilon$. The following theorem is Landau's [17]. He takes μ from the unit circle $\mathbb{T}$ and is able to say much more in this case.

Theorem 10. *Given $\varepsilon > 0$, $\mu \in \mathbb{T}$, and $C > 0$, there exists a $\tau_0 > 0$ such that for every $\tau > \tau_0$ there are at least $C\tau$ functions $u_{\tau,n}$ which form an orthonormal system in $L^2(0, \tau)$ and satisfy $\|W_\tau(\sigma)u_{\tau,n} - \mu u_{\tau,n}\| \leq \varepsilon$. Moreover, if μ_1 and μ_2 are distinct points on $\mathbb{T}$, then these functions corresponding to μ_1 and μ_2 can be chosen to be mutually orthogonal.*

Landau [17] writes that this theorem "shows that for large Fresnel number ω the laser cannot be expected to settle to a single mode." Physical features of the pseudospectrum of the Fox–Li operator are also discussed in the work by Sir Michael Berry and his co-workers; see, e.g., [2, 3].

7 Challenges

So what are the big open problems for the Fox–Li operator we are, all progress notwithstanding, left with? Here are a few of them. (a) Determine the absolute value of the outmost or better of the outmost and next eigenvalues. (b) Prove that the eigenvalues cluster, in some sense, along a spiral. (c) Prove that this spiral migrates towards the unit circle as $\tau \to \infty$. (d) Determine the shape of the spiral. Is it as conjectured by Vainshtein (6), is it related to theta-three as tabled in [6], or is it something completely different? (e) Describe the density of the eigenvalue distribution along the spiral. (f) Determine the eigenfunctions: numerical indications in [10] are that the eigenfunctions corresponding to leading eigenvalues are trigonometric functions superimposed with low-amplitude rapid oscillation, while for small eigenvalues the eigenfunctions are wave packets.

These questions are of course also of interest for the operator with the original kernel function (2).

We should emphasize that, with the exception of problems (d) and (e), these questions have all been solved numerically. Approaching the Fox–Li operator numerically is not a triviality, since this involves working with highly oscillatory integrals. That Cochran and Hinds [12] were able to show us the spirals as early as 1974 must in this light be appreciated as an admirable feat. Since then numerical methods for highly oscillatory integral equations have been elaborated by many mathematicians, and by now the apparatus is well developed to overcome nearly all subtleties caused by high frequencies. We refer to the recent papers [9, 10] and the references therein for more on the computational mathematics for the Fox–Li and related operators.

Finally, we repeat that two peculiarities of the Fox–Li operator are that its kernel is complex-symmetric and that it depends only on the difference of the arguments. To gain deeper insight into the Fox–Li operator it seems therefore reasonable first to attain greater command of simpler operators with such kernels. In [8], we accordingly considered Wiener–Hopf operators with even and rational symbols. These are given by (5) where $\varrho(t)$ is a finite sum of terms of the form $p_n(|t|)\mathrm{e}^{-\gamma_n|t|}$ with polynomials p_n and complex numbers γ_n such that $\operatorname{Re}\gamma_n > 0$. The symbol $a = \hat{\varrho}$ is an even and rational function in $L^1(\mathbb{R}) \cap C(\dot{\mathbb{R}})$. Hence, by what was outlined in Sect. 5, $\operatorname{sp} W_\tau(a)$ converges to the curve $\mathcal{R}(a)$ formed by the range of a in the Hausdorff metric. However, in the case at hand we can say more. There are explicit formulae for the Fredholm determinants of Wiener–Hopf operators with rational symbols. Given a and under additional technical assumptions, we used these formulae to construct a certain function $b{:}(0,\infty) \to \mathbb{C}$ and to prove that there is a numbering $\{\mu_n\}_{n=1}^{\infty}$ of the eigenvalues of $W_\tau(a)$ such that, with $\xi_n := n\pi/\tau$,

$$\mu_n = a(\xi_n) + \frac{1}{2\tau}a'(\xi_n)\arg b(\xi_n) - \frac{\mathrm{i}}{2\tau}a'(\xi_n)\log|b(\xi_n)| + O(1/\tau^2).$$

Note that the tangent to $\mathcal{R}(a)$ through $a(\xi_n)$ has the parametric representation $\mu = a(\xi_n) + a'(\xi_n)t$, $t \in \mathbb{R}$, and increasing values of the parameter t provide the tangent with an orientation. The point $a(\xi_n) + (1/2\tau)a'(\xi_n)\arg b(\xi_n)$ lies on this tangent. It follows that, up to the $O(1/\tau^2)$ term, the eigenvalue μ_n is located on the right of the tangent if $|b(\xi_n)| > 1$, while it is on the left of the tangent if $|b(\xi_n)| < 1$. Furthermore, the eigenfunctions for an eigenvalue μ_n are shown to be linear combinations of $\mathrm{e}^{\mathrm{i}z_j x}$ where $z_j \in \mathbb{C}$ ranges over the finite set of solutions of the algebraic equation $a(z) = \mu_n$.

Acknowledgements The second author acknowledges support by CONACYT grants 80503 and 102800.

References

1. E.L. Basor, E.M. Landesman, Biography of H. Widom. Oper. Theory Adv. Appl. **71**, ix–xi (1994)
2. M. Berry, Mode degeneracies and the Petermann excess-noise factor for unstable lasers. J. Mod. Opt. **50**, 63–81 (2003)
3. M. Berry, C. Storm, W. van Saarlos, Theory of unstable laser modes: edge waves and fractality. Opt. Commun. **197**, 393–402 (2001)
4. A. Böttcher, B. Silbermann, *Analysis of Toeplitz Operators*, 2nd edn. (Springer, Berlin/Heidelberg/New York, 2006)
5. A. Böttcher, H. Widom, Two remarks on spectral approximations for Wiener–Hopf operators. J. Integral Equ. Appl. **6**, 31–36 (1994)
6. A. Böttcher, H. Brunner, A. Iserles, S. Nørsett, On the singular values and eigenvalues of the Fox–Li and related operators. N.Y. J. Math. **16**, 539–561 (2010)
7. A. Böttcher, S. Grudsky, D. Huybrechs, A. Iserles, First-order Trace Formulae for the Iterates of the Fox-Li Operator. Panor. Mod. Operator Theory Relat. Top. 207–224 (2012). Springer
8. A. Böttcher, S. Grudsky, A. Iserles, Spectral theory of large Wiener–Hopf operators with complex-symmetric kernels and rational symbols. Math. Proc. Camb. Philos. Soc. **151**, 161–191 (2011)
9. H. Brunner, A. Iserles, S. P. Nørsett, The spectral problem for a class of highly oscillatory Fredholm integral equations. IMA J. Numer. Anal. **30**, 108–130 (2010)
10. H. Brunner, A. Iserles, S.P. Nørsett, The computation of the spectra of highly oscillatory Fredholm integral operators. J. Integral Equ. Appl. (2010), to appear
11. J.A. Cochran, The existence of eigenvalues for the integral equations of laser theory. Bell Syst. Tech. J. **44**, 89–110 (1965)
12. J.A. Cochran, E.W. Hinds, Eigensystems associated with the complex-symmetric kernels of laser theory. SIAM J. Appl. Math. **26**, 776–786 (1974)
13. A.G. Fox, T. Li, Resonance modes in a maser interferometer. Bell Syst. Tech. J. **40**, 453–488 (1961)
14. U. Grenander, G. Szegő, *Toeplitz Forms and Their Applications* (University of California Press, Berkeley/Los Angeles, 1958)
15. P. Hartman, A. Wintner, The spectra of Toeplitz's matrix. Am. J. Math. **76**, 867–882 (1954)
16. H. Hochstadt, On the eigenvalues of a class of integral equations arising in laser theory. SIAM Rev. **8**, 62–65 (1966)
17. H. Landau, The notion of approximate eigenvalues applied to an integral equation of laser theory. Q. Appl. Math. **35**, 165–172 (1977/1978)
18. H. Landau, H. Widom, Eigenvalue distribution of time and frequency limiting. J. Math. Anal. Appl. **77**, 469–481 (1980)
19. D.J. Newman, S.P. Morgan, Existence of eigenvalues of a class of integral equations arising in laser theory. Bell Syst. Tech. J. **43**, 113–126 (1964)
20. D. Slepian, Some asymptotic expansions for prolate spheroidal wave functions. J. Math. Phys. **44**, 99-140 (1965)
21. P. Tilli, Some results on complex Toeplitz eigenvalues. J. Math. Anal. Appl. **239**, 390–401 (1999)
22. L.N. Trefethen, M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators* (Princeton University Press, Princeton, 2005)
23. L.A. Vainshtein, Open resonators for lasers. Sov. Phys. JETP **40**, 709–719 (1963)
24. H. Widom, On a class of integral operators with discontinuous symbol. Oper. Theory Adv. Appl. **4**, 477–500 (1982)
25. H. Widom, Eigenvalue distribution of nonselfadjoint Toeplitz matrices and the asymptotics of Toeplitz determinants in the case of nonvanishing index. Oper. Theory Adv. Appl. **48**, 387–421 (1990)

Kähler Metrics with Cone Singularities Along a Divisor

S.K. Donaldson

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract We develop some analytical foundations for the study of Kähler metrics with cone singularities in codimension one. The main result is an analogue of the Schauder theory in this setting. In the later parts of the paper we discuss connections with the existence problem for Kähler–Einstein metrics,in the positive case.

1 Introduction

Let D be a smooth divisor in a complex manifold X. In this paper we study Kähler metrics on $X \setminus D$ with cone singularities of cone angle $2\pi\beta$ transverse to D, where $0 < \beta < 1$. The case we have primarily in mind is when X is a Fano manifold, D is an anticanonical divisor and the metrics are Kähler–Einstein; the motivation being the hope that one can study the existence problem for *smooth* Kähler–Einstein metrics on X (as a limit when β tends to 1) by deforming the cone angle. This can be seen as a variant of the standard "continuity method". We will make some more remarks about this programme in Sect. 6 but it is clear that, at the best, a substantial amount of work will be needed to carry this through—adapting much of the standard theory to the case of cone singularities. This paper is merely a first step along this road. Our goal is to set up a linear theory and apply it to the problem of deforming the cone angle (Theorem 2 below). In further papers with X-X Chen, we will study more advanced and sophisticated questions.

S.K. Donaldson (✉)
Department of Mathematics, Imperial College, Queen's Gate, London SW7 2AZ, UK
e-mail: s.donaldson@imperial.ac.uk

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_4, © Springer-Verlag Berlin Heidelberg 2012

There are several precedents for this line of work. First and foremost, singular metrics of this kind have been considered before by Jeffres [6, 7] and Mazzeo [11]. Some applications to algebraic geometry are outlined by Tian in [4]. Mazzeo considers the case of negative first Chern class, but this makes no difference in the elementary foundational questions we consider here (until Sect. 6). As in this paper, Mazzeo's main emphasis is on the linear theory, and he outlines an approach using the "edge calculus". However this assumes some specialised background, some complications with the choice of function spaces are reported and [11] does not give quite enough detail for those not expert in the techniques to easily fill in the proofs. Thus we have decided to make a fresh start here on the analysis, using elementary methods. This means that we are very probably re-deriving many results that are well-known to experts, and our conclusions are entirely consistent with those described by Mazzeo. It is very likely that the edge calculus, or similar technology, will be important in developing more refined analytical results.

A second precedent occurs in the study of 3-dimensional hyperbolic manifolds. Here again one can consider metrics with cone singularities transverse to a knot. A strategy, similar to ours in Kähler geometry, for constructing nonsingular hyperbolic metrics via deformation of the cone angle was proposed by Thurston and there are a number of papers in the literature developing the theory and the relevant analysis ([5, 12, 13] for example). A third precedent occurs in gauge theory and the work of Kronheimer, Mrowka and others on connections with codimension-2 singularities [8]. In the case when the underlying manifold is complex, this is related to the theory of holomorphic bundles with parabolic structures [2] and there are some closer parallels with our situation.

The general scheme of this paper mimics the development of standard theory for smooth manifolds. We begin by considering a "flat model" for a cone singularity and in Sect. 2 we obtain an estimate in Hölder spaces for the Laplace operator, as in the usual Schauder theory. This depends on certain properties of the Green's function which are derived in Sect. 3, using Bessel functions and classical methods. In Sect. 4 we introduce complex structures, considering first a flat model and then a general class of singular metrics on a pair (X, D). What we achieve is roughly, a parallel to the standard theory of Hölder continuous Kähler metrics. This degree of regularity suffices to give a Fredholm theory linearising the Kähler–Einstein equation, and in particular we can proceed to study the problem of deforming the cone angle. Naturally we expect that it will be possible to say much more about the local structure of these solutions but we mainly leave this for future papers. Sections 5 and 6 are intended to provide context. In Sect. 5 we use the Gibbons–Hawking construction, combined with our study of the Green's function, to produce certain almost-explicit Ricci-flat metrics with cone singularities, analogous to ALE spaces in the usual theory. In Sect. 6 we outline what one might expect when X is the complex projective plane blown up at one or two points—when no smooth Kähler–Einstein metrics exist—and discuss connections with work of Szekelyhidi and Li.

The author is grateful to Xiu-Xiong Chen, Mark Haskins and Jared Wunsch for discussions related to this work.

2 A Schauder Estimate

For $\alpha \in (0, 1)$ and for a function f on $\mathbf{R}^m$ we define

$$[f]_\alpha = \sup_{p,q} \frac{|f(p) - f(q)|}{|p - q|^\alpha}, \tag{1}$$

where $\sup = \infty$ is allowed. Write $\mathbf{R}^m = \mathbf{R}^2 \times \mathbf{R}^{m-2}$ and let $S = \{0\} \times \mathbf{R}^{m-2}$. Take polar co-ordinates r, θ on $\mathbf{R}^2$ and standard co-ordinates s_i on $\mathbf{R}^{m-2}$. Fix $\beta \in (0, 1)$ and consider the singular metric

$$g = dr^2 + \beta^2 r^2 d\theta^2 + \sum ds_i^2. \tag{2}$$

This is the *standard cone metric* with cone angle $2\pi\beta$ and a singularity along S. We want to consider the Green's operator of the Laplacian $\Delta = \Delta_g$. To fix a definition, let H be the be the completion of C_c^∞ under the Dirichlet norm $\|\nabla f\|_{L^2}$. Since the metric g is uniformly equivalent to the standard Euclidean one, we get the same space H using either metric. The Sobolev inequality implies that for $q = 2m/(m + 2)$ and any $\rho \in L^q$ the linear form

$$f \mapsto \int f\rho, \tag{3}$$

is bounded with respect to the H norm, so there is a unique $G\rho \in H$ such that

$$\int f\rho = \int (\nabla f, \nabla G\rho)_g, \tag{4}$$

which is to say that $\phi = G\rho$ is a weak solution of the equation $\Delta_g \phi = \rho$. Thus we define a linear map $G : L^q \to H$.

Proposition 1. *There is a locally-integrable kernel function $G(x, y)$ such that*

$$G\rho(x) = \int G(x, y)\rho(y)dy, \tag{5}$$

for $\rho \in C_c^\infty$. The function $G(x, y)$ is smooth away from the diagonal and points $x, y \in S$.

This follows from standard theory, but in the next section we will give an "explicit" formula for G.

Let D be one of the differential operators

$$\frac{\partial^2}{\partial s_i \partial s_j} \quad \frac{\partial^2}{\partial r \partial s_i}\ , \ \frac{1}{r}\frac{\partial^2}{\partial \theta \partial s_i}. \tag{6}$$

We define $T = D \circ G$. Let $\mu = \beta^{-1} - 1$. The main result of this section is

Theorem 1. *Fix α with $0 < \alpha < \mu$. Then there is a constant C depending on β, n, α such that for all functions $\rho \in C^\infty_c(\mathbf{R}^m)$ we have*

$$[T\rho]_\alpha \leq C[\rho]_\alpha. \tag{7}$$

(The statement should be interpreted as including the assertion that $T\rho$ is continuous, so its value at each point is defined.)

Note When we refer to the distance $d(x, y) = |x - y|$ between points in $\mathbf{R}^m$ we always mean the standard Euclidean distance. However this is uniformly equivalent to the distance defined by the singular metric.

The proof of the Theorem uses an integral representation for T. Let $K(x, y) = D_x G(x, y)$ where the notation means that the differentiation is applied to the first variable. Then we have

Proposition 2. *If $\rho \in C^\infty_c$ and $\rho(x_0) = 0$ for some $x_0 \in \mathbf{R}^m$ then $K(x_0, \ )\rho(\ )$ is integrable and*

$$(T\rho)(x_0) = \int K(x_0, y)\rho(y)dy. \tag{8}$$

Of course the subtlety is that if ρ does not vanish at x_0 then $K(x_0, \ )\rho(\ )$ is not integrable and the formula has to be interpreted as a singular integral, but we will not need to use this approach. What we do need is some more detailed information about the kernel K, summarised in the next Proposition. We write $\pi : \mathbf{R}^2 \times \mathbf{R}^{m-2} \to \mathbf{R}^2$ for the projection map.

Proposition 3. *There are $\kappa_1, \kappa_2, \kappa_3, \kappa_4$ with the following properties.*

- *If $|z| = 1$ then*
$$|K(0, z)| \leq \kappa_1 \tag{9}$$
- *If $|z| = 1$ then*
$$|K(w_1, z) - K(w_2, z)| \leq \kappa_2|w_1 - w_2|^\mu \tag{10}$$
for any w_1, w_2 with $|w_i| \leq 1/2$
- *If $|z| = 1$ and $|\pi(z)| \geq 1/2$ then*
$$|K(z, w)| \leq \kappa_3|z - w|^{-n}, \tag{11}$$
for w with $|w| \leq 5$.
- *If $|z| = 1$ and $|\pi(z)| \geq 1/2$ then*
$$|(\nabla K)(z, w)| \leq \kappa_4|z - w|^{-n-1}, \tag{12}$$
for w with $|w| \leq 5$. Here the derivative ∇K is taken with respect to the first variable.

Propositions 2 and 3 will be established in the next section but now, assuming them, we go on to the proof of Theorem 1. This is a variant of the standard proof of the Schauder estimate for the ordinary Laplace operator.

What is crucially important is that T commutes with dilations. Thus, given $\lambda > 0$ and a function ρ on $\mathbf{R}^n$ we define $\rho_\lambda(x) = \rho(\lambda^{-1}x)$ and we have $(T\rho)_\lambda = T(\rho_\lambda)$. This implies that

$$K(\lambda x, \lambda y) = \lambda^{-n} K(x, y). \tag{13}$$

Note also that the Hölder seminorm scales by dilation as $[f_\lambda]_\alpha = \lambda^{-\alpha}[f]_\alpha$ so our problem is scale invariant.

Fix a smooth function ψ supported in the unit ball, with $\Delta_g \psi$ and $D\psi$ both smooth and with $\Delta_g \psi = 1$ on the δ-ball for some fixed $\delta > 0$. For example we can take $\psi = a(r)b(s)$ where $a(r)$ is equal to 1 for small r and b is a suitable function of s. Set $\chi = \Delta_g \psi$ so χ has compact support, is equal to 1 on the δ-ball and $T\chi = D\psi$ is smooth. We write $[\chi]_\alpha = c_0$, $[T\chi]_\alpha = c_1$.

By scale invariance and linearity, it suffices to show that if $\rho \in C_c^\infty$ has $[\rho]_\alpha = 1$ and if $x_1, x_2 \in \mathbf{R}^m$ with $|x_1 - x_2| = 1$ then $|\rho(x_1) - \rho(x_2)| \leq C$. Let d be a minimum of $(|\pi(x_1)|, |\pi(x_2)|)$. We consider two cases: Case A, when $d \leq 2$, and Case B, when $d > 2$.

Case A. Let x_1', x_2' be the projections of x_1, x_2 to S. Then we can write

$$T\rho(x_1) - T\rho(x_2) = \big(T\rho(x_1) - T\rho(x_1')\big) + \big(T\rho(x_1') - T\rho(x_2')\big) + \big(T\rho(x_2') - T\rho(x_2)\big) \tag{14}$$

and $|x_1 - x_1'|, |x_1' - x_2'|, |x_2' - x_2|$ are all bounded by 3. Using this, and translation and scale invariance, it suffices to consider two sub-cases

Sub-case A1 $x_1 = 0, |x_2| = 1,\ x_2 \in S$.

Sub-case A2 $x_1 = 0, |x_2| = 1,\ x_2' = 0$. (That is, x_2 lies in $\mathbf{R}^2 \times \{0\}$.)

But to begin with the same discussion applies to either sub-case. We define $\sigma_0 = \rho(x_2)\chi_\lambda$ where

$$\lambda = \max(\delta^{-1}, |\rho(x_2)^{1/\alpha}|). \tag{15}$$

We also define

$$\sigma_1 = (\rho(0) - \rho(x_2))\chi. \tag{16}$$

Then $[\sigma_0]_\alpha, [\sigma_1]_\alpha \leq c_0$ and $[T\sigma_0]_\alpha, [T\sigma_1]_\alpha \leq c_1$, using the fact that $|\rho(x_2) - \rho(0)| \leq [\rho]_\alpha = 1$, by hypothesis. Now set $\rho' = \rho - \sigma_0 - \sigma_1$. Thus ρ' vanishes at 0 and x_2 and we have

$$[\rho']_\alpha \leq [\rho]_\alpha + 2c_0 = 1 + 2c_0\ ,\ \ |T\rho(x_2) - T\rho(0)| \leq |T\rho'(x_2) - T\rho'(0)| + 2c_1. \tag{17}$$

This means that, simplifying notation, we can reduce to the situation where ρ vanishes at x_2 and 0. Thus, in this situation, we want to estimate

$$\int K(0,y)\rho(y)dy - \int K(x_2,y)\rho(y)dy \tag{18}$$

which is dominated by

$$I = \int |K(x_2,y) - K(0,y)|\,|\rho(y)|dy. \tag{19}$$

Consider the contribution from the region $|y| \geq 2$. By the homogeneity we have

$$K(x_2,y) - K(0,y) = |y|^{-n}\left(K\left(\frac{x_2}{|y|}, \frac{y}{|y|}\right) - K\left(0, \frac{y}{|y|}\right)\right), \tag{20}$$

and the second item in Proposition 3 gives

$$|K(x_2,y) - K(0,y)| \leq \kappa_2 |y|^{-\mu-n}. \tag{21}$$

Then we get a bound on the contribution to I from $\{|y| \geq 2\}$ in the form

$$\int_2^\infty \kappa_2 R^{-\mu-n} R^\alpha R^{n-1} dR, \tag{22}$$

which is finite since $\alpha < \mu$.

Next we have to estimate the contribution to I from $\{|y| \leq 2\}$. First we consider

$$I_1 = \int_{|y|\leq 2} |K(0,y)\rho(y)|dy. \tag{23}$$

By the homogeneity and the first item of Proposition 3 we have

$$|K(0,y)| \leq \kappa_1 |y|^{-n}, \tag{24}$$

and $|\rho(y)| \leq |y|^\alpha$ so

$$I_1 \leq \kappa_1 \int_{|y|\leq 2} |y|^{-n+\alpha} \tag{25}$$

which is finite. The final step is to estimate

$$I_2 = \int_{|y|\leq 2} |K(x_2,y)\rho(y)|dy. \tag{26}$$

This is where we use different arguments in the two sub-cases. In sub-case A1, when x_2 lies in S, the estimate is just the same as for I_1 above, using translation invariance in the $\mathbf{R}^{m-2}$ factor. In sub-case A2, when x_2 is in the orthogonal complement of S, we use the third item of Proposition 3 to get

$$|K(x_2,y)| \leq \kappa_3 |y - x_2|^{-n} \tag{27}$$

when $|y| \leq 2$ and so

$$|K(x_2, y)||\rho(y)| \leq \kappa_3 |y - x_2|^{\alpha - n} \tag{28}$$

and we can proceed as before. This completes the proof for Case A.

Case B. Recall that we have x_1, x_2 with $|x_1 - x_2| = 1$ and $|\pi(x_i)| > 2$. Set

$$\lambda = \max(|x_1|, |x_2|, |\rho(x_2)|^{1/\alpha}) \tag{29}$$

and define $\sigma_0 = \rho(x_2)\chi_\lambda$. Then $\sigma_0(x_1) = \sigma_0(x_2) = \rho(x_2)$ and we have bounds on $[\sigma_0]_\alpha, [T\sigma_0]_\alpha$ as before. It is clear that we can choose a function $\tilde{\psi}$, supported in the unit ball centred at x_1, with $\Delta_g \tilde{\psi}$ equal to 1 in a small neighbourhood of x_1 and in such a way that $[\tilde{\psi}]_\alpha, [\Delta\tilde{\psi}]_\alpha$ are bounded by fixed constants, independent of x_1 provided only that $|\pi(x_1)| > 2$ (that is, x_1 stays well away from the singular set). Then we put $\tilde{\chi} = \Delta_g \tilde{\psi}$ and

$$\rho' = \rho - (\sigma_0 + (\rho(x_1 - \rho(x_2))\tilde{\chi}. \tag{30}$$

Arguing just as before, we are reduced to the situation where $\rho(x_1) = \rho(x_2) = 0$.

Now we can obviously suppose that x_1 is the point closest to S and by translation we can suppose that $|x_1| = d$. We have to estimate the integral I, as before. We consider the contribution from three regions

- Points y with $|y| > 2d$. This goes just as in Case A, using the second item in Proposition 3, and rescaling.
- Points y with $|y - x_1| \leq 2$. This goes just as before using the third item in Proposition 3 and rescaling.
- Points y with $|y| \leq 2d$ and $|y - x_1| > 2$.

Here we use the fourth item in Proposition 3. Set $z_i = x_i/d$ and $w = y/d$. The fourth item in Proposition 3 gives a bound on the derivative of $K(z, w)$ with respect to z for all points z on the segment joining z_1, z_2. For such points the distance $|z - w|$ is comparable to $|z_1 - w|$ so, integrating the bound gives

$$|K(z_1, w) - K(z_2, w)| \leq \kappa d^{-1} |w - z_1|^{-n-1}. \tag{31}$$

since the distance $|z_1 - z_2|$ is d^{-1}. Scaling back we have

$$|K(x_1, y) - K(x_2, y)| \leq \kappa |y - x_1|^{-n-1}. \tag{32}$$

Now we can bound the contribution to I from this region by

$$\kappa \int_2^{3d} R^{-n-1} R^\alpha R^{n-1} dR < \infty, \tag{33}$$

where $R = |y - x_1|$.

3 Representation of the Green's Functions by Bessel Functions

Write $c = \beta^{-1}$ and consider the map $\iota : \mathbf{R}^2 \times \mathbf{R}^{m-2} \to \mathbf{R}^2 \times \mathbf{R} \times \mathbf{R}^{m-2}$ defined by $\iota(r\cos\theta, r\sin\theta, s) = (r^c \cos\theta, r^c \sin\theta, r^2, s)$. For an open subset $\Omega \subset \mathbf{R}^m$ we say that function f on Ω is β-smooth if each point of Ω has a neighbourhood $N \subset \Omega$ such that the restriction of f to N is the composite of ι and a smooth function in the ordinary sense on a neighbourhood of $\iota(N)$. We define the notion of convergence of β-smooth functions similarly. For fixed y write $\Gamma_y = G(\ , y)$. Then we have

Proposition 4. *If y is not in Ω then Γ_y is β-smooth on Ω and Γ_y varies continuously with y, with respect to the topology of β-smooth functions on Ω.*

If the point y is not in S then we can identify the metric g in a neighbourhood of y with the usual Euclidean metric and it follows from standard theory that Γ_y differs from the usual Newton potential by a smooth (in fact harmonic) function. It is straightforward to deduce from Proposition 4, this observation, and the symmetries and scaling behaviour of the Green's function that our kernel K satisfies the criteria stated in Proposition 3. Likewise for the proof of Proposition 2. The main point of interest is the second item of Proposition 3: this is the only place where the number μ, and hence the restriction on the range of the Hölder exponent, appears. The derivative of the map

$$(r\cos\theta, r\sin\theta) \mapsto (r^c\cos\theta, r^c\sin\theta) \tag{34}$$

is Hölder continuous with exponent μ. Then the chain rule shows that for a β-smooth function f the derivatives $\frac{\partial f}{\partial r}$ and $r^{-1}\frac{\partial f}{\partial \theta}$ are Hölder continuous with this exponent. It follows that, for each choice of differential operator D, the derivative $D\Gamma_y$ is $C^{,\mu}$ near the singular set (the derivatives in the s_i variables being harmless).

Granted the assertions above, we will focus for the rest of this Section on the proof of Proposition 4. We achieve this by showing that the Green's function has a "polyhomogeneous expansion" around the singular set. This must be considered a standard fact. Knowing the Green's function in our problem is essentially the same as knowing the Green's function for the Dirichlet problem for the ordinary Laplace equation on the product of a wedge of angle $2\pi\beta$ in $\mathbf{R}^2$ with $\mathbf{R}^{m-2}$ and, at least when $m = 3$, this is a topic with a large classical literature (see for example [3]). Equally, such polyhomogeneous expansions are prominent in the general theory of edge operators, as applied in [11]. But, lacking an elementary reference for exactly the result we want, we will include a proof here. The proof involves traditional methods of separation of variables and a check on convergence.

We pause for a moment to recall some facts about Bessel functions. Our main reference is [17]. We fix $\nu \geq 0$. The Bessel equation for $f(z)$ is

$$f'' + z^{-1}f' + (1 - \nu^2 z^{-2})f = 0. \tag{35}$$

The Bessel function $J_\nu(z)$ is defined by a series expansion

$$J_\nu(z) = \sum_{j=0}^{\infty} \frac{(-1)^j (z/2)^{\nu+2j}}{j!(\nu+j)!}, \tag{36}$$

and satisfies the Bessel equation. (Here and below we use the notation $a! = \Gamma(a+1)$ for the generalised factorial function). The asymptotic behaviour for large real z is $J_\nu \sim \sqrt{\frac{2}{\pi z}} \cos z$. We define $J_{-\nu}$ by the same formula with ν replaced by $-\nu$. Then $J_\nu, J_{-\nu}$ are two solutions of the Bessel equation. They can be seen as roughly analogous to $\cos z, \sin z$. The linear combination

$$h_\nu(z) = e^{\nu\pi i/2} J_{-\nu}(z) - e^{-\nu\pi i/2} J_\nu(z), \tag{37}$$

has the property that it decays rapidly at infinity on the upper half-plane: it is roughly analogous to e^{iz}. We write $I_\nu(z) = e^{-\nu\pi i/2} J_\nu(iz)$ and

$$K_\nu(z) = \frac{\pi}{2\sin(\nu\pi)} h_\nu(iz) = \frac{\pi}{2\sin\nu\pi} \left(I_{-\nu}(z) - I_\nu(z)\right). \tag{38}$$

This formula can be extended to the case when ν is an integer by taking a suitable limit.

From (4), we have a convergent expansion

$$I_\nu(z) = \sum_{j=0}^{\infty} \frac{1}{j!(\nu+j)!} \left(\frac{z}{2}\right)^{\nu+2j}, \tag{39}$$

and I_ν has asymptotic behaviour for large positive z

$$I_\nu(z) \sim \frac{e^z}{\sqrt{2\pi z}}. \tag{40}$$

The function K_ν has the asymptotic behaviour for large positive z

$$K_\nu(z) \sim \sqrt{\frac{\pi}{2z}} e^{-z} \tag{41}$$

but is unbounded near $z = 0$. For $\nu > 0$.

$$K_\nu(z) \sim \frac{(\nu-1)!}{2} \left(\frac{z}{2}\right)^{-\nu} \quad z \to 0, \tag{42}$$

and $K_0(z) \sim -\log z$. Our main tool will be the integral representation,

$$K_\nu(z) = \frac{1}{2} \int_{-\infty}^{\infty} e^{-z\cosh u + \nu u} du. \tag{43}$$

With this background in place, we proceed to analyse the Green's function by separation of variables. To begin with we argue formally, but in the end when we check convergence it will be clear that everything is watertight. Note that on grounds of symmetry we can write

$$G(r,\theta,s;r',\theta',s')=\sum_{k\geq 0}G_k(r,r',R)\ \cos k(\theta-\theta'), \tag{44}$$

where $R=|s-s'|$. We want to find formulae for the G_k and we will usually write $\nu=ck$. Our Laplace operator can be written as $\Delta_g\phi=\Delta_\beta\phi+\Delta_{\mathbf{R}^{m-2}}\phi$ where $\Delta_{\mathbf{R}^{m-2}}$ is the ordinary Laplacian on $\mathbf{R}^{m-2}$ and Δ_β is the operator in the plane defined by

$$\Delta_\beta\phi=\phi_{rr}+\frac{1}{r}\phi_r+\frac{1}{\beta^2r^2}\phi_{\theta\theta}. \tag{45}$$

This means that $\phi=J_\nu(\lambda r)e^{ik\theta}$ is an eigenfunction for Δ_β, with $\Delta_\beta\phi=-\lambda^2\phi$. The Fourier-Bessel representation of a general function in terms of these eigenfunctions leads to a formula for the heat kernel associated to the operator Δ_β as

$$\sum_{k=0}^{\infty}H_k\cos k(\theta-\theta') \tag{46}$$

where

$$H_k(r,r')=\pi^{-1}\int_0^\infty e^{-\lambda^2t}J_\nu(\lambda r)J_\nu(\lambda r')d\lambda. \tag{47}$$

Now the heat kernel on a product is the product of the heat kernels, so the heat kernel of the Laplacian Δ_g on $\mathbf{R}^m$ is

$$(2\pi t)^{1-m/2}e^{-R^2/4t}\left(\sum H_k(r,r')\cos k(\theta-\theta')\right). \tag{48}$$

We assume that $m\geq 3$. Then the Green's function can be obtained by integrating the heat kernel with respect to the time parameter. Thus

$$G_k(r,r',R)=\int_0^\infty\int_0^\infty(2\pi t)^{1-m/2}e^{-\lambda^2t-R^2/4t}J_\nu(\lambda r)J_\nu(\lambda r')\ d\lambda dt. \tag{49}$$

Changing variable by $t=\frac{R}{2\lambda}e^u$ and using (9) we see that

$$G_k=\frac{1}{(2\pi)^m}R^{2-m/2}g_k \tag{50}$$

where

$$g_k=2\int_0^\infty\lambda^{m/2-2}K_{m/2-2}(R\lambda)J_\nu(r\lambda)J_\nu(r'\lambda)d\lambda. \tag{51}$$

The integral is convergent for all r, r' provided that $R > 0$. We get another representation by rotating the integration path. Suppose that $r < r'$ and write

$$\sin(\nu\pi)J_\nu(r'\lambda) = \mathrm{Im}(e^{-\nu\pi i/2}h_\nu(r'\lambda)). \tag{52}$$

Thus

$$g_k = \mathrm{Im}\left[\int_0^\infty \lambda^{m/2-2}K_{m/2-2}(2R\lambda)\frac{e^{-\nu\pi i/2}}{\sin\nu\pi}h_\nu(r'\lambda)J_\nu(r\lambda)d\lambda\right]. \tag{53}$$

Because of the rapid decay of h_ν over the upper half plane we can rotate the integration path to the positive imaginary axis, which is the same as replacing λ by $i\lambda$ in the integral. We get another expression

$$g_k = 2\int_0^\infty \lambda^{m/2-2}J_{m/2-2}(R\lambda)K_\nu(r'\lambda)I_\nu(r\lambda)d\lambda. \tag{54}$$

This integral converges for any R, provided that $r < r'$.

We will now derive polyhomogeneous expansions for the Green's function in appropriate regions. We need two elementary lemmas.

Lemma 1. *For $p, q \geq 0$ we have*

$$\int_0^\infty K_p(2x)x^{p+q}dx \leq (p+q)!. \tag{55}$$

To see this, use the integral formula (9) to write the integral as

$$I = \frac{1}{2}\int_0^\infty\int_{-\infty}^\infty e^{-2x\cosh u+pu}x^{p+q}dxdu. \tag{56}$$

Now change the order of integration and perform the x integral to get

$$I = \frac{(p+q)!}{2}\int_{-\infty}^\infty \frac{e^{pu}}{(2\cosh u)^{p+q+1}}du. \tag{57}$$

Divide the integral into the two ranges $u \leq 0, u \geq 0$ and use the inequality $2\cosh u \geq e^{-u}$ on the first and $2\cosh u \geq e^u$ on the second. We get

$$I \leq \frac{(p+q)!}{2}\left(\int_{-\infty}^0 e^{(2p+q+1)u}du + \int_0^\infty e^{-(q+1)u}du\right). \tag{58}$$

The right hand side is

$$\frac{(p+q)!}{2}\left(\frac{1}{q+1}+\frac{1}{2p+q+1}\right) \leq (p+q)!. \tag{59}$$

Lemma 2. *There is a universal constant C such that for all $p, q \geq 1$ we have*

$$\frac{(p+q)!}{p!q!} \leq C 2^{p+q}. \tag{60}$$

When p, q are integers this follows immediately from the binomial theorem, with $C = 1$. The same argument applies when one of p, q is an integer. Very likely we can always take $C = 1$ but the author has not found this in texts so we give an ad hoc argument. Set $f(x) = (1+x)^p(1-x)^q$. We have an identity

$$\int_{-1}^{1} f(x)dx = \frac{2^{p+q+1}}{p+q+1}\frac{p!q!}{(p+q)!}. \tag{61}$$

(See [17], p. 225.) Since $(1+h^{-1})^h$ converges as $h \to \infty$ there is a universal constant $\delta > 0$ such that if $0 \leq x \leq 1/2q$ we have $(1-x)^q \geq \delta$, and if $-1/2p \leq x \leq 0$ we have $(1+x)^q \geq \delta$. So if $-1/2p \leq x \leq 1/2q$ we have $f(x) \geq \delta$. Thus the integral of $f(x)$ is at least $\frac{\delta}{2}(p^{-1} + q^{-1})$. This gives

$$2^{p+q}\frac{p!q!}{(p+q)!} \geq \frac{\delta}{4}(p+q+1)(p^{-1}+q^{-1}) \geq \delta/2. \tag{62}$$

First consider the representation arising from (11), when $r < r'$. It is convenient to normalise to $r' = 2$. Write $J_{m/2-2}(x) = x^{m/2-2}F(x)$, so F is bounded for positive real x. Then using the series representation (6) for I_ν and integrating term-by-term we get

$$G = \sum_{j,k} a_{j,k}(R) r^{\nu+2j} \cos k(\theta - \theta') \tag{63}$$

where

$$a_{j,k}(R) = \frac{1}{2^{\nu+2j}}\frac{1}{j!(\nu+j)!}\int_0^\infty \lambda^{\nu+2j+m-4}F(R\lambda)K_\nu(2\lambda)d\lambda. \tag{64}$$

Thus, by Lemma 1,

$$|a_{j,k}(R)| \leq C'\frac{1}{2^{\nu+2j}}\frac{(\nu+2j+m-4)!}{j!(\nu+j)!}, \tag{65}$$

where $C' = \sup|F|$.

Now, using Lemma 2,

$$|a_{j,k}(R)| \leq CC'\frac{(\nu+2j+m-4)!}{(\nu+2j)!}. \tag{66}$$

It is then elementary that the sum on the right hand side of (12) does converge absolutely provided that $r < 1$. For general r' we can use the scaling behaviour to deduce that $G = \left(\frac{2}{r'}\right)^{m-3} \sum a_{j,k}\left(\frac{2R}{r'}\right)\left(\frac{r}{2r'}\right)^{\nu+2j} \cos k(\theta-\theta')$ and the sum converges absolutely if $r < r'/2$.

For the other representation (10), we normalise to $R = 1$. Then we expand both the Bessel functions $J_\nu(r\lambda)$, $J_\nu(r'\lambda)$ in powers of λ and, arguing in a similar way, we have to consider a sum $\sum_{j,j',k} M_{j,j',k}$ where

$$M_{j,j',k} = \left(\frac{r}{2}\right)^{\nu+2j} \left(\frac{r'}{2}\right)^{\nu+2j'} \frac{(p+2\nu+2j+2j')!}{j!j'!(\nu+j)!(\nu+j')!}. \tag{67}$$

For $A, B, C, D > 1$ we can write

$$\begin{aligned} \frac{(A+B+C+D)!}{A!B!C!D!} &= \frac{(A+B+C+D)!}{(A+B)!(C+D)!} \frac{(A+B)!}{A!B!} \frac{(C+D)!}{C!D!} \\ &\le C^3 2^{2(A+B+C+D)}, \end{aligned} \tag{68}$$

by three applications of Lemma 2. Thus

$$\frac{(2\nu+2j+2j')!}{j!j'!(\nu+j)!(\nu+j')!} \le C^3 2^{2(2\nu+2j+2j')}. \tag{69}$$

Again, elementary arguments show that the sum of $M_{j,j',k}$ converges absolutely provided $r, r' < 1/2$. Scaling back: in the region $r, r' < R/2$ we get a convergent polyhomogeneous expansion

$$G = \sum b_{j,j',k}(R) r^{\nu+2j} (r')^{\nu+2j'} \cos k(\theta-\theta'). \tag{70}$$

We can now finish the proof of Proposition 4. We consider an open set Ω and y not in Ω. We know from standard elliptic regularity that we only need to verify the β-smoothness condition at points x_0 in $\Omega \cap S$. Let Ω' be the ball of radius $d/10$ about x_0 where d is the distance to the boundary of Ω. We want to prove that for any y not in Ω the function Γ_y restricted to Ω' is the composite of ι and a smooth function. But for any such y at least one of the two expansions above is valid over Ω'. It is straightforward to see that the formal series we have considered define weak solutions of the equation characterising Γ_y and since we have verified local convergence it follows that the sums represent valid formulae pointwise. Now the series obviously define functions on appropriate neighbourhoods in $\mathbf{R}^{m-2} \times \mathbf{R}^2 \times \mathbf{R}$. That is, a polyhomogeneous series $\sum a_{j,k}(s) r^{\nu+2j} \cos k(\theta)$ defines a smooth function of (s, ζ, ρ) by

$$\sum a_{j,k}(s) \rho^j \mathrm{Re}(\zeta^k), \tag{71}$$

which gives the required factorisation when we set $\rho = r^2, \zeta = r^c e^{i\theta}$. Similar considerations show that Γ_y varies continuously in the desired sense as y varies.

4 Application to Kähler–Einstein Equations

4.1 Flat Model

Write $\mathbf{C}_\beta$ for the Riemannian manifold with underlying space $\mathbf{C}$, on which we take a standard co-ordinate ζ, and with the singular metric associated to the 2-form $\beta^2|\zeta|^{2(\beta-1)} i d\zeta d\overline{\zeta}$. Then the map $\zeta = r^c e^{i\theta}$ (recall that we write $c = \beta^{-1}$) gives an isometry from the standard cone metric $dr^2 + \beta^2 r^2 d\theta^2$ to $\mathbf{C}_\beta$. Likewise, when $m = 2n$ we get an isometry between the singular metric we considered above on $\mathbf{R}^{2m}$ and the Riemannian product $\mathbf{C}_\beta \times \mathbf{C}^{n-1}$. Let $\sigma_1, \dots, \sigma_{n-1}$ be standard complex co-ordinates on $\mathbf{C}^{n-1}$. Thus we have two natural systems of co-ordinates (r, θ, σ_a) and (ζ, σ_a).

We consider the $i\partial\overline{\partial}$-operator on the complement of S, mapping functions to (1,1) forms. Set $\epsilon = dr + i\beta r d\theta$. Then, up to a factor of $\sqrt{2}$, the forms $\epsilon, d\sigma_1, \dots, d\sigma_{m-1}$ give an orthonormal basis for the $(1,0)$ forms at each point. We should keep in mind that ϵ is *not* a holomorphic 1-form , although $c r^{c-1} e^{i\theta} \epsilon = d\zeta$ is. Now take a trivialisation of the $(1,1)$ forms by sections

$$d\sigma_a \wedge d\overline{\sigma}_b \ , \ d\sigma_a \wedge \epsilon \ , \ d\overline{\sigma}_a \wedge \overline{\epsilon} \ , \ \epsilon \wedge \overline{\epsilon}. \tag{72}$$

Up to scale factor, this is a unitary trivialisation. With respect to this trivialisation the components of $i\partial\overline{\partial}$ are all operators D of the kind considered above, except for

$$D_0 = i\left(r^{-1}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \beta^{-2} r^{-1}\frac{\partial^2}{\partial\theta^2}\right). \tag{73}$$

which is the $i\epsilon \wedge \overline{\epsilon}$ component of $i\partial\overline{\partial}$. Of course this is just the Laplacian Δ_β in the $\mathbf{R}^2$ variable, with respect to the singular metric. So

$$\Delta_g = D_0 + \Delta_{\mathbf{C}^{n-1}} \tag{74}$$

in an obvious notation. Since, by definition, $\Delta_g G\rho = \rho$ we can write

$$D_0 G\rho = \rho - \Delta_{\mathbf{C}^{n-1}} \rho. \tag{75}$$

The operator $\Delta_{\mathbf{C}^{n-1}}$ is a sum of terms of the form allowed in Sect. 2 so we get a Hölder estimate on $D_0 G\rho$ and hence on $i\partial\overline{\partial} G\rho$. So we have

Corollary 1. *Suppose $\alpha < \mu = (\beta^{-1} - 1)$. Then there is a constant C depending only on m, β, α such that for all $\rho \in C^\infty_c$ we have*

$$[i\partial\overline{\partial}(G\rho)]_\alpha \leq C[\rho]_\alpha, \tag{76}$$

where the left hand side is interpreted using the trivialisation above.

Notice that it follows from our discussion of the Green's function that the components of $i\partial\overline{\partial}G\rho$ corresponding to the basis elements $\epsilon \wedge d\sigma_a$ tend to zero on the singular set S.

4.2 Further Local Theory

Corollary 1 expresses the essential fact that we are after, but for applications we need a variety of other statements which will be set out here. One detail is that the smooth functions are not dense in Hölder spaces. But any $C^{,\alpha}$ function can be approximated by smooth functions in the norm of $C^{,\underline{\alpha}}$ for any $\underline{\alpha} < \alpha$. So in the end this complication becomes irrelevant and we will ignore it. Suppose that ρ is a $C^{,\alpha}$ function with support in the unit ball $B \subset \mathbf{C}_\beta \times \mathbf{C}^{n-1}$. Then $i\partial\overline{\partial}G\rho$ is $C^{,\alpha}$ and the same estimate as in Corollary 1 holds. As in our discussion of $i\partial\overline{\partial}$, we say that the derivative of a function f is in $C^{,\alpha}$ if the components $\frac{\partial f}{\partial r}, r^{-1}\frac{\partial f}{\partial \theta}$ and $\frac{\partial f}{\partial s_i}$ are $C^{,\alpha}$. Similar arguments to those of Sect. 2 (but easier)show that in the situation above $G\rho$ and $\nabla G\rho$ are in $C^{,\alpha}$. In fact the same argument show that $G\rho, \nabla G\rho$ are in $C^{,\overline{\alpha}}$ for any $\overline{\alpha}$ with $\overline{\alpha} < \mu$ and we have an estimate

$$[G\rho]_{\overline{\alpha}} + [\nabla G\rho]_{\overline{\alpha}} \leq C[\rho]_\alpha. \tag{77}$$

Taking $\overline{\alpha} > \alpha$ we get a compactness result: for a sequence ρ_i, supported on B and bounded in $C^{,\alpha}$, there is a subsequence $\{i'\}$ such that $G\rho_{i'}$ and $\nabla G\rho_{i'}$ converge in $C^{,\alpha}$ over compact sets. Notice also that, as in the remark following Corollary 1, the components of $\nabla G\rho$ corresponding to the derivatives $\frac{\partial f}{\partial r}, r^{-1}\frac{\partial f}{\partial \theta}$ tend to zero on the singular set.

Now consider the situation where we have a function $\phi \in C^{,\alpha}(B)$ such that $\Delta\phi$, defined pointwise outside S, is also $C^{,\alpha}$. Applying standard elliptic estimates in small balls in the complement of S we see that $|\nabla\phi| = O(r^{-1+\alpha})$ near the singular set. One easy consequence is that $\Delta\phi$, defined pointwise as above, agrees with the weak, distributional, notion. For another we take a smooth cut-off function χ of compact support in B, equal to 1 on some interior region B' and with $\Delta\chi$ smooth. Then $\Delta(\chi\phi)$ is in L^q so $G\Delta(\chi\phi)$ is defined. It follows that $\chi\phi = G\rho_1 + G\rho_2$ where $\rho_1 = (\Delta\chi)\phi + \chi\Delta\phi$ and $\rho_2 = 2\nabla\chi.\nabla\phi$. Away from the support of $\nabla\chi$ it is clear that $G\rho_2$ is in $C^{,\alpha}$. Thus we see that $i\partial\overline{\partial}\phi$ is locally in $C^{,\alpha}$ and we obtain interior estimates of the form

$$[i\partial\overline{\partial}\phi]_{\alpha,B'} \leq C\left([\Delta\phi]_{\alpha,B} + [\phi]_{\alpha,B}\right), \tag{78}$$

where C depends on B'. Similarly we get

$$[\nabla\phi]_{\overline{\alpha},B'} \leq C\left([\Delta\phi]_{\alpha,B} + [\phi]_{\alpha,B}\right) \tag{79}$$

for any $\overline{\alpha} < \mu$.

Now let η be a $C^{,\alpha}$ section of the bundle of $(1,1)$ forms, in the sense we have defined, and consider the operator $\Delta_\eta\phi = \Delta\phi + \eta.i\partial\overline{\partial}\phi$. We suppose first that η is supported on B and is sufficiently small in $C^{,\alpha}$. It follows from the usual Neumann series argument that we can invert Δ_η and that an estimate corresponding to Corollary 1 holds. Then we can extend all the results above to Δ_η. As usual, if we have any η which vanishes at the origin we can reduce to the situation where η is small and of compact support by dilation and multiplying by a cut-off function and thus obtain the interior estimate near the origin.

4.3 Global Set-Up

Let X be a compact Kähler manifold and $D \subset X$ be a smooth hypersurface. Let $\Lambda \to X$ be the holomorphic line bundle associated to D, so there is a section s of Λ cutting out D. Let h_Λ be any smooth hermitian metric on Λ and write

$$\chi = i\partial\overline{\partial}|s|^{2\beta}_{h_\Lambda}. \tag{80}$$

Let Ω_0 be a smooth Kähler metric on X. Then we have

Lemma 3. *For sufficiently small $\delta > 0$ the $(1,1)$ form $\omega_0 = \Omega_0 + \delta\chi$ is positive on $X \setminus D$. The metric we obtain is independent of the choices of $\Omega_0, h_\Lambda, \delta$ up to quasi-isometry.*

This is elementary to check and we omit the proof. If we choose standard complex co-ordinates ζ, σ_a around a point of D, so that D is defined by the equation $\zeta = 0$, then $|s|^2_{h_\Lambda} = F|\zeta|^2$ where F is a smooth positive function of ζ, σ_a. Thus

$$\chi = (i\partial\overline{\partial}F^\beta)|\zeta|^{2\beta} + i\beta|\zeta|^{2(\beta-1)}\left(\zeta\partial F^\beta d\overline{\zeta} - \overline{\zeta\partial}F^\beta d\zeta\right) + \beta^2F^\beta|\zeta|^{2(\beta-1)}id\zeta d\overline{\zeta}. \tag{81}$$

Lemma 3 implies that there is a well-defined notion of a Hölder continuous function, with exponent α, on $X \setminus D$, using the singular metric. If we take a standard local complex co-ordinate system ζ, σ_a as above and then set $z = \zeta|\zeta|^{\beta-1}$ then this becomes the ordinary notion of Hölder continuity in terms of the co-ordinates z, σ_a. We write $C^{,\alpha,\beta}$, or sometimes just $C^{,\alpha}$ for these functions on X. Now we want to go on to define Hölder continuous differential forms. With a fixed metric h_λ as above, define the $(1,0)$-form

$$\eta = \partial|s|^\beta. \tag{82}$$

Then we say that a $(1,0)$ form on $X \setminus D$ is Hölder continuous with exponent α if and only if it can be written as

$$f_0\eta + f_1\pi_1 + \cdots + f_N\pi_N, \tag{83}$$

where $f_i \in C^{,\alpha,\beta}$ and π_i are smooth forms (in the ordinary sense) on X.

Lemma 4. *If $\alpha < \mu$ this notion is independent of the choice of metric h_Λ.*

If we make another choice of metric we get another form $\eta' = \partial(f|s|^\beta)$ for a smooth positive function f. Then

$$\eta' = \partial f|s|^\beta + f\eta. \tag{84}$$

The function $|s|^\beta$ is in $C^{,\alpha}$ so η' can be written in the stated form, and the result follows immediately. Similarly ones sees that, in standard co-ordinates $z = re^{i\theta}$, the Hölder continuous (1,0)- forms are just those of the shape $f_0\epsilon + \sum_{a=1}^{n-1} f_a d\sigma_a$, where $\epsilon = dr + i\beta r d\theta$, as in the previous subsection, the co-efficients $f_0, f_1 \dots$ are in $C^{,\alpha}$ and f_0 vanishes on the singular set. Similarly, we can give a global definition of a space of Hölder continuous $(1,1)$ forms which reduces in local co-ordinates (r, θ, σ_a) to those of the shape

$$mi\epsilon\overline{\epsilon} + \sum m_{ab} d\sigma_a d\overline{\sigma}_b + m_a \epsilon d\overline{\sigma}_a + \overline{m}_a \overline{\epsilon} d\sigma_a \tag{85}$$

where m, m_{ab}, m_a are $C^{,\alpha}$ and the m_a vanish on the singular set.

Now we define $\mathcal{C}^{2,\alpha,\beta}$ to be the space of (real-valued) functions f on $X \setminus D$ with $f, \partial f, i\partial\overline{\partial} f$ all Hölder continuous with exponent α. This is the analogue of the usual Hölder space $C^{2,\alpha}$ but there is an important difference that we are not asserting that *all* second derivatives of f are in $C^{,\alpha}$. We can define norms on $C^{,\alpha,\beta}, \mathcal{C}^{2,\alpha,\beta}$ in the usual way, making them Banach spaces. If ω_0 is a singular metric, as constructed above, we have a space of *Hölder continuous Kähler metrics* of the form $\omega_\phi = \omega_0 + i\partial\overline{\partial}\phi$ where $\phi \in \mathcal{C}^{2,\alpha,\beta}$ and we require that $\omega_\phi \geq \kappa\omega_0$ on $X \setminus D$, for some $\kappa > 0$. It is easy to check that this space of metrics is independent of the choice of ω_0.

Let ω be a Hölder continuous Kähler metric as above. In local co-ordinates the metric is described by co-efficients as in (17). All of these have limits along the singular set and by definition the limits of the m_a are zero. The limits of the m_{ab} obviously define a $C^{,\alpha}$ metric on D and the limit of the power $m^{1/\beta}$ is intrinsically a $C^{,\alpha}$ Hermitian metric on the restriction of the line bundle Λ to D (which is identified with the normal bundle of D in X). Given any point $p \in D$ it is clear that we can choose a standard co-ordinate system centred at p so that the $m = 1$ and $m_{\alpha\beta} = \delta_{\alpha\beta}$ at this point. Now write Δ for the Laplace operator of the metric ω. Since it is given by an algebraic contraction of $i\partial\overline{\partial}$ it appears, in these local co-ordinates, in the form Δ_η considered in the previous subsection, and η vanishes at p. So we can apply the results there to obtain interior estimates and inversion operators in sufficiently small balls about this point. From here we can carry through the usual arguments to obtain a parametrix for Δ over all of X. In this way we obtain

Proposition 5. *If $\alpha < \mu = (\beta^{-1} - 1)$ the inclusion $\mathcal{C}^{2,\alpha,\beta} \to C^{,\alpha,\beta}$ is compact. If ω is a $\mathcal{C}^{,\alpha,\beta}$ Kähler metric on (X, D) then the Laplacian of ω defines a Fredholm map $\Delta : \mathcal{C}^{2,\alpha,\beta} \to C^{,\alpha,\beta}$.*

From now on we restrict attention to the case when X is a Fano manifold, $[\Omega_0] = 2\pi c_1(X)$ and D is in the linear system $-K_X$. We can regard Ω_0 as the curvature form associated to a smooth metric on the dual of K_X. Then ω is the curvature form of a singular metric h_0 on this line bundle and any $\phi \in \mathcal{C}^{2,\alpha,\beta}$ defines another metric $|\ |_\phi = e^\phi h_0$. We identify Λ_D with K_X^{-1}, so s is a section of K_X^{-1}. If ω_ϕ is any Kähler metric on $X \setminus D$ its Riemannian volume form can be regarded as an element of $K_X \otimes \overline{K}_X$ so we get a function $s \otimes \overline{s}\mathrm{Vol}_\omega$ on $X \setminus D$. We say the metric is Kähler–Einstein if

$$s \otimes \overline{s}\mathrm{Vol}_\omega = |s|_\phi^{2\beta}. \tag{86}$$

If this holds then, by standard elliptic regularity, ϕ is smooth on $X \setminus D$ and satisfies $\mathrm{Ric}(\omega_\phi) = \beta\omega_\phi$.

We expect there to be a detailed regularity theory for these Kähler–Einstein metrics around the singular divisor, as outlined by Mazzeo in [11]. We will leave most of the discussion of this to another paper but we want to observe here that the metrics are "smooth in tangential directions". In a local co-ordinate system (z, σ_a) we can choose a local Kähler potential ψ so that the equation becomes

$$(i\partial\overline{\partial}\psi)^n = e^{\beta\psi}. \tag{87}$$

Let ψ' be a derivative with respect to the real or imaginary part of any σ_a. Then ψ' satisfies a linear equation $(\Delta + \beta)\psi' = 0$, so it follows that $i\partial\overline{\partial}\psi'$ is $C^{,\alpha}$. Repeating the argument, we find that all multiple derivatives in these directions satisfy this condition. In particular, the induced metric on D and the metric induced on the restriction of Λ to D are both smooth.

Another simple fact is that a solution of our Kähler–Einstein equation which is in $\mathcal{C}^{2,\alpha,\beta}$ for some $\alpha > 0$ lies in $\mathcal{C}^{2,\overline{\alpha},\beta}$ for all $\overline{\alpha} < \mu = \beta^{-1} - 1$: thus the theory is independent of the choice of exponent α.

4.4 Deforming the Cone Angle

For a Fano manifold X and smooth $D \in |-K_X|$ as above we have:

Theorem 2. *Let $\beta_0 \in (0, 1), \alpha < \mu_0 = \beta_0^{-1} - 1$ and suppose there is a $\mathcal{C}^{2,\alpha,\beta_0}$ solution ω to the Kähler–Einstein equation (18) on (X, D), with $\beta = \beta_0$. If there are no nonzero holomorphic vector fields on X which are tangent to D then for β sufficiently close to β_0 there is a $\mathcal{C}^{2,\alpha,\beta}$ solution to (18) for this cone angle.*

It seems likely that the condition on holomorphic vector fields is always satisfied, by general results from algebraic geometry, but the author has not gone into this. In any case it is not a serious restriction.

The proof of the theorem follows standard general lines. Having set up a linear theory, we can deform the solutions to the nonlinear equation using an implicit function theorem, provided that the linearised operator is invertible. However there are some complications, for example due to the fact that the function spaces depend on β. We have seen that the solution ω defines a smooth metric on Λ over D. We extend this to a smooth metric, which we will write as $\| \|$, on Λ over X. This is not to be confused with the singular metric, which we will write as $| |$, whose curvature is ω. Now for β near to β_0 we define

$$\omega_\beta = \omega + i\partial\overline{\partial}(\|s\|^\beta - \|s\|^{\beta_0}), \tag{88}$$

so $\omega_{\beta_0} = \omega$. In other words, ω_β is the curvature form of the singular metric on K_X^{-1} with

$$|s|^2_\beta = \exp(\|s\|^\beta - \|s\|^{\beta_0})|s|^2. \tag{89}$$

Set

$$k_\beta = |s|_\beta^{-2\beta} s \otimes \overline{s}\ \mathrm{Vol}_{\omega_\beta}. \tag{90}$$

Thus $k_{\beta_0} = 1$, since ω solves the Kähler–Einstein equation. We state three Propositions.

Proposition 6.

$$\|k_\beta - 1\|_{C^{,\alpha,\beta}} \to 0 \tag{91}$$

as $\beta \to \beta_0$.

Write Δ_β for the Laplace operator of ω_β.

Proposition 7. *If* $\Delta_{\beta_0} + \beta_0 : \mathcal{C}^{2,\alpha,\beta_0} \to C^{,\alpha,\beta_0}$ *is invertible then for* β *close to* β_0 *the operator* $\Delta_\beta + \beta : \mathcal{C}^{2,\alpha,\beta} \to C^{,\alpha,\beta}$ *is also invertible and the operator norm of its inverse is bounded by a fixed constant independent of* β.

The statements of Propositions 6 and 7 are not completely precise. There are many ways of defining norms on $C^{\alpha,\beta}, \mathcal{C}^{2,\alpha,\beta}$, all of which are equivalent for fixed β. But what we need here is a definite family of norms, for example defined using a fixed system of co-ordinate charts. But we hope that the details of such a definition will be clear to the reader and do not need to be spelled out.

Proposition 8. *If* $\Delta_{\beta_0} + \beta_0$ *is not invertible then there is a non-trivial holomorphic vector field on* X *tangent to* D.

Given these three results, the proof of Theorem 2 is a standard application of the implicit function theorem.

We begin with the proof of Proposition 6. This is completely elementary, but the set-up is a little complicated. As a first simplification we reduce to considering convergence with respect to the Hölder norm defined by the fixed parameter β_0. That is to say, for any β we are considering a standard chart χ_β mapping a neighbourhood of 0 in $\mathbf{C} \times \mathbf{C}^{n-1}$ to X and the functions in $C^{,\alpha,\beta}$ are those which pull back by χ_β to ordinary $C^{,\alpha}$ functions. The composite $\eta_{\beta,\beta_0} = \chi_\beta^{-1} \circ \chi_\beta$ is the map defined by $(re^{i\theta}, s) \mapsto (r^\lambda e^{i\theta}, s)$, where $\lambda = \beta_0/\beta$. If $\beta > \beta_0$ this is not Lipschitz so the notions of Hölder continuity are different. However, η_{β,β_0} is β_0/β-Hölder and this means that it pulls $C^{,\alpha}$ functions back to $C^{,\alpha\beta_0/\beta}$ functions. Since we are always free to adjust α a little and since we can take β_0/β arbitrarily close to 1, we see that it suffices to prove that

$$\|k_\beta - 1\|_{C^{,\alpha,\beta_0}} \to 0 \tag{92}$$

as $\beta \to \beta_0$.

We will use another, similar, elementary observation below. Supppose that f_i is a sequence of functions on the ball in $\mathbf{C} \times \mathbf{C}^{n-1}$ converging to a limit f_∞ in $C^{,\alpha}$ and with f_i all vanishing on the singular set $\{r = 0\}$. Suppose that $0 < \epsilon_i \leq \epsilon < \alpha$ and $\epsilon_i \to 0$ as $i \to \infty$. Then the functions $r^{-\epsilon_i} f_i$ are Hölder with exponent $\alpha - \epsilon$ and converge in this sense to f_∞ as $i \to \infty$.

With these remarks in place we can begin the proof. We work in a standard local co-ordinate system ζ, σ_a chosen so that section s is given by

$$s = \zeta(d\zeta d\sigma_1 \dots d\sigma_n)^{-1}. \tag{93}$$

Then $\|s\|^2 = F|\zeta|^2$, where F is smooth strictly positive function of ζ, σ_a. Now write, as in (16),

$$i\partial\overline{\partial}(F^\beta|\zeta|^{2\beta}) = F^\beta|\zeta|^{2\beta-2}\tau + V_\beta, \tag{94}$$

say, where $\tau = i d\zeta d\overline{\zeta}$. Of course we can write down a formula for V_β, although it is a little complicated. The point to emphasise is that this just depends on the smooth function F and β. All we need to know is that the (1,1)-forms V_β are $C^{,\alpha,\beta_0}$ forms for β close to β_0, they all vanish on the singular set and they converge to V_{β_0} in this Hölder space sense as $\beta \to \beta_0$. We leave the reader to verify these assertions by straightforward calculation.

Now we can write

$$\omega_\beta = F^\beta|\zeta|^{2\beta-2}\tau + V_\beta + \Omega, \tag{95}$$

where Ω is independent of β. Thus in our standard co-ordinates r, θ, σ_a the form Ω has $C^{,\alpha}$ co-efficients and all co-efficients tend to zero on the singular set except those involving $d\sigma_a d\overline{\sigma}_b$.

Recall that $k_\beta = |s|_\beta^{-2\beta} s \otimes \overline{s}\mathrm{Vol}(\omega_\beta)$. By the definition of our class of Holder continuous metrics we can write

$$|s|^2_{\beta_0} = \|s\|^2 \exp(\psi + \|s\|^{2\beta_0}), \tag{96}$$

where ψ is C^{2,α,β_0}. From this we get

$$\frac{k_\beta}{k_{\beta_0}} = \|s\|^{2(\beta_0-\beta)} \exp((\beta-\beta_0)\psi + \beta\|s\|^{2\beta} - \beta_0\|s\|^{2\beta_0}) \frac{\mathrm{Vol}(\omega_\beta)}{\mathrm{Vol}(\omega_{\beta_0})}. \tag{97}$$

Writing $\|s\|^2 = F|\zeta|^2$ we get

$$\frac{k_\beta}{k_{\beta_0}} = |\zeta|^{2(\beta_0-\beta)} H_\beta \frac{\mathrm{Vol}(\omega_\beta)}{\mathrm{Vol}(\omega_{\beta_0})}, \tag{98}$$

where H_β tends to 1 in $C^{,\alpha,\beta_0}$ as $\beta \to \beta_0$. So it suffices to prove that $|\zeta|^{2(\beta_0-\beta)} \frac{\mathrm{Vol}(\omega_\beta)}{\mathrm{Vol}(\omega_{\beta_0})}$ also tends to 1 in this sense.

For simplicity, to explain the argument, let us suppose that $n = 2$. Write $V_\beta + \Omega = \Omega_\beta$. So Ω_β are (1, 1) forms which vary continuously in $C^{,\alpha,\beta_0}$ for β close to β_0. We take the standard volume form J_0 in our co-ordinates (r, θ, σ_0) to be $J_0 = \tau \wedge d\sigma_1 \dots d\sigma_{n-1} d\overline{\sigma}_1 \dots d\overline{\sigma}_{n-1}$. Then $\omega_{\beta_0}^2 / J_0 \geq \kappa > 1$. Now since $\tau^2 = 0$ we have

$$\omega_\beta^2 = F^\beta |\zeta|^{2(\beta-\beta_0)} \tau \wedge \Omega_\beta + \Omega_\beta^2, \tag{99}$$

so, writing $r = |\zeta|^\beta$,

$$|\zeta|^{2(\beta_0-\beta)} \omega_\beta^2 = F^\beta \tau \wedge \Omega_\beta + r^{2(1-\beta/\beta_0)} \Omega_\beta^2. \tag{100}$$

The crucial thing is that Ω_β^2 / J_0 vanishes on the singular set. Thus we can apply the observation about multiplication above to see that, after slightly adjusting α, the product $r^{2(1-\beta/\beta_0)} \Omega_\beta^2 / J_0$ converges in $C^{,\alpha,\beta_0}$ as β tends to β_0. This completes the proof of Proposition 6.

Next we consider Proposition 8. The first step is to establish a Fredholm alternative: if $(\Delta_{\beta_0} + \beta_0)$ has no kernel in $\mathcal{C}^{2,\alpha,\beta_0}$ then it is surjective (i.e. the Fredholm index is zero). For the corresponding L^2 theory this is straightforward, so what one needs to know is that if ρ is in $C^{,\alpha}$ and f is a weak solution of the equation $(\Delta + \beta_0) f = \rho$ then f is in $\mathcal{C}^{2,\alpha,\beta_0}$. By the results of (Sect. 4.2), this will be true if we can show that f is in $C^{,\alpha,\beta_0}$ and this follows from the general theory developed in [4], Chap. 8. Granted this, the proof of Proposition 8 comes down to showing that a non-trivial solution of $(\Delta_{\beta_0} + \beta_0) f = 0$ defines a non-trivial holomorphic vector field on X, tangent to D. Of course this is standard material in the ordinary, non-singular, case.To simplify notation write $\Delta_{\beta_0} = \Delta$. We write $\mathcal{D}$ for the operator $\overline{\partial} \circ \mathrm{grad}$ over $X \setminus D$, where $\mathrm{grad} f$ denotes the gradient vector field of f with respect to the metric and $\overline{\partial}$ is the $\overline{\partial}$-operator on vector fields. The fact that the Ricci curvature of ω_{β_0} is $\beta_0 \omega_{\beta_0}$ gives an identity

$$\overline{\partial}^* \mathcal{D} f = \mathrm{grad}(\Delta f + \beta_0 f). \tag{101}$$

For $\epsilon > 0$, let X_ϵ be the complement of a tubular neighbourhood of D in X, modelled in standard local co-ordinates on the region $\{r \geq \epsilon\}$. Suppose that $(\Delta + \beta_0) f = 0$. We take the inner product of (19) with $\mathrm{grad} f$ and integrate by parts to get

$$\int_{X_\epsilon} |\mathcal{D} f|^2 = \int_{\partial X_\epsilon} \mathcal{D} f * \mathrm{grad} f, \tag{102}$$

where $*$ denotes a certain bilinear algebraic operation. What we need to see is that the boundary term tends to 0 with ϵ. From that we see that $\mathrm{grad} f$ is a holomorphic vector field on $X \setminus D$. We know that the radial derivative $\frac{\partial f}{\partial r}$ is $O(r^\alpha)$ for and this translates into the fact that the $\frac{\partial}{\partial \zeta}$ component of the vector field, in holomorphic co-ordinates, is $O(|\zeta|^{\alpha\beta-\beta+1})$. The $\frac{\partial}{\partial \sigma_a}$ components are bounded. If α is sufficiently close to $\mu = \beta^{-1} - 1$ then $\alpha\beta - \beta + 1$ is positive and this implies that the vector field extends holomorphically across D and is tangent to D.

So the real task is to check that the boundary term in (20) tends to zero with ϵ. For this we use

Lemma 5. *With the notation above, $|\mathcal{D}f| = O(r^{\alpha-1})$, where r is the distance (in the metric ω_{β_0}) to the divisor D.*

Assuming this Lemma it follows that the integrand $\mathcal{D}f * \mathrm{grad} f$ is $O(r^\alpha)$, because $\mathrm{grad} f$ is bounded so the boundary integral is $O(r^\alpha)$ and the volume of ∂X_ϵ is $O(r)$.

To prove the Lemma we can work in a local chart and there is no loss in taking r to be the radial co-ordinate as before. Given a point p with radial co-ordinate r_0 we consider a small ball B_0 of radius hr_0 centred at p on which we can identify the model cone metric with the flat metric (so h is a fixed small number depending on β_0). We re-scale this small ball to a unit ball $B \subset \mathbf{C}^n$. The Kähler–Einstein metric ω_{β_0} re-scales to a Kähler–Einstein metric $\tilde{\omega}$ on B. The fact that ω_{β_0} is $C^{,\alpha}$ means that the C^0 difference between $\tilde{\omega}$ and a Euclidean metric on B is $O(r_0^\alpha)$. Now standard elliptic regularity for the Kähler–Einstein equations implies that the derivative of $\tilde{\omega}$ is also $O(r_0^\alpha)$ on an interior ball. Scaling back, we see that the derivative of ω_{β_0} is $O(r_0^{\alpha-1})$ at p.

Now consider our function f with $(\Delta + \beta_0)f = 0$. We know that the radial derivative of f is $O(r^\alpha)$ and the tangential derivatives are in $C^{,\alpha}$. Given p as above, let f_0 be the $\mathbf{R}$-linear function of the co-ordinates σ_a defined by the tangential derivative of f at p. Thus the derivative of $g = f - f_0$ is $O(r_0^\alpha)$ over B_0 and the variation of g over B_0 is $O(r_0^{\alpha+1})$. We also have $\Delta g = -\beta_0 f$ since $\Delta f_0 = 0$. By the same kind of argument as before, rescaling and using standard elliptic estimates, we see that $\mathcal{D}g$ is $O(r_0^{\alpha-1})$ at p. On the other hand $\mathcal{D}f_0 = \overline{\partial}(\mathrm{grad} f_0)$ and the definition of $\mathrm{grad} f_0$ involves the metric tensor ω_{β_0}. From this we see that $|\mathcal{D}f_0|$ is bounded by a fixed multiple of the derivative of the metric tensor and so is $O(r_0^{\alpha-1})$ by the preceding discussion. Hence $\mathcal{D}f = \mathcal{D}g + \mathcal{D}f_0$ is $O(r_0^{\alpha-1})$ as required.

Finally we turn to Proposition 8, but here we will be very brief since nothing out of the ordinary is involved. By the Fredholm alternative, it suffices to show that if β_i is a sequence converging to β_0 and if f_i are functions with $\| f_i \|_{C^{2,\alpha,\beta_i}} = 1$ but $\|(\Delta_{\beta_i} + \beta_i) f_i\|_{C^{,\alpha,\beta_i}} \to 0$ as $i \to \infty$ then there is a nontrivial solution to the equation $(\Delta_{\beta_0} + \beta_0)f = 0$. To do this one applies elementary observations about the family of metrics ω_β, like those in the proof of Proposition 6, and standard arguments to get uniform estimates, independent of i.

5 Model Ricci-Flat Solutions

5.1 Digression in Four-Dimensional Riemannian Geometry

Suppose that we have six 2-forms $\omega_1, \omega_2, \omega_3, \theta_1, \theta_2, \theta_3$ on a 4-manifold which satisfy the equations

$$\omega_i \wedge \omega_j = V\delta_{ij} \ , \ \theta_i \wedge \theta_j = -V\delta_{ij} \ , \ \omega_i \wedge \theta_j = 0 \tag{103}$$

where V is a fixed volume form. There is a unique Riemannian metric such that the ω_i form an orthonormal basis for the self-dual forms Λ^+ and θ_j for the anti-self-dual forms Λ^-. We want to discuss the Levi-Civita connection of this metric, viewed as a pair of connections on the bundles Λ^+, Λ^- (that is, using the local isomorphism between $SO(4)$ and $SO(3) \times SO(3)$). Changing orientation interchanges the two bundles so we can work with either and we fix on Λ^-.

Write $d\theta_i = \psi_i$ and consider the linear equations for 1-forms T_i

$$\psi_i = T_j \wedge \theta_k - T_k \wedge \theta_j. \tag{104}$$

(Here, and below, we use the convention that (ijk) runs over the cyclic permutations of (123).) It is a fact that this system of linear equations has a unique solution. This fact is essentially the same as the usual characterisation of the Levi-Civita connection in that the covariant derivative on Λ^- is

$$\nabla\theta_i = T_j \otimes \theta_k - T_k \otimes \theta_j. \tag{105}$$

The solution of (22) is

$$-2T_i = *\psi_i - \theta_j \wedge (*\psi_k) + \theta_k \wedge (*\psi_j), \tag{106}$$

where $*$ is the $*$-operator of the metric. The T_i are connection forms for Λ^- in the local orthonormal trivialisation θ_i. The components of the curvature tensor of Λ^- are the forms

$$F_i = dT_i + T_j \wedge T_k. \tag{107}$$

This gives a way to compute the Riemann curvature tensor which is useful in some situations, such as that below. In particular we can take the anti-self-dual components F_i^- of the F_i and express them in terms of the given basis so

$$F_i^- = \sum_j W_{ij}\theta_j, \tag{108}$$

Then the matrix W_{ij} represents the anti-self-dual Weyl tensor of the Riemannian metric. It is a general fact that this is symmetric and trace-free.

A particular case of this is when the forms θ_i are all closed. Then the T_i vanish and we see that Λ^- is flat. This means that locally we have a hyperkähler metric, although to fit with standard conventions we should change orientation, so we are considering closed forms ω_i. In this situation the only non-vanishing component of the Riemann curvature tensor is the anti-self-dual Weyl tensor, so we can use a basis θ_i to compute the curvature tensor, as above.

5.2 *The Gibbons–Hawking Construction*

We review this well-known construction. We start with a positive harmonic function f on a domain Ω in $\mathbf{R}^3$ and an S^1 bundle P over Ω with a connection whose curvature is $- * df$. Let α be the connection 1-form over P and dx_i the pull-back of the standard 1-forms on $\mathbf{R}^3$. Then we have

$$d\alpha = -\sum f_i dx_j dx_k. \tag{109}$$

(We will write f_i, f_{ij} etc. for the partial derivatives of f.) Set

$$\omega_i = \alpha \wedge dx_i + f dx_j \wedge dx_k. \tag{110}$$

Then $d\omega_i = -f_i dx_i dx_j dx_k + f_i dx_i \wedge dx_j \wedge dx_k = 0$ and it is clear that the forms satisfy $\omega_i \wedge \omega_j = \delta_{ij} V$, with $V = f\alpha \wedge dx_1 \wedge dx_2 \wedge dx_3$, so we have a hyperkähler structure.

One basic example is when $\Omega = \mathbf{R}^3 - \backslash\{0\}$ and $f = 4\pi^{-1}|x|^{-1}$. Then the manifold we construct is $\mathbf{R}^4 \setminus \{0\}$ with the flat metric. If we identify $\mathbf{R}^4$ with $\mathbf{C}^2$ in the usual way, the circle action can be taken to be $(z, w) \mapsto (\lambda z, \lambda^{-1} w)$ and the map from $\mathbf{C}^2$ minus the origin to $\mathbf{R}^3$ given by the identification with P is

$$(z, w) \mapsto (\mathrm{Re}(zw), \mathrm{Im}(zw), |z|^2 - |w|^2). \tag{111}$$

Like the metric, this map extends smoothly over the origin but we get a fixed point of the action, corresponding to the pole of f. In general if we start with a hyperkähler 4-manifold $(M, \omega_1, \omega_2, \omega_3)$ with a circle action which is Hamiltonian with respect to the three symplectic forms then the Hamiltonians $x_i : M \to \mathbf{R}$ define a map $\underline{x} : M \to \mathbf{R}^3$ and we recover the structure on M (at least locally) from a harmonic function with poles.

Now we want to compute the curvature tensor of such a hyperkähler 4-manifold. Set $\theta_i = \alpha \wedge dx_i - f dx_j \wedge dx_k$. Then θ_i form an orthonormal basis for Λ^- as considered in (Sect. 5.1) and

$$d\theta_i = -2 f_i dx_1 dx_2 dx_3. \tag{112}$$

One finds then that the 1-forms T_i are

$$T_i = -\frac{f_i}{f}\alpha + \frac{1}{f^2}(f_j dx_k - f_k dx_j). \tag{113}$$

Computing $dT_i + T_j \wedge T_k$, one finds that the curvature tensor (W_{ij}) in this orthonormal basis is the trace-free part of the matrix

$$\left(\frac{f_{ij}}{f^2} - 3\frac{f_i f_j}{f^3}\right). \tag{114}$$

This can also be written as $2f$ times the trace-free part of the Hessian of the function f^{-2} which checks with the fact that when $f = |x|^{-1}$ the construction yields the flat metric on $\mathbf{R}^4$. For then $f^{-2} = |x|^2$, the Hessian of f^{-2} is twice the identity matrix and so its trace-free part is zero.

5.3 Cone Singularities

Now return to our cone metric on $\mathbf{R}^2 \times \mathbf{R}$ and let f be the Green's function $f(x) = \Gamma_p(x) = G(x, p)$ where $p = (1, 0, 0)$. Locally, away from the singular set, we can identify domains in $\mathbf{R}^2_\beta \times \mathbf{R}$ with domains in $\mathbf{R}^3$ and it is clear that the construction above yields a Ricci-flat metric on an S^1 bundle P over the complement of the singular set and the point p. Another useful way to think about this is to cut the plane along the negative real axis and identify the corresponding cut 3-space with a wedge-shaped region U in standard Euclidean 3-space. We perform the Gibbons–Hawking construction in the usual way over U, with the pole of f yielding a fixed point of the action. Then we reverse the cut we made and glue appropriate points on the boundary of the 4-manifold to get our metric with cone singularity. Either way, the upshot is that we get a 4-manifold $\overline{P}$ with an S^1 action having a single fixed point, a map $\pi : \overline{P} \rightarrow \mathbf{R}^2 \times \mathbf{R}$ and a metric g on $\overline{P}$ with a cone singularity along $\pi^{-1}(S)$.

The metric g is locally hyperkähler but not globally. It has a global Kähler structure ω_1 corresponding to the direction of the edge of the wedge. If we choose local structures ω_2, ω_3 then parallel transport around the singular set takes the complex form $\Theta = \omega_2 + i\omega_3$ to $e^{2\pi(\beta-1)i}\Theta$.

Now we claim that, with this global complex structure, $\overline{P}$ can be identified with $\mathbf{C}^2$ and the singular set $\pi^{-1}(S)$ corresponds to the complex curve $C = \{(z, w) \in \mathbf{C}^2 : zw = 1\}$. For this we begin by going back to the general Gibbons–Hawking construction with a harmonic function f on a domain Ω which we suppose to be the product of a domain in the plane $x_1 = 0$ with an interval about 0 in the x_1 co-ordinate. Trivialise the bundle P by parallel transport in the x_1 direction, so the connection 1-form is $a = a_2 dx_2 + a_3 dx_3$. Write ψ for the angular co-ordinate on the fibres of P. We seek a holomorphic function h on P, for the complex structure corresponding to x_1. In the trivialisation this amounts to solving the equations

$$\frac{\partial h}{\partial x_1} = -if\frac{\partial h}{\partial \psi} \quad , \quad \frac{\partial h}{\partial x_2} + \frac{\partial h}{\partial x_3} = (a_2 + ia_3)h. \tag{115}$$

We look for a solution which has weight 1 for the circle action, $h(\lambda z) = \lambda h(z)$. In this case $\frac{\partial h}{\partial \psi} = ih$ so the first equation gives

$$h(x_1, x_2, x_3, \psi) = e^u h(0, x_2, x_3) e^{i\psi} \tag{116}$$

where

$$u(x_1, x_2, x_3) = \int_0^{x_1} f(t, x_2, x_3) dt. \tag{117}$$

Conversely if we find a solution $h(0, x_2, x_3)$ of the second equation in (24) over the slice $x_1 = 0$ and define h using (25) and (26) then the integrability condition for the complex structure implies that we obtain a solution of (24). In particular suppose that we are in the case when $\frac{\partial f}{\partial x_1}$ vanishes on the plane $x_1 = 0$. This means that we can choose a_2, a_3 to vanish on this plane. Thus, on this plane, the second equation in (24) is the ordinary Cauchy–Riemann equation. Given any holomorphic function $h_0(x_2 + ix_3)$ the formulae (25) and (26) define a holomorphic function h on P. The same discussion applies if we seek a function $\tilde{h}$ which transforms with weight -1. We get another holomorphic function

$$\tilde{h}(x_1, x_2, x_3, \psi) = e^{-u} h_0(0, x_2, x_3) e^{-i\psi}. \tag{118}$$

Thus we get a pair of holomorphic functions $(h, \tilde{h})$ on P with

$$\tilde{h}h = h_0(x_2 + ix_3), \tag{119}$$

or in other words a holomorphic map from P to $\mathbf{C}^2$. This maps the lifts $\psi = 0, \pi$ in P of the 2-dimensional domain in $\{x_1 = 0\}$ to the diagonal $\{z = w\}$ in $\mathbf{C}^2$. It also maps the subset in P lying over any line $x_2 = \xi_2, x_3 = \xi_3$ to the plane curve $zw = h_0^2(\xi_2 + i\xi_3)$.

We apply this discussion to the case when f is the Green's function Γ on $\mathbf{R}^2 \times \mathbf{R}$. Of course we can only immediately fit in with the discussion above locally but we hope that the picture will be clear to the reader. By symmetry, the $\mathbf{R}$ derivative of Γ vanishes on the plane $s = 0$ and we are in the position above. Moreover the symmetry taking s to $-s$ lifts to a symmetry interchanging $h, \tilde{h}$. Of course one has to consider how the local construction above works around the pole, but this is just the same as in the model case of the ordinary Green's function on $\mathbf{R}^3$. In terms of our usual co-ordinates (r, θ) on $\mathbf{R}^2$ we define $h_0 = 1 - r^c e^{i\theta}$. This is holomorphic with respect to the given complex structure on the plane and vanishes at the pole of Γ. The construction above produces global holomorphic functions $h, \tilde{h}$ on $\overline{P}$ with $h = \tilde{h}$ on a (real) 2-plane in P which maps to the plane $s = 0$ in $\mathbf{R}^2 \times \mathbf{R}$ as a double branched cover, branched over the origin. The functions satisfy $h\tilde{h} = 1$ on the singular set. So we get a holomophic map from $\overline{P}$ to $\mathbf{C}^2$ taking the circle action on P to the action $(z, w) \mapsto (\lambda z, \lambda^{-1} z)$ and mapping the singular set to the curve $zw = 1$. The fact that $\Gamma(r, \theta, s)$ decays like s^{-1} as $s \to \infty$, so its indefinite integral with respect to s is unbounded, implies that this map is bijective, by a straightforward argument.

We can also start from the opposite point of view with the complex manifold $\mathbf{C}^2$ and the $\mathbf{C}^*$-action $(z, w) \to (\lambda z, \lambda^{-1} w)$. We consider a locally-defined holomorphic 2-form

$$\Theta = \beta(1-zw)^{\beta-1} dz dw. \tag{120}$$

This is preserved by the $\mathbf{C}^*$-action and, locally, there is a holomorphic Hamiltonian map $H_{\mathbf{C}}(z,w) = (1-zw)^{\beta}$. Although this is not well-defined globally the power $H_{\mathbf{C}}^{1/\beta}$ is so, and this gives the $\mathbf{R}^2$ component of the map from $\mathbf{C}^2$ to $\mathbf{R}^2 \times \mathbf{R}$ which arises from the identification of $\mathbf{C}^2$ with $\overline{P}$.

It would take a little work to check that the metrics we have studied here really do give metrics with cone singularities of the kind we defined in Sect. 4—analysing the local representation in complex co-ordinates, but it seems to the author that this should not be hard.

There are several possible variants of this construction. For example, we can use finite sums of Green's functions to get Ricci-flat Kähler metrics with cone singularities on ALE spaces. The example we have constructed above furnishes a plausible model for certain degenerations of metrics with cone singularities on compact manifolds. Consider a compact complex surface X and a family of curves D_ϵ which converge as $\epsilon \to 0$ to a singular curve D_0 with one ordinary double point at $p \in X$. Suppose there are Kähler–Einstein metrics ω_ϵ with fixed cone angle β along D_ϵ, for $\epsilon \neq 0$. We should expect that, after re-scaling small balls about p, the rescaled metrics converge to the Ricci-flat metric we have discussed above. Thus these kind of Ricci-flat, non-compact model solutions should play the same role in the theory of metrics with cone singularities that the ordinary ALE spaces play in the standard theory.

Another interesting application of these ideas is to supply models for the behaviour of the metrics around the singular set. In particular we can study the growth of the curvature. Looking at (23), we see that the curvature will be dominated by the Hessian of the harmonic function f and from the discussion in Sect. 3 we see that this will typically be $O(r^{c-2}) = O(r^{\beta^{-1}-2})$. Since $\beta^{-1} > 1$ the curvature is, at least locally, in L^2 but if $\beta > 1/2$ the curvature is unbounded. We expect that the like will hold for general Kähler–Einstein metrics with cone singularities.

6 Conjectural Picture

In this final section we discuss what one might expect about the existence problem for metrics with cone singularities on a Fano manifold. It is natural to think of such a metric as a solution of a distributional equation

$$\mathrm{Ric}(\omega) = \beta\omega + 2\pi(1-\beta)[D]. \tag{121}$$

But in writing this equation we emphasise that we mean solutions of the kind we have defined precisely in Sect. 4. This equation can be compared with the equation studied in the standard "continuity method"

$$\mathrm{Ric}(\omega) = \beta\omega + (1-\beta)\rho \tag{122}$$

where ρ is a prescribed closed $(1,1)$ form representing $2\pi c_1(X)$. There are good reasons for believing that the cone singularity problem will always have solutions for *small* positive β. In one direction, Tian and Yau established the existence of a complete Ricci-flat Kähler metric on the non-compact manifold $X \setminus D$ [16] and one could expect that this is the limit of solutions ω_β as β tends to 0 (this idea, in the negative case, is mentioned by Mazzeo in [11]). In another direction, it is known that, at least if X has no holomorphic vector fields, solutions to (28) exist for small β and one could perhaps view (27) as a limiting case. Szekelyhidi [15] introduced an invariant $R(X)$, defined to be the supremum of numbers μ such there is a Kähler metric Ω in the class $2\pi c_1(M)$ with $\mathrm{Ric}(\Omega) \geq \mu\Omega$, pointwise on the manifold. He showed that for any choice of ρ this is also the supremum of the values $\beta \leq 1$ such that a solution of (28) exists. Further, it is known that $R(X) \geq \frac{n+1}{n}\alpha(X)$ where $\alpha(X)$ is Tian's invariant. The natural conjecture then is

Conjecture 1. There is a cone-singularity solution ω_β to (27) for any parameter β in the interval $(0, R(X))$. If $R(X) < 1$ there is no solution for parameters β in the interval $(R(X), 1)$

Note that if $\beta = \nu^{-1}$ for an integer ν, our metrics with cone singularities are *orbifold* metrics, so a great deal of standard theory can be brought to bear. See the recent work [14] of Ross and Thomas, for example.

Suppose that we are in a case when solutions exist for small cone angles but not for cone angles close to 1. We would like to understand how the solutions can break down at some critical cone angle. This leads into a large discussion involving notions of "stability" which we only want to touch on here. Recall that in the established theory one defines the *Futaki invariant* of a Kähler manifold Y with a fixed circle action. One definition is to take any invariant metric in the Kähler class and then set

$$\mathrm{Fut}(Y) = \int_Y (S - \hat{S})H \tag{123}$$

where S is the scalar curvature, $\hat{S}$ is the average value of the scalar curvature and H is the Hamiltonian of the circle action. The key point is that in fact the Futaki invariant does not depend on the choice of metric, in a fixed Kähler class. There are other definitions which generalise to singular spaces and schemes. What is visible from the formula (29) is that if Y admits an invariant metric of constant scalar curvature, in the given class, then the Futaki invariant vanishes, since in that case $S = \hat{S}$.

Now let $\Delta \subset Y$ be a divisor invariant under the circle action and $0 < \beta \leq 1$. We define a Futaki invariant of the data by

$$\mathrm{Fut}(Y, \Delta, \beta) = \mathrm{Fut}(Y) - (1-\beta)\left(\int_\Delta H - \frac{\mathrm{Vol}(\Delta)}{\mathrm{Vol}(X)}\int_X H\right). \tag{124}$$

This definition can be motivated, in the framework of metrics with cone singularity along Δ, by adding a suitable distributional term to the scalar curvature, in the manner of (27) and substituting into (29). Under plausible assumptions about the behaviour around the singular set, the definition implies that if there is an invariant constant scalar curvature metric with cone angle β along D then $\mathrm{Fut}(Y, \Delta, \beta) = 0$. In particular this should apply in the Kähler–Einstein situation.

Conjecture 2. Let X be a Fano manifold and D a smooth divisor in $-K_X$. Suppose $\beta_0 \leq 1$ and there are Kähler–Einstein metrics with cone angle β along D for $\beta < \beta_0$ but not for cone angle β_0. Then the pair (X, D) can be degenerated to a pair (Y, Δ), which has an S^1 action, and $\mathrm{Fut}(Y, \Delta, \beta_0) = 0$.

This conjecture really needs to be fleshed out. In one direction, we should discuss pairs (Y, Δ) with singularities. In another direction, what is really relevant is that the Futaki invariant $\mathrm{Fut}(Y, \Delta, \gamma)$ *decreases* to 0 as β increases to β_0 where the sign of H is linked to the degeneration of (X, D) to (Y, Δ). But the statement conveys the general idea.

We can illustrate this, albeit still at the conjectural level, by considering two rational surfaces X_1, X_2: the blow-ups of $\mathbf{CP}^2$ in one or two points respectively. It is well-known that these do not admit Kähler–Einstein metrics and we will see that the calculation of certain Futaki invariants reproduces explicit known values of the invariants $R(X_i)$ obtained by Szekelyhidi [15] and Li [9].

We begin with the case of X_2, which take to be the blow-up of $\mathbf{CP}^2$ at the points $p = [1, 0, 0], q = [0, 1, 0]$. We take a smooth cubic C in $\mathbf{CP}^2$ through these two points, so the proper transform D of C is a canonical divisor in X_2. In this case the degeneration of the pair (X_2, D) will only involve D, so $Y = X_2$. To obtain Δ we consider the $\mathbf{C}^*$-action on X_2 induced by $[u, v, w] \mapsto [\lambda u, \lambda v, w]$ on $\mathbf{CP}^2$. We define Δ to be the limit of D under the action as $\lambda \to 0$. This is the proper transform of a singular curve C' in $\mathbf{CP}^2$ which is the union of three lines through $r = [0, 0, 1]$ (the lines $\overline{pr}, \overline{qr}$ and one other line). We take the circle action on $Y = X_2$ to be the obvious one defined by the above $\mathbf{C}^*$-action. It is then a straightforward exercise to compute the Futaki invariant $\mathrm{Fut}(X_2, \Delta, \beta)$ as a function of β. To fix signs and constants we take the Hamiltonian H to vanish at r and to take the value 3 on the line at infinity $\{[u, v, 0]\}$. The calculation of $\mathrm{Fut}(X_2)$ is easiest using a toric description. One finds that

$$\mathrm{Fut}(X_2) = -2/3. \tag{125}$$

Likewise

$$\mathrm{Vol}(X_2) = 7/2\,, \mathrm{Vol}(\Delta) = 7 \tag{126}$$

$$\int_{X_2} H = 19/3, \int_{\Delta} H = 17/2 \tag{127}$$

Thus $\mathrm{Fut}(X_2, \Delta, \beta) = -\frac{2}{3} + \frac{25(1-\beta)}{6}$ and this vanishes when $1 - \beta = 4/25$. This fits in with the result of Chi-Li that $R(X_2) = 21/25$. In fact this is not too surprising because the calculation in [9] involves essentially the same ingredients, but from a different point of view.

Notice that this discussion ties in with that in Sect. 5 because the curve Δ is singular. We expect that re-scaling the metrics for parameters $\beta < \beta_0$ around the point r we will get a limit which is a Ricci-flat metric on $\mathbf{C}^2$ with cone angle β_0 along the affine part of C.

There is a similar discussion for X_1. We define this to be the blow-up of $\mathbf{CP}^2$ at r and now we take C to be a smooth cubic through r. This time we degenerate in the opposite direction, taking $\lambda \to \infty$. The limit Δ is a divisor which is the sum of the proper transform of a line through r and the line at infinity, taken with multiplicity 2. With H normalised to be equal to 1 on the exceptional divisor and 3 on the proper transform of the line at infinity, the calculation now yields

$$\mathrm{Fut}(X_1, \Delta, \beta) = \frac{2}{3} - \frac{14(1-\beta)}{3} \tag{128}$$

and we find the critical value $\beta_0 = 6/7$, agreeing with [9, 15]. (The fact that the coefficient of $(1-\beta)$ has different signs in the two cases is connected with the fact that we take limits in opposite directions $\lambda \to 0, \infty$.) The expected behaviour of the Kähler–Einstein metrics as $\beta \to \beta_0$ is less clear in this case and we leave that discussion for another place.

Remark. Towards the end of the writing of this paper, preprints by Berman [1] and Li [10] appeared. These both seem to be very relevant to the discussion of this section, and to give additional evidence for the conjectural picture above.

References

1. R. Berman, A thermodynamic formalism for Monge-Ampère equations, Moser-Trudinger inequalities and Kähler–Einstein metrics. arxiv 1011.3976
2. O. Biquard, Sur les fibrées parabolique sur une surface complexes. J. Lond. Math. Soc. **253**(2) 302–316 (1996)
3. H.S. Carslaw, The Green's function for a wedge of any angle and other problems in the conduction of heat. Proc. Lond. Math. Soc. **8**, 365–374 (1910)
4. D. Gilbarg, N. Trudinger, *Elliptic Partial Differential Equations of Second Order* (Springer, Berlin/New York, 1983)
5. C. Hodgson, S. Kerchoff, Rigidity of hyperbolic cone manifolds and hyperbolic Dehn surgery. J. Differ. Geom. **48**, 1–59 (1998)
6. T. Jeffres, Uniqueness of Kähler–Einstein cone metrics. Publ. Mat. **44**(2), 437–448 (2000)
7. T. Jeffres, Schwarz lemma for Kähler cone metrics. Int. Math. Res. Not. **2000**(7), 371–382 (2000)
8. P. Kronheimer, T. Mrowka, Gauge theory for embedded surfaces, I. Topology **32**(4), 773–826, (1993)
9. C. Li, Greatest lower bounds on Ricci curvature for toric Fano manifolds. arxiv 0909.3443
10. C. Li, On the limit behaviour of metrics in continuity method to Kähler–Einstein problem in toric Fano case. arxiv 1012.5229
11. R. Mazzeo, Kähler–Einstein metrics singular along a smooth divisor. J. Équ. aux Deriv. Partielles **VI**, 1–10 (1999)

12. R. Mazzeo, G. Montcouquiol, Infinitesimal rigidity of cone-manifolds and the Stoker problem for hyperbolic and Euclidean polyhdedra. arxiv 0908.2981
13. G. Montcouquiol, H. Weiss, Complex twist flows on surface group representations and the local structure of the deformation space of hyperbolic cone 3-manifolds. arxiv 1006.5582
14. J. Ross, R. Thomas, Weighted projective embeddings, stbility of orbifolds and constant scalar curvature Kähler metrics. J. Differ. Geom. **88**, 109–159 (2011)
15. G. Szekelyhidi, Greatest lower bounds on the Ricci curvature of Fano manifolds. arxiv 0909.5504
16. G. Tian, S-T. Yau, Complete Kähler metrics with zero Ricci curvature I. J. Am. Math. Soc. **3**, 579–609 (1990)
17. E.T. Whittaker, G.N. Watson, *A Course in Modern Analysis* (Cambridge University Press, Cambridge, 1927)

The Space of Framed Functions is Contractible

Y.M. Eliashberg and N.M. Mishachev

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract According to Igusa (Ann Math 119:1–58, 1984) a *generalized Morse function* on M is a smooth function $M \to \mathbb{R}$ with only Morse and birth-death singularities and a *framed* function on M is a generalized Morse function with an additional structure: a framing of the negative eigenspace at each critical point of f. In (Igusa, Trans Am Math Soc 301(2):431–477, 1987) Igusa proved that the space of framed generalized Morse functions is $(\dim M - 1)$-connected. Lurie gave in (arXiv:0905.0465) an algebraic topological proof that the space of framed functions is contractible. In this paper we give a geometric proof of Igusa-Lurie's theorem using methods of our paper (Eliashberg and Mishachev, Topology 39:711–732, 2000).

1 Framed Igusa Functions

1.1 Main Theorem

This paper is written at a request of Kazhdan and Hinich who asked us whether we could adjust our proof in [5] of Igusa's h-principle for generalized Morse functions from [9] to the case of framed generalized Morse functions considered by Igusa in

Y.M. Eliashberg (✉)
Stanford University, Stanford, CA 94305, USA
e-mail: eliash@math.stanford.edu

N.M. Mishachev
Lipetsk Technical University, Lipetsk, 398055, Russia
e-mail: mishachev@lipetsk.ru

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_5, © Springer-Verlag Berlin Heidelberg 2012

his paper [10] and more recently by Lurie in [11]. We are very happy to devote this paper to Stephen Smale whose geometric construction in [12] plays the central role in our proof (as well as in the proofs of many other h-principle type results).

Given an n-dimensional manifold W, a *generalized Morse function*, or as we call it in this paper *Igusa function*, is a function with only Morse (A_1) and birth-death (A_2) type singularities. A *framing* ξ of an Igusa function $\varphi : W \to \mathbb{R}$ is a trivialization of the negative eigenspace of the Hessian quadratic form at A_1-points which satisfy certain extra conditions at A_2-points, see a precise definition below.

If the manifold W is endowed with a foliation $\mathcal{F}$ then we call $\varphi : (W, \mathcal{F}) \to \mathbb{R}$ a *leafwise Igusa* function if restricted to leaves it has only Morse or birth-death type singularities. A *framing* ξ of a leafwise Igusa function $\varphi : (W, \mathcal{F}) \to \mathbb{R}$ is a leafwise framing; see precise definitions below.

The following theorem is the main result of the paper. We use Gromov's notation $\mathcal{O}p\, A$ for an unspecified open neighborhood of a closed subset $A \subset W$.

Theorem 1 (Extension theorem). *Let W be an $(n+k)$-dimensional manifold with an n-dimensional foliation $\mathcal{F}$. Let $A \subset W$ be a closed (possibly empty) subset and (φ_A, ξ_A) a framed leafwise Igusa function defined on $\mathcal{O}p\, A \supset A$. Then there exists a framed leafwise Igusa function (φ, ξ) on the whole W which coincides with (φ_A, ξ_A) on $\mathcal{O}p\, A$.*

Theorem 1 is equivalent to the fact that the space of framed Igusa functions is contractible, which is a content of J. Lurie's extension (see Theorem 3.4.7 in [11]) of Igusa's result from [10]. Indeed, any family, parameterized by a manifold Q, of framed Igusa functions on a manifold M can be viewed as a framed *leafwise* Igusa function on the manifold $M \times Q$ endowed with a foliation by the fibers of the projection $M \times Q \to Q$. Hence, Theorem 1 implies the contractibility of the space of framed Igusa functions. Conversely, Theorem 1 can be deduced from the contractibility result via the standard h-principle technique, see [8]. The current form of the theorem allows us do not discuss the topology on this space, comp. [10].

1.2 Framed Igusa Functions

Objects associated with a leafwise Igusa function. Let $T\mathcal{F}$ denote the n-dimensional subbundle of TW tangent to the leaves of the foliation $\mathcal{F}$. Let us fix a Riemannian metric on W. Given a leafwise Igusa function (LIF) φ we associate with it the following objects:

- $V = V(\varphi)$ is the set of all its leafwise critical points, i.e. the set of zeros of the leafwise differential $d_{\mathcal{F}}\varphi : W \to T^*\mathcal{F}$.
- $\Sigma = \Sigma(\varphi)$ is the set of A_2-points. Generically, V is a k-dimensional submanifold of W which is transversal to $\mathcal{F}$ at the set $V \setminus \Sigma$ of A_1-points and has the fold type tangency to $\mathcal{F}$ along a $(k-1)$-dimensional submanifold $\Sigma \subset V$ of leafwise A_2-critical points of φ.
- Vert is the restriction bundle $T\mathcal{F}|_V$.

- $d^2_{\mathcal{F}}\varphi$ is the leafwise quadratic differential of φ. It is invariantly defined at each point $v \in V$. $d^2_{\mathcal{F}}\varphi$ can be viewed as a homomorphism Vert $\to$ Vert*. Using our choice of a Riemannian metric we identify the bundles Vert and Vert* and view $d^2_{\mathcal{F}}\varphi$ as a self-adjoint operator Vert $\to$ Vert. This operator is non-degenerate at the points of $V \setminus \Sigma$, and has a 1-dimensional kernel $\lambda \subset \mathrm{Vert}|_{\Sigma}$. Note that λ is tangent to V, and thus we have $\lambda = \mathrm{Vert} \cap TV|_{\Sigma}$.
- $d^3_{\mathcal{F}}\varphi$ is the invariantly defined third leafwise differential, which is a cubic form on λ. For a leafwise Igusa function φ this cubic form is non-vanishing, and hence the bundle λ is trivial and can be canonically oriented by choosing the direction in which the cubic function $d^3_{\mathcal{F}}\varphi$ increases. We denote by λ^+ the unit vector in λ which defines its orientation.

Decomposition of $V(\varphi)$ and splitting of Vert. The index of the leafwise quadratic differential $d^2_{\mathcal{F}}\varphi\,(v)$, $v \in V$, may takes values $0, 1, \dots, n$ for $v \in V \setminus \Sigma$ and $0, 1, \dots, n-1$ for $v \in \Sigma$. Let

$$V \setminus \Sigma = V^0 \cup \cdots \cup V^n \quad \text{and} \quad \Sigma = \Sigma^1 \cup \cdots \cup \Sigma^{n-1}$$

be the decompositions of $V \setminus \Sigma$ and Σ according to the index. Note that Σ^i is the intersection of the closures of V^i and V^{i+1}. Then for $v \in V^i$ we have the splitting

$$T_v\mathcal{F} = \mathrm{Vert}(v) = \mathrm{Vert}^i_+(v) \oplus \mathrm{Vert}^i_-(v)$$

where $\mathrm{Vert}^i_+(v)$ and $\mathrm{Vert}^i_-(v)$ are the positive and the negative eigenspaces of $d^2_{\mathcal{F}}\varphi(v)$, and for any $\sigma \in \Sigma^i$ we have the splitting

$$T_\sigma\mathcal{F} = \mathrm{Vert}(\sigma) = \mathrm{Ver}(\sigma) \oplus \lambda(\sigma) \ = \mathrm{Ver}^i_+(\sigma) \oplus \mathrm{Ver}^i_-(\sigma) \oplus \lambda(\sigma)$$

(Ver $\neq$ Vert !), where $\mathrm{Ver}^i_+(v)$ and $\mathrm{Ver}^i_-(v)$ are the positive and the negative eigenspaces of $d^2_{\mathcal{F}}\varphi\,(\sigma)$. For $\sigma \in \Sigma^i$ and $v \in V^i$ we have

$$\lim_{v \to \sigma} \mathrm{Vert}^i_+(v) = \mathrm{Ver}^i_+(\sigma) \oplus \lambda(\sigma) \quad \text{and} \quad \lim_{v \to \sigma} \mathrm{Vert}^i_-(v) = \mathrm{Ver}^i_-(\sigma)\,.$$

For $\sigma \in \Sigma^i$ and $v \in V^{i+1}$ we have

$$\lim_{v \to \sigma} \mathrm{Vert}^{i+1}_-(v) = \mathrm{Ver}^i_-(\sigma) \oplus \lambda(\sigma) \quad \text{and} \quad \lim_{v \to \sigma} \mathrm{Vert}^{i+1}_+(v) = \mathrm{Ver}^i_+(\sigma)\,.$$

Framing of a leafwise Igusa function. A framing of a leafwise Igusa function φ is an ordered set $\xi = (\xi^1, \dots, \xi^n)$ of unit vector fields in Vert(V) such that:

- ξ^i is defined (only) over the union $\Sigma^{i-1} \cup V^i \cup \cdots \cup \Sigma^{n-1} \cup V^n$;
- $\xi^i|_{\Sigma^{i-1}} = \lambda^+|_{\Sigma^{i-1}}$;
- $(\xi^1, \dots, \xi^i)|_{V^i}$ is an orthonormal framing for Vert^i_-.

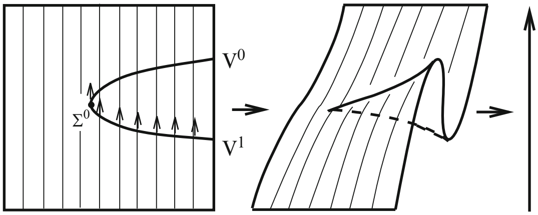


Fig. 1 Framed leafwise Igusa function

In particular, ξ^n is defined only on $\Sigma^{n-1} \cup V^n$ and ξ^1 is defined only on $V \setminus V^0$. The pair (φ, ξ) is called a *framed leafwise Igusa function* (see Fig. 1).

The motivation for adding a framing is discussed in [10].

1.3 Framed Formal Leafwise Igusa Functions

A *formal leafwise Igusa function* (FLIF) is a quadruple $\Phi = (\Phi^0, \Phi^1, \Phi^2, \lambda^+)$ where:

- $\Phi^0 : W \to \mathbb{R}$ is any function;
- $\Phi^1 : W \to T\mathcal{F}$ is a vector field tangent to $\mathcal{F}$, vanishing on a subset $V = V(\Phi) \subset W$;
- Φ^2 is a self-adjoint operator Vert $\to$ Vert, which has rank $n-1$ over a subset $\Sigma = \Sigma(\Phi) \subset V$ and rank n over $V \setminus \Sigma$;
- λ^+ is a unit vector field in the line bundle where $\lambda := \mathrm{Ker}\,(\Phi^2|_{TV|_\Sigma})$.

A leafwise 3-jet of a genuine Igusa function can be viewed as a formal Igusa function Φ, where $\Phi^0 = \varphi$, $\Phi^1 = \nabla_{\mathcal{F}}\varphi$, $\Phi^2 = d^2_{\mathcal{F}}\varphi$ and λ^+ is the unit vector field in Ker $d^2_{\mathcal{F}}$ oriented by the third differential $d^3_{\mathcal{F}}\varphi$. We denote this FLIF Φ by $J(\varphi)$. A FLIF Φ of the form $J(\varphi)$ is called *holonomic*. Thus we can view a genuine Igusa function as a holonomic formal Igusa function. Usually we will not distinguish between leafwise holonomic functions and corresponding holonomic FLIFs.

Given a FLIF Φ we will use the notation similar to the holonomic case. Namely,

- $V^i \subset V \setminus \Sigma$ is the set of points $v \in V \setminus \Sigma$ where the index (dimesion of the negative eigenspace) of Φ^2_v is equal to i , $i = 0, \dots, n$;
- $\Sigma^i \subset \Sigma$ is the set of points $\sigma \in \Sigma$ such that the index of Φ^2_σ is equal to i , $i = 0, \dots, n-1$;
- $T_v\mathcal{F} = \mathrm{Vert}(v) = \mathrm{Vert}^i_+(v) \oplus \mathrm{Vert}^i_-(v)$ where $\mathrm{Vert}^i_+(v)$ and $\mathrm{Vert}^i_-(v)$ are the positive and the negative eigenspaces of Φ^2_v, $v \in V$;
- $T_\sigma\mathcal{F} = \mathrm{Vert}(\sigma) = \mathrm{Ver}(\sigma) \oplus \lambda(\sigma) = \mathrm{Ver}^i_+(\sigma) \oplus \mathrm{Ver}^i_-(\sigma) \oplus \lambda(\sigma)$ where $\mathrm{Ver}^i_+(v)$ and $\mathrm{Ver}^i_-(v)$ are the positive and the negative eigenspaces of Φ^2_σ, $\sigma \in \Sigma^i$.

As in the holonomic case, for $\sigma \in \Sigma^i$ and $v \in V^i$ we have

$$\lim_{v \to \sigma} \mathrm{Vert}^i_+(v) = \mathrm{Ver}^i_+(\sigma) \oplus \lambda(\sigma) \quad \text{and} \quad \lim_{v \to \sigma} \mathrm{Vert}^i_-(v) = \mathrm{Ver}^i_-(\sigma)\,,$$

and so on.

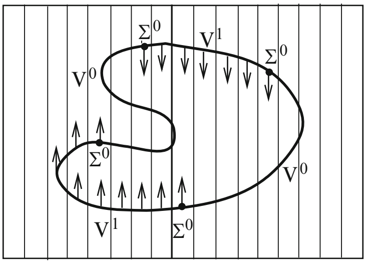


Fig. 2 Framed FLIF

A *framing* for a formal leafwise Igusa function Φ is an ordered set $\xi = (\xi^1, \dots, \xi^n)$ of unit vector fields in Vert(V) such that:

- ξ^i is defined (only) over the union $\Sigma^{i-1} \cup V^i \cup \dots \cup \Sigma^{n-1} \cup V^n$;
- $\xi^i|_{\Sigma^{i-1}} = \lambda^+|_{\Sigma^{i-1}}$;
- $(\xi^1, \dots, \xi^i)|_{V^i}$ is an orthonormal framing for Vert$^i_-$.

The pair (Φ, ξ) is called a *framed* formal leafwise Igusa function (framed FLIF).

As in the holonomic case, for a generic FLIF Φ the set V is a k-dimensional manifold and Σ its codimension 1 submanifold. However, Σ has nothing to do with tangency of V to $\mathcal{F}$, and moreover there is no control of the type of the tangency singularities between V and $\mathcal{F}$ (see Fig. 2).

In what follows we will need to consider FLIFs for different foliations on W. We will say that Φ is an $\mathcal{F}$-FLIF when we need to emphasize the corresponding foliation $\mathcal{F}$. Moreover, the notion of a FLIF can be generalized without any changes to an arbitrary, not necessarily integrable n-dimensional distribution $\zeta \subset TW$. We will call such an object a ζ-FLIF. In the case when a distribution ζ is integrable and integrates into a foliation $\mathcal{F}$ we will use as synonyms both terms: ζ-FLIF and $\mathcal{F}$-FLIF.

Push-forward operation for FLIFs. Let $\zeta, \widetilde{\zeta}$ be two n-dimensional distributions in TW. Let $f : W \to W$ be a diffeomorphism covered by an isomorphism $F : \zeta \to \widetilde{\zeta}$. Let Φ be a ζ-FLIF. Then we define the push-forward $\widetilde{\zeta}$- FLIF $\widetilde{\Phi} = (f, F)_*\Phi = (\widetilde{\Phi}^0, \widetilde{\Phi}^1, \widetilde{\Phi}^2, \widetilde{\lambda}^+)$ as

- $\widetilde{\Phi}^0(f(x)) := \Phi^0(x),\ x \in W$;
- $\widetilde{\Phi}^1_{f(x)}(F(Z)) = \Phi^1_x(Z),\ x \in W,\ Z \in \zeta_x$;
- $\widetilde{\Phi}^2_{f(x)}(F(Z)) = F(\Phi^2_x(Z)),\ x \in V,\ Z \in \mathrm{Vert}_x = \zeta_x$;
- $\widetilde{\lambda}^+(f(x)) = F(\lambda^+(x)),\ x \in \Sigma$.

If Φ is framed then the push-forward operator $(f, F)_*$ transforms its framing ξ to a framing $\widetilde{\xi}$ of $\widetilde{\Phi}$ in a natural way:

- $\widetilde{\xi}^i(f(x)) = F(\xi^i(x)),\ x \in V$.

Note that if ζ and $\widetilde{\zeta}$ are both integrable, i.e. tangent to foliations $\mathcal{F}$ and $\widetilde{\mathcal{F}}$, $F = df$ and Φ is holonomic i.e. $\Phi = J(\varphi)$ then $\widetilde{\Phi}$ is also holonomic, $\widetilde{\Phi} = J(\widetilde{\varphi})$, where $\widetilde{\varphi} = \varphi \circ f^{-1}$.

1.4 Outline of the Proof and Plan of the Paper

Any framed leafwise Igusa function can be extended from $\mathcal{O}p\,A$ to W *formally*, i.e. as a framed FLIF (Φ, ξ), see Theorem 2. This is, essentially, an original Igusa's observation from [10]. We then gradually improve (Φ, ξ) to make it holonomic. Note that unlike the holonomic case, the homotopical data associated with Φ^1 and Φ^2 are essentially unrelated. We formulate the necessary so-called *balancing* homotopical condition for a FLIF to be holonomic, see Sect. 3.4, and show that one can always make a FLIF (Φ, ξ) balanced via a modification, called *stabilization*, see Sect. 3.6.

Our next task is to arrange that $V(\varphi)$ has fold type tangency with respect to the foliation $\mathcal{F}$, as it is supposed to be in the holonomic case. FLIFs satisfying this property, together with certain additional coorientation conditions over the fold, are called *prepared*, see Sect. 3.1. We observe that for a prepared FLIF one can define a stronger necessary homotopical condition for holonomicity. We call prepared FLIFs satisfying this stronger condition *well balanced*, see Sect. 3.4.

Given any FLIF Φ one can associate with it a *twisted normal bundle* (also called *virtual vertical bundle*) ${}^{\Phi}\mathrm{Vert}$ over $V = V(\Phi)$ which is a subbundle of $TW|_V$ obtained by twisting the normal bundle of V in W near $\Sigma = \Sigma(\Phi)$, see Sect. 3.2. In the holonomic case we have ${}^{\Phi}\mathrm{Vert} = \mathrm{Vert}$, see Lemma 7. A crucial observation is that the manifold V has fold type tangency to any extension ζ of the bundle ${}^{\Phi}\mathrm{Vert}$ to a neighborhood of V, see Lemma 5. Moreover, if Φ is balanced then there exist a global extension ζ of ${}^{\Phi}\mathrm{Vert}$ and a bundle isomorphism $F : \mathrm{Vert} \to {}^{\Phi}\mathrm{Vert}$ homotopic to the identity $\mathrm{Vert} \to \mathrm{Vert}$ through injective bundle homomorphisms into $TW|_V$ such that the push-forward framed ζ-FLIF $(\widetilde{\Phi}, \widetilde{\xi}) = (\mathrm{Id}, F)_*(\Phi, \xi)$ is well balanced, see Lemma 11. If Φ is holonomic on $\mathcal{O}p\,A$ then the bundle ζ and the framed FLIF $(\widetilde{\Phi}, \widetilde{\xi})$ coincide with $T\mathcal{F}$ and (Φ, ξ) over $\mathcal{O}p\,A$.

The homotopy of the homomorphism F generates a homotopy of distributions ζ_s connecting ζ and $T\mathcal{F}$. If it were possible to construct a fixed on $\mathcal{O}p\,A$ isotopy V_s of V in W keeping V_s folded with respect to ζ_s then one could cover this homotopy by a fixed on $\mathcal{O}p\,A$ homotopy of framed well balanced ζ_s- FLIFs $(\widetilde{\Phi}_s, \widetilde{\xi}_s)$ beginning with $(\widetilde{\Phi}_0, \widetilde{\xi}_0) = (\widetilde{\Phi}, \widetilde{\xi})$. Though this is, in general, impossible, the wrinkling embedding theorem from [6] allows us to do that after a certain additional modification of V, called *pleating*, see Lemma 2. We then show that the pleating construction can be extended to the class of framed well balanced FLIFs, see Sect. 3.5. Thus we get a framed well balanced FLIF $(\widehat{\Phi}, \widehat{\xi})$ extending the local framed leafwise Igusa function (φ_A, ξ_A).

The proof now is concluded in two steps. First, we show, see Lemma 20, that a framed well balanced FLIF can be made holonomic near V, and then use the wrinkling theorem from [3] to construct a holonomic extension to the whole manifold W, see Step 4 in Sect. 4.

The paper has the following organization. In Sect. 2.1 we discuss the notion of fold tangency of a submanifold with respect to a not necessarily integrable distribution, define the pleating construction for submanifolds and formulate the main technical result, Lemma 2, which is an analog for folded maps of Gromov's directed embedding theorem, see [8]. This is a corollary of the results of [6]. Section 3 is the main part of the paper. We define and study there the notions and properties of balanced, prepared and well balanced FLIFs, and gradually realize the described above program of making a framed FLIF well balanced. We also prove here Igusa's result about existence of a formal extension for framed FLIFs, see Theorem 2, and local integrability of well balanced FLIFs, see Lemma 20. Finally, in Sect. 4 we just recap the main steps of the proof.

2 Tangency of a Submanifold to a Distribution

In this section we always denote by V an n-dimensional submanifold of an $(n+k)$-dimensional manifold W, by Σ a codimension 1 submanifold of V and by $\mathrm{Norm} = \mathrm{Norm}(V)$ the normal bundle of V.

2.1 Submanifolds Folded with Respect to a Distribution

Let ζ be an n-dimensional distribution, i.e. a subbundle $\zeta \subset TW$. The *non-transversality* condition of V to ζ defines a variety Σ_ζ of the 1-jet space $J^1(V, W)$. We say that V has at a point $p \in V$ a tangency to ζ of *fold* type if

- $\operatorname{Corank} \pi^\zeta|_{T_pV} = 1$;
- $J^1(j) : V \to J^1(V, W)$, where $j : V \hookrightarrow W$ is the inclusion, is transverse to Σ_ζ; we denote $\Sigma := (J^1(j))^{-1}(\Sigma_\zeta)$;
- $\pi^\zeta|_{T_p\Sigma} : T_p\Sigma \to TW_p/\zeta$ is injective.

If ζ is integrable, and hence locally is tangent to an affine foliation defined by the projection $\pi : \mathbb{R}^{n+k} \to \mathbb{R}^k$, these conditions are equivalent to the requirement that the restriction $\pi|_V$ has fold type singularity, and in this case one has a normal form for the fold tangency.

If V has fold type tangency to ζ along Σ then we say that V is *folded* with respect to V along Σ. The fold locus $\Sigma \subset V$ is a codimension one submanifold, and at each point $\sigma \in \Sigma$ the 1-dimensional line field $\lambda = \operatorname{Ker} \pi^\zeta|_{TV} = \zeta|_V \cap TV$ is transverse to Σ.

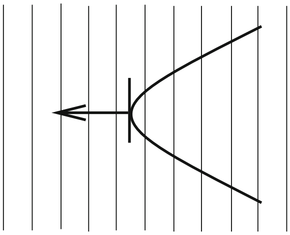


Fig. 3 Characteristic coorientation of the fold

The hyperplane field $T\Sigma \oplus \zeta|_\Sigma$ can be canonically cooriented. In the case when ζ is integrable this coorientation can be defined as follows. The leaves of the foliation trough points of Σ form a hypersurface which divides a sufficiently small tubular neighborhood Ω of Σ in W into two parts, $\Omega = \Omega_+ \cup \Omega_-$, where Ω_- is the part which contains $V \cap \Omega$. Then the *characteristic coorientation* of the fold Σ is the coorientation of the hyperplane $T\Sigma \oplus \zeta|_\Sigma$ determined by the outward normal vector field to Ω_- along Σ, see Fig. 3.

For a general ζ take a point $\sigma_0 \in \Sigma$, a neighborhood U of σ_0 in V, an arbitrary unit vector field $\nu^+ \in (T\Sigma \oplus \zeta|_\Sigma)^\perp$ and consider an embedding $g : U \times (-\epsilon, \epsilon) \to W$ such that $g(x, 0) = x$, $x \in U$ and $\frac{\partial g}{\partial t}(\sigma, 0) = \nu^+, \sigma \in \Sigma \cap U$, where $t \in (-\epsilon, \epsilon)$ is the coordinate corresponding to the second factor. Consider the line field $L = d\,g(\zeta)$. Note that $L|_{(\Sigma\cap U)\times 0} = \lambda$. The line field L integrates to a 1-dimensional foliation on $U \times (-\epsilon, \epsilon)$ which has a tangency of fold type to $U \times 0$ along $\Sigma \times 0$. Hence, $U \times 0 \subset U \times (-\epsilon, \epsilon)$ can be cooriented, as in the integrable case, which gives the required coorientation of $T\Sigma \oplus \zeta|_\Sigma$, see Fig. 3.

It is important to note that *the property that V has a fold type tangency to ζ along Σ depends only on $\zeta|_V$, and not on its extension to $\mathcal{O}p\,V$.* Similarly, the above definition of the characteristic coorientation of $T\Sigma \oplus \zeta|_\Sigma$ is independent of all the choices and depends only on $\zeta|_V$ and not on its extension to $\mathcal{O}p\,V$.

The following simple lemma (which we do not use in the sequel) clarifies the geometric meaning of the fold tangency.

Lemma 1 (Local normal form for fold type tangency to a distribution). *Suppose $V \subset W$ is folded with respect to ζ along V and the fold Σ is cooriented. Denote $\lambda := \zeta|_\Sigma \cap TV|_\Sigma$ and $\eta := (\zeta|_V)/\lambda$. Consider the pull-back $\widetilde{\eta}$ of the bundle η to $\Sigma \times \mathbb{R}^2$ and denote by E the total space of this bundle. Then there exists a neighborhood Ω of $\Sigma \times 0$ in E, a neighborhood $\Omega' \supset \Sigma$ in W, and a diffeomorphism $\Omega \to \Omega'$ introducing coordinates (σ, x, z, y) in Ω', $\sigma \in \Sigma$, $(x, z) \in \mathbb{R}^2$, $y \in \eta$, such that in these coordinates the manifold V is given by the equations $z = x^2, y = 0$ and the bundle $\zeta|_V$ coincides with the restriction to V of the projection $(\sigma, x, z, y) \to (\sigma, z)$.*

Lemma 1 implies, in particular, that if V is folded with respect to ζ then $\zeta|_V$ always admits an integrable extension to a neighborhood of V.

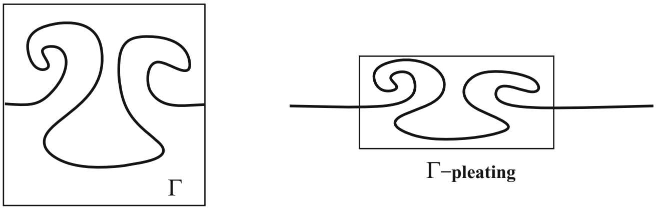


Fig. 4 Γ-pleating

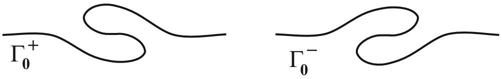


Fig. 5 Curves $\Gamma_0^{\pm}$

2.2 Pleating

Suppose V is folded with respect to ζ along Σ. Let $S \subset V \setminus \Sigma$ be a closed codimension 1 submanifold and $\nu^+ \in \zeta$ be a vector field defined over $\mathcal{O}p\, S \subset W$. For a sufficiently small $\epsilon, \delta > 0$ there exists an embedding $g : S \times [-\delta, \delta] \times [-\epsilon, \epsilon] \to W$ such that

- $\frac{\partial g}{\partial u}(s,t,u) = \nu^+(g(s,t,u)),\ (s,t,u) \in S \times [-\delta,\delta] \times [-\epsilon,\epsilon]$,
- $g|_{S\times 0\times 0}$ is the inclusion $S \hookrightarrow V$,
- $g|_{S\times[-\delta,\delta]\times 0}$ is a diffeomorphism onto the tubular δ-neighborhood $U \supset S$ in V, which sends intervals $s \times [-\delta,\delta] \times 0$, $s \in S$, to geodesics normal to S.

Let $\Gamma \subset P := [-1,1] \times [-1,1]$ be an embedded connected curve which near ∂P coincides with the line $\{u = 0\}$. Here we denote by t, u the coordinates corresponding to the two factors. We assume that Γ is folded with respect to the foliation defined by the projection $(t,u) \mapsto t$ (this is a generic condition). We denote by $\Gamma_{\delta,\epsilon}$ the image of Γ under the scaling $(t,u) \mapsto (\delta t, \epsilon u)$. Consider a manifold $\widetilde{V}$ obtained from V by replacing the neighborhood U by a deformed neighborhood $\widetilde{U}_\Gamma = g(S^{n-1} \times \Gamma_{\delta,\epsilon})$. We say $\widehat{V}$ is the result of *Γ-pleating* of V over S in the direction of the vector field ν^+, see Fig. 4.

The Γ_0^+-pleating with the curve Γ_0^+ shown on Fig. 5 will be referred simply as *pleating*.

Lemma 2 (Pleated isotopy). *Suppose $V \subset (W, \zeta)$ is folded with respect to ζ along $\Sigma \subset V$. Let ζ_s, $s \in [0,1]$, be a family of n-dimensional distributions over a neighborhood $\Omega \supset V$. Then there exist*

– *A manifold $\widetilde{V} \subset \Omega$ obtained from V by a sequence of pleatings over boundaries of small embedded balls in the direction of vector fields which extend to these balls, and*
– *A C^0-small isotopy $h_s : \widetilde{V} \to \Omega$,*

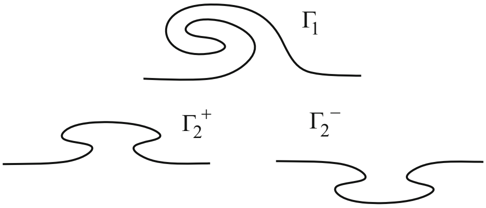


Fig. 6 Curves Γ_1 and $\Gamma_2^\pm$

such that for each $s \in [0,1]$ the manifold $h_s(\widetilde{V})$ has only fold type tangency to ζ_s. If $\widetilde{\Sigma} = \Sigma \cup \Sigma'$ is the fold of $\widetilde{V}$ with respect to ζ_0 then $h_s(\widetilde{\Sigma})$ is the fold of $h_s(\widehat{V})$ with respect to ζ_s. If the homotopy ζ_s is fixed over a neighborhood OpA of a closed subset $A \subset V$ then one can arrange that $V \cap \mathcal{O}p\,A = \widetilde{V} \cap \mathcal{O}p\,A$ and the isotopy h_s is fixed over $\mathcal{O}p\,A$.

Lemma 2 is a version of the wrinkled embedding theorem from [6], see Theorem 3.2 in [6] and the discussion in Sects. 3.2 and 3.3 in that paper on how to replace the wrinkles by spherical double folds and how to generalize Theorem 3.2 to the case of not necessarily integrable distributions. Another cosmetic difference between the formulations in [6] and Lemma 2 is that the former one allows not only double folds, but also their embryos, i.e. the moments of death-birth of double folds. This can be remedied by preserving the double folds till the end in the near-embryo state, rather than killing them, and similarly by creating the necessary number of folds by pleating at the necessary places before the deformation begins.

Remark 1. If $\widetilde{V}$ satisfies the conclusion of Lemma 2 then any manifold $\widetilde{\widetilde{V}}$ obtained from $\widetilde{V}$ by an additional Γ-pleating with *any* Γ will also have this property. For our purposes we will need to pleat with three special curves Γ_1 and $\Gamma_2^\pm$ shown on Fig. 6.

As it clear from this picture, a pleating with any of these curves can be viewed as a result of a Γ_0^+-pleating followed by a second Γ_0^--pleating. Hence, in the formulation of Lemma 2 one can pleat with any of the curves Γ_1 and $\Gamma_2^\pm$ instead of Γ_0^+.

3 Geometry of FLIFs

3.1 Homomorphisms Γ_Φ and Π_Φ

Given a ζ-FLIF Φ we will associate with it several objects and constructions.

Isomorphism Γ_Φ : Norm $\to$ Vert. This isomorphism is determined by Φ^1. The tangent bundle $T(\zeta|_V)$ to the total space of the bundle $\zeta|_V$ canonically splits as Vert $\oplus\, TW|_V$, and hence the bundle of tangent planes to the section Φ^1 along its 0-set V can be viewed as a graph of a homomorphism $\widehat{\Gamma}_\Phi{:}TW|_V \to$ Vert vanishing

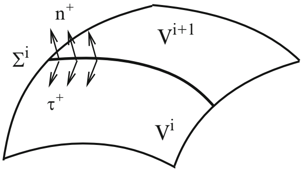


Fig. 7 The vector fields n^+ and τ^+

on TV. The restriction of this homomorphism to Norm will be denoted by Γ_Φ. The transversality of the section Φ^1 to the 0-section ensures that Ker $\widehat{\Gamma}_\Phi = TV$ and hence Γ_Φ is an isomorphism.

By an *index* coorientation of Σ in V we will mean its coorientation by a normal vector field τ^+ pointing in the direction of **decreasing** of the index, i.e. on Σ^i it points into V^i. We will denote by n^+ the vector field $\Gamma_\Phi^{-1}(\lambda^+) \in \mathrm{Norm}(V)$, see Fig. 7.

In the holonomic situation the index coorientation is given by the vector field λ^+ and the vector field n^+ determines the characteristic coorientation of the fold.

We call a ζ-FLIF Φ *prepared* if

- $V(\Phi)$ is folded with respect to ζ with the fold along $\Sigma(\Phi)$;
- $TV \cap \mathrm{Vert}_\Sigma = \lambda$ and the vector field λ^+ determines the *index* coorientation of the fold;
- The vector field n^+ determines the *characteristic* coorientation of the fold Σ.

Thus, any *holonomic* FLIF (when, in particular, ζ is integrable) is prepared.

Isomorphism Π_Φ : Norm → Vert. Given a *prepared* ζ-FLIF Φ, let us denote by K the restriction of the orthogonal projection $TW|_V \to \mathrm{Norm}$ to the subbundle $\mathrm{Vert} = \zeta|_V \subset TW|_V$. The homomorphism K is non-degenerate over $V \setminus \Sigma$ and has a 1-dimensional kernel λ over Σ.

Lemma 3 (Definition of Π_Φ). *The composition $\Phi^2 \circ K^{-1} : \mathrm{Norm}|_{V\setminus\Sigma} \to \mathrm{Vert}|_{V\setminus\Sigma}$ continuously extends to a non-degenerate homomorphism $\Pi_\Phi : \mathrm{Norm} \to \mathrm{Vert}$.*

Proof. Let us prove the extendability of the inverse operator $K \circ \left(\Phi^2\right)^{-1}$. There exists a canonical extension of the vector field λ^+ as a unit Φ^2-eigenvector field $\widetilde{\lambda}^+$ on $\mathcal{O}p\,\Sigma \subset V$. Then

$$\Phi^2(\widetilde{\lambda}^+(v)) = c(v)\widetilde{\lambda}^+(v)\,,\ v \in \mathcal{O}p\,\Sigma\,,$$

where the eigenvalue function $c : \mathcal{O}p\,\Sigma \to \mathbb{R}$ has Σ as its regular 0-level. Denote $\widetilde{\mathrm{Ver}} = \widetilde{\lambda}^\perp(v)$ the orthogonal eigenspace of $\Phi^2(v)$. Denote $\widetilde{\mathrm{Nor}} := K(\widetilde{\mathrm{Ver}})$. The operator $K \circ \left(\Phi^2\right)^{-1}$ is well defined on $\mathrm{Ver} = \widetilde{\mathrm{Ver}}|_\Sigma \subset \mathrm{Vert}|_\Sigma$ and maps it isomorphically onto $\mathrm{Nor} = \widetilde{\mathrm{Nor}}|_\Sigma$. It remains to prove existence of a non-zero limit

$$\lim_{v\to v_0\in\Sigma} K\left(\left(\Phi^2\right)^{-1}(\widetilde{\lambda}^+)\right) = \lim_{v\to v_0\in\Sigma} \frac{1}{c(v)} K(\widetilde{\lambda}^+(v))\,.$$

The vector-valued function $K(\widetilde{\lambda}^+(v))$ vanishes on Σ while the function $c(v)$ has no critical points on Σ. Hence, the above limit exists. On the other hand, the transversality condition for the fold implies that $||K(\widetilde{\lambda}^+(v))|| \geq a\,\mathrm{dist}(v,\Sigma)$, while $|c(v)| \leq b\,\mathrm{dist}(v,\Sigma)$ for some positive constants $a, b > 0$, and therefore $\lim\limits_{v\to v_0\in\Sigma} \frac{1}{c(v)} K(\widetilde{\lambda}^+(v)) \neq 0$. □

Lemma 4. *If Φ is holonomic then $\Pi_\Phi = \Gamma_\Phi$.*

Proof. Indeed, recall that $\Gamma_\Phi = \widehat{\Gamma}_\Phi|_{\mathrm{Norm}}$, where $\widehat{\Gamma}_\Phi : TW|_V \to \mathrm{Vert}$ is the homomorphism defined by the section Φ^1 linearized along its zero-set V. In the holonomic situation one has over $V \setminus \Sigma$ the equality

$$\widehat{\Gamma}_\Phi|_{\mathrm{Vert}} = d^2\varphi = \Phi^2\,,$$

where $\varphi = \Phi^0$. But $\widehat{\Gamma}_\Phi|_{\mathrm{Vert}}$ and $\widehat{\Gamma}_\Phi|_{\mathrm{Norm}}$ are related by a projection along the kernel of $\widehat{\Gamma}_\Phi$ which is equal to TV. Hence, $\Gamma_\Phi = \Phi^2 \circ K^{-1} = \Pi_\Phi$. By continuity, the equality $\Pi_\Phi = \Gamma_\Phi$ holds everywhere. □

3.2 Twisted Normal Bundle and the Isomorphism Δ_Φ

Given any ζ-FLIF Φ we define here a *twisted normal bundle*, or as we also call it *virtual vertical bundle* ${}^\Phi\mathrm{Vert} \subset TW_V$ over V. As we will see later (see Lemma 7), in the holonomic case ${}^\Phi\mathrm{Vert}$ coincides with Vert.

Let $U = \Sigma \times [-\epsilon, \epsilon]$ be the tubular neighborhood of Σ in V of radius $\epsilon > 0$. We assume that the splitting is chosen in such a way that the vector field $\frac{\partial}{\partial t}$, where t is the coordinate corresponding to the second factor, defines the index coorientation of Σ in V, and hence coincides with τ^+. We denote

$$U_+ := \Sigma \times (0,\epsilon],\;\; U_- := \Sigma \times [-\epsilon, 0)\,.$$

Denote by $\widetilde{\lambda}^+ \in \mathrm{Vert}$ the unit eigenvector field of $\Phi^2|_U$ which extends $\lambda^+ \in \mathrm{Vert}|_\Sigma$. If ϵ is small enough then such extension is uniquely defined. Let $\widetilde{\mathrm{Ver}} := \widetilde{\lambda}^\perp$ be the complementary Φ^2-eigenspace. We have $\widetilde{\mathrm{Ver}}|_\Sigma = \mathrm{Ver}$. Denote $\widetilde{n}^+ := \Gamma_\Phi^{-1}(\widetilde{\lambda}^+)$, $\widetilde{\mathrm{Nor}} = \Gamma_\Phi^{-1}(\widetilde{\mathrm{Ver}})$, $\widetilde{\tau}^+ := \frac{\partial}{\partial t}$. Choose a function $\theta : U \to [-\frac{\pi}{2}, \frac{\pi}{2}]$ which has Σ as its regular level set $\{\theta = 0\}$, and which is equal to $\pm\frac{\pi}{2}$ near $\Sigma \times (\pm\epsilon)$.

We define the bundle ${}^\Phi\mathrm{Vert}$ in the following way. Over $V \setminus U$ it is equal to Norm. The fiber over a point $v \in U$ is equal to $\mathrm{Span}(\widetilde{\mathrm{Nor}}, \mu(v))$, where the line $\mu(v)$ is generated by the vector

$$\mu^+(v) = \sin\theta(v)\widetilde{n}^+(v) + \cos\theta(v)\widetilde{\tau}^+(v),$$

see Fig. 8.

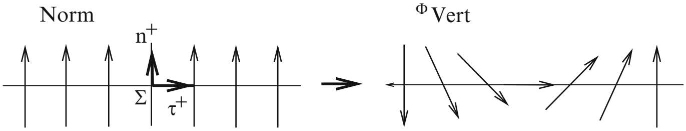


Fig. 8 Twisting the normal bundle

Isomorphism Δ_Φ : Vert → ${}^{\Phi}$Vert. Let $c : U \to \mathbb{R}$ be the eigenvalue function corresponding to the Φ^2-eigenvector field $\widetilde{\lambda}^+ \in$ Vert on U, i.e. we have $\Phi^2(\widetilde{\lambda}^+(v)) = c(v)\widetilde{\lambda}^+(v)$, $v \in U$. The function c is positive on U_+ and negative on U_-. Let $\widetilde{c} : U \to \mathbb{R}$ be any *positive* function which is equal to c on $\partial U_+ = \Sigma \times \epsilon$ and equal to $-c$ on $\partial U_- = \Sigma \times (-\epsilon)$.

We then define the operator

$$\Delta_\Phi : \mathrm{Vert} \to {}^{\Phi}\mathrm{Vert}$$

by the formula

$$\Delta_\Phi(Z) = \begin{cases} \Gamma_\Phi^{-1}(\Phi^2(Z)), & \text{over } V \setminus U,\ Z \in \mathrm{Vert}\,; \\ \Gamma_\Phi^{-1}(\Phi^2(Z)), & \text{over } U,\ Z \in \widetilde{\mathrm{Ver}}\,; \\ \widetilde{c}(v)\left(\sin\theta(v)\widetilde{n}^+ + \cos\theta(v)\widetilde{\tau}^+\right), & Z = \widetilde{\lambda}^+(v),\ v \in U\,. \end{cases} \tag{1}$$

It will be convenient for us to keep some ambiguity in the definition of ${}^{\Phi}$Vert and Δ_Φ. However, we note that the space of choices we made in the definition is contractible, and hence the objects are defined in a homotopically canonical way. Let us extend ${}^{\Phi}$Vert and Δ_Φ to a neighborhood $\mathcal{O}p\,V \subset W$. We will keep the same notation for the extended objects.

Lemma 5. *For any ζ-FLIF Φ the* ${}^{\Phi}$Vert-*FLIF*

$$\Phi^{\mathrm{Norm}} = (\mathrm{Id}, \Delta_\Phi)_*\Phi$$

on $\mathcal{O}p\,V$ is prepared.

Proof. Denote $\widehat{\Phi} := \Phi^{\mathrm{Norm}}$. We have $V(\widehat{\Phi}) = V(\Phi) = V$. First of all we observe (see Fig. 9) that V is folded with respect to ${}^{\Phi}$Vert along Σ and the vector field $n^+ = n^+(\Phi)$ defines the characteristic coorientation of the fold. On the other hand, $\lambda^+(\widehat{\Phi}) = \Delta_\Phi(\lambda_+(\Phi)) = \tau^+(\Phi) = \tau^+(\widehat{\Phi})$ and

$$n^+(\widehat{\Phi}) = \Gamma_{\widehat{\Phi}}^{-1}(\lambda^+(\widehat{\Phi})) = \Gamma_{\widehat{\Phi}}^{-1}(\tau^+) = \Gamma_\Phi^{-1}(\Delta_\Phi^{-1}(\tau_+)) = \Gamma_\Phi^{-1}(\lambda^+) = n^+(\Phi),$$

and hence $n^+(\widehat{\Phi})$ defines the characteristic coorientation of the fold Σ. Thus $\widehat{\Phi}$ is prepared. □

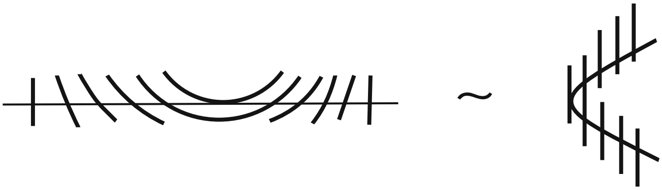


Fig. 9 V is folded with respect to ${}^{\Phi}\mathrm{Vert}$.

Lemma 6. *For any FLIF Φ the diagram*

$$\begin{tikzcd}
\mathrm{Norm} \arrow[rr, "\Gamma_\Phi"] \arrow[dr, "\Pi_{\widehat{\Phi}}"'] & & \mathrm{Vert} \arrow[dl, "\Delta_\Phi"] \\
& {}^{\Phi}\mathrm{Vert} &
\end{tikzcd}$$

commutes for appropriate choices in the definition of ${}^{\Phi}\mathrm{Vert}$ *and* Γ_Φ.

Proof. We need to check that $\Pi_{\widehat{\Phi}} = \Delta_\Phi \circ \Gamma_\Phi$. First, we check the equality over $V \setminus U$. We have $\mathrm{Norm}|_{V\setminus U} = {}^{\Phi}\mathrm{Vert}|_{V\setminus U}$, and hence $K_{\widehat{\Phi}} = \mathrm{Id}$. Furthermore, over $V \setminus U$ we have $\Pi_{\widehat{\Phi}} = K_{\widehat{\Phi}}^{-1} \circ \widehat{\Phi}^2 = \Delta_\Phi \circ \Phi^2 \circ \Delta_\Phi^{-1} = \Gamma_\Phi^{-1} \circ \Phi^2 \circ \Phi^2 \circ \left(\Phi^2\right)^{-1} \Gamma_\Phi = \Gamma_\Phi^{-1} \circ \Phi^2 \circ \Gamma_\Phi = \Delta_\Phi \circ \Gamma_\Phi$. Similarly, we check that $\Pi_{\widehat{\Phi}}|_{\widetilde{\mathrm{Nor}}} = \Delta_\Phi \circ \Gamma_\Phi|_{\widetilde{\mathrm{Nor}}}$. Finally, evaluating both parts of the equality on the vector field n_+ we get: $\Pi_{\widehat{\Phi}}(n^+) = \lambda^+ = \Delta_\Phi(\Gamma_\Phi(n^+))$. Then this implies $\Pi_{\widehat{\Phi}}(\widetilde{n}^+) = \Delta_\Phi(\Gamma_\Phi(\widetilde{n}^+))$ for an appropriate choice of the function $\widetilde{c} > 0$ in the definition of the homomorphism Δ_Φ. □

3.3 The Holonomic Case

We will need the following normal form for a leaf-wise Igusa function φ near Σ (see [1, 2]).

Consider the pull-back of the bundle $\mathrm{Ver} = \mathrm{Ver}_+ \oplus \mathrm{Ver}_-$ defined over Σ to $\Sigma \times \mathbb{R} \times \mathbb{R}$ via the projection $\Sigma \times \mathbb{R} \times \mathbb{R} \to \Sigma$. Let E be the total space of this bundle. The submanifold $\Sigma \times 0 \times 0$ of the 0-section of this bundle we will denote simply by Σ. Consider a function $\theta : E \to \mathbb{R}$ given by the formula

$$\theta(\sigma, x, z, y_+, y_-) = x^3 - 3zx + \frac{1}{2}(||y_+^2|| - ||y_-^2||)\,; \tag{2}$$

$(\sigma, x, z) \in \Sigma \times \mathbb{R} \times \mathbb{R},\ y_\pm \in (\mathrm{Ver}_\pm)_\sigma$.

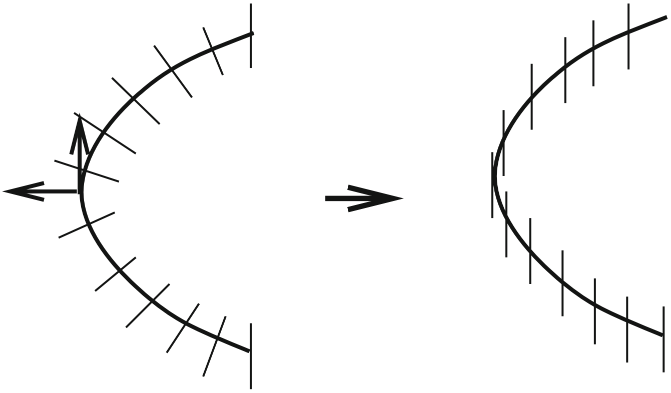


Fig. 10 Holonomic case: the bundle ${}^{\Phi}$Vert coincides with Vert

Consider the projection $p : E \to \Sigma \times \mathbb{R}$ defined by the formula

$$p(\sigma, x, z, y_+, y_-) = (\sigma, z).$$

There exists an embedding $g : \mathcal{O}p\,\Sigma \to W$, where $\mathcal{O}p\,\Sigma$ is a neighborhood of Σ in E, such that

- $g(\sigma) = \sigma, \sigma \in \Sigma$;
- g maps the fibers of the projection p to the leaves of the foliation $\mathcal{F}$.
- $\varphi \circ g = \theta$.

Via the parameterization map g we will view (σ, x, z, y_+, y_-) as coordinates in $\mathcal{O}p\,\Sigma \subset W$. In these coordinates the function φ has the form (2), the manifold V is given by the equations $z = x^2$, $y_\pm = 0$, the foliation $\mathcal{F}$ is given by the fibers of the projection p, the vector field $-\frac{\partial}{\partial z}$ defines the characteristic coorientation of the fold Σ, and the vector field $\frac{\partial}{\partial x} \in TV|_\Sigma$ defines the index coorientation.

The normal form (2) can be extended to a neighborhood of V using the parametric Morse lemma. However, we will not need it for our purposes.

Lemma 7. *If Φ is holonomic then for appropriate auxilliary choices the virtual vertical bundle* ${}^{\Phi}$Vert *coincides with* Vert *(Fig. 10) and the isomorphism*

$$\Delta_\Phi : \mathrm{Vert} \to {}^{\Phi}\mathrm{Vert} = \mathrm{Vert}$$

is the identity.

Proof. Let Φ be holonomic and $\Phi^0 = \varphi$. The bundle Vert is transverse to V over $V \setminus U$, and over U it splits as $\widetilde{\mathrm{Ver}} \oplus \widetilde{\lambda}$. We have $\widetilde{\mathrm{Nor}} \cap TV = \{0\}$, the bundle $\widetilde{\lambda}$ is tangent to V along Σ and $\lambda = \widetilde{\lambda}|_\Sigma$ is transverse to Σ . Let us choose a metric such that the transversality condition for the bundles $\mathrm{Vert}|_{V \setminus U}$, $\widetilde{\mathrm{Ver}}|_U$, $\lambda|_U$ are replaced by the orthogonality one. Then the operator Γ_Φ^{-1}, and hence Δ_Φ leaves invariant the bundles $\mathrm{Vert}|_{V\setminus U}$ and $\widetilde{\mathrm{Ver}}|_U$. Moreover, on both these bundles the operators $\Phi^2 = d^2\varphi$ and Γ_Φ coincide, and hence $\Delta_\Phi = \mathrm{Id}$.

It remains to analyze $\Delta_\Phi|_{\widetilde{\lambda}^+}$. By definition,

$$\Delta_\Phi(\widetilde{\lambda}^+(v)) = \widetilde{c}(v)\left(\cos\theta(v)\tau^+(v) + \sin\theta(v)\widetilde{n}^+(v)\right), \; v \in U,$$

where $\widetilde{n}^+ = \Gamma_\Phi^{-1}(\widetilde{\lambda}^+)$. It is sufficient to ensure that the line $\Delta_\Phi|_{\widetilde{\lambda}}$ coincides with $\widetilde{\lambda}^+$ because then the similar equality for vectors could be achieved just by choosing an appropriate amplitude function $\widetilde{c}$ in the definition of the operator Δ_Φ. Note that we have $\Delta_\Phi(\widetilde{\lambda}(v)) = \lambda(v)$ for $v \in \partial U$ or $v \in \Sigma$. To ensure this equality on the rest of U we need to further specify our choices. As it was explained above in Sect. 3.3 we can assume that the function φ in a neighborhood $\Omega \supset U$ in W is given by the normal form (2). Choosing $\Omega = \{|x|, |z| \leq \epsilon\}$ we have

$$U := V \cap \Omega = \{z = x^2, y_\pm = 0, |x| \leq \epsilon\},$$

and bundles Vert, $\widetilde{\mathrm{Ver}}$ and $\widetilde{\lambda}$ are given, respectively, by restriction to V of the projections $(\sigma, x, z, y_+, y_-) \mapsto (\sigma, z)$, $(\sigma, x, z, y_+, y_-) \mapsto (\sigma, x, z)$ and $(\sigma, x, z, y_+, y_-) \mapsto (\sigma, z, y_-, y_+)$. Let us choose the tangent to V vector field $\frac{\partial}{\partial x} + 2z\frac{\partial}{\partial z}$ as $\widetilde{\tau}^+$ and recall that we have chosen a metric for which the vectors $\tau^+(v)$ and $\widetilde{\lambda}^+(v)$ for $v \in \partial U$ are orthogonal. Let us choose any vector field $\widehat{\nu} \in P := \mathrm{Span}(\frac{\partial}{\partial x}, \frac{\partial}{\partial z})$ such that

- $\widehat{\nu}^+|_{\partial U_+} = \widetilde{\lambda}^+|_{\partial U_+}$;
- $\widehat{\nu}^+|_{\partial U_-} = -\widetilde{\lambda}^+|_{\partial U_-}$;
- $\widehat{\nu}^+|_\Sigma = -\frac{\partial}{\partial z}$ defines the characteristic coorientation;
- The vector field $\lambda^+|_{\mathrm{Int}\,U+}$ belongs to the positive cone generated by $\widetilde{\tau}^+$ and $\widehat{\nu}^+$;
- The vector field $\lambda^+|_{\mathrm{Int}\,U-}$ belongs to the positive cone generated by $\widetilde{\tau}^+$ and $-\widehat{\nu}^+$.

Let us pick a metric on P for which the vector fields $\widetilde{\tau}^+$ and $\widehat{\nu}^+$ are orthogonal and the vector fields $\widetilde{\tau}^+$ and $\widetilde{\lambda}^+$ have length 1. By rescaling, if necessary, the vector field $\widehat{\nu}^+$ we can arrange that it has length 1 as well. Let us denote by $\theta(v)$ the angle between the vectors τ^+ and λ^+ in this metric. If we construct the virtual vertical bundle (Fig. 10) ${}^\Phi|$ with this choice of the metric and the angle function θ, then the condition $\Delta_\Phi(\widetilde{\lambda}) = \widetilde{\lambda}$ will be satisfied. □

In all our results below concerning an extension of a holonomic FLIF from a neighborhood of a closed set A we will always assume that over $\mathcal{O}p\,A$ all the necessary special choices are made to ensure the conclusion of Lemma 7: the virtual vertical bundle ${}^\Phi\mathrm{Vert}$ coincides with Vert and the isomorphism $\Delta_\Phi : \mathrm{Vert} \to {}^\Phi\mathrm{Vert} = \mathrm{Vert}$ is the identity, and hence, according to Lemma 6, we have $\Gamma_\Phi = \Pi_{\widehat{\Phi}}$, where $\widehat{\Phi} = \Phi^{\mathrm{Norm}}$.

3.4 Balanced and Well Balanced FLIFs

We call a FLIF Φ *balanced* if the compositions

$$\mathrm{Norm} \xrightarrow{\Gamma_\Phi} \mathrm{Vert} \hookrightarrow TW|_V \quad \text{and} \quad \mathrm{Norm} \xrightarrow{\Pi_{\widehat{\Phi}}} {}^{\Phi}\mathrm{Vert} \hookrightarrow TW|_V$$

are homotopic in the space of injective homomorphisms Vert → $TW|_V$. Here we denote by $\widehat{\Phi}$ the FLIF Φ^{Norm}. If Φ is holonomic over $\mathcal{O}p\,A \subset W$ then we say that Φ is *balanced relative* A if the homotopy can be made fixed over A.

Lemma 6 shows that the balancing condition is equivalent to the requirement that the composition $\mathrm{Vert} \xrightarrow{\Delta_\Phi} {}^{\Phi}\mathrm{Vert} \hookrightarrow TW|_V$ is homotopic to the inclusion $\mathrm{Vert} \hookrightarrow TW|_V$ in the space of injective homomorphisms Vert → $TW|_V$.

Lemma 4 shows that a holonomic Φ is balanced. Moreover, it is balanced relative to any closed subset $A \subset W$.

We say that a FLIF Φ is *well balanced* if it is prepared and the isomorphisms Π_Φ, Γ_Φ : Norm → Vert are homotopic as isomorphisms. Similarly we define the notion of a FLIF well balanced relative to a closed subset A.

It is not immediately clear from the definition that a well balanced FLIF is balanced. The next lemma shows that this is still the case.

Lemma 8. *A well balanced FLIF is balanced.*

Proof. We need to check that over $V \setminus \Sigma$ we have $\Pi_\Phi = \Phi^2 \circ K^{-1}$ and $\Pi_{\widehat{\Phi}} = \Phi^2 \circ \widehat{K}^{-1}$, where K is the projection Norm → Vert and $\widehat{K}$ is the projection Norm → ${}^{\Phi}$Vert. We have $\widehat{K} = T \circ K$, where T : Vert → ${}^{\Phi}$Vert is the projection along TV. Hence, we have $\Pi_{\widehat{\Phi}} = \Pi_\Phi \circ T$ which implies, in particular, that the projection T is non-degenerate over the whole V. Hence, the composition of the projection operator T with the inclusion ${}^{\Phi}\mathrm{Vert} \overset{i}{\hookrightarrow} TW|_V$ is homotopic to the inclusion $\mathrm{Vert} \overset{j}{\hookrightarrow} TW|_V$ as injective homomorphisms, and so do the compositions $i \circ \Pi_{\widehat{\Phi}}$ and $j \circ \Pi_\Phi$. □

Note that for the codimension 1 case, i.e. when $n = 1$ the well balanced condition for a prepared FLIF is very simple:

Lemma 9 (Well-balancing criterion in codimension 1). *Suppose* $\dim \zeta = 1$. *Then any prepared ζ-FLIF Φ is well balanced if and only if at one point $v \in V \setminus \Sigma$ of every connected component of V the map*

$$(\Pi_\Phi)_v \circ (\Gamma_\Phi)_v^{-1} : \mathrm{Vert}_v \to \mathrm{Vert}_v$$

is a multiplication by a positive number. The same statement holds also in the relative case.

Lemma 10 (Well balanced FLIFs and folded isotopy). *Let Φ be a well balanced FLIF. Let $h_s : W \to W$ be a diffeotopy, ζ_s a family of n-dimensional distributions on W, and $\Theta_s : \zeta_0 \to \zeta_s$ a family of bundle isomorphisms covering h_s, $s \in [0,1]$, such that $h_0 = \mathrm{Id}$ and for each $s \in [0,1]$*

- *Submanifold $V_s := h_s(V) \subset W$ is folded with respect to ζ_s along $\Sigma_s := h_s(\Sigma)$;*
- *$dh_s(\zeta_0 \cap TV)) = dh_s(\zeta_s) \cap TV_s$;*
- *$dh_s|_{\zeta_0 \cap TV} = \Theta_s|_{\zeta_0 \cap TV}$.*

Then the push-forward ζ_s-FLIF $\Phi_s := (h_s, \Theta_s)_* \Phi$, $s \in [0,1]$, is well balanced.

Proof. By assumption $V(\Phi_s)$ is folded with respect to ζ_s. Next, we observe that all co-orientations cannot change in the process of a continuous deformation, and similarly, the isomorphisms Π_{Φ_s} and Γ_{Φ_s} vary continuously, and hence remain homotopic as bundle isomorphisms $\mathrm{Norm}(\Phi_s) \to \mathrm{Vert}(\Phi_s)$. Thus the well balancing condition is preserved. □

Note that if Φ is balanced then the homomorphism $\Delta_\Phi : \zeta|_V \to {}^{\Phi}\mathrm{Vert}$ composed with the inclusion ${}^{\Phi}\mathrm{Vert} \hookrightarrow TW$ extends to an injective homomorphism $F : \zeta \to TW$. Then $(\mathrm{Id}, F)_* \Phi$ is a ν-FLIF extending the local ν-FLIF $\widehat{\Phi}$. Here we denoted by $\nu := F(\zeta)$.

Lemma 11. *The ν-FLIF $\widehat{\Phi} = \Phi^{\mathrm{Norm}}$ on $\mathcal{O}p\, V$ is well balanced.*

Proof. We already proved in Lemma 5 that $\widehat{\Phi}$ is prepared. Let us show that $\Pi_{\widehat{\Phi}} = \Gamma_{\widehat{\Phi}}$. According to the definition of the push-forward operator we have $\Gamma_{\widehat{\Phi}} = \Delta_\Phi \circ \Gamma_\Phi$. But according to Lemma 6 we have $\Delta_\Phi \circ \Gamma_\Phi = \Pi_{\widehat{\Phi}}$. □

Consider a ζ-FLIF Φ. Suppose there exists a $(k+1)$-dimensional submanifold $Y \subset W$, $Y \supset V$, such that

- Y is transverse to ζ;
- The line field $\mu|_V \subset \mathrm{Vert}$ is an eigenspace field for Φ^2, where we denoted $\mu := \zeta \cap TY$;
- $\Phi^2|_{N := \mu^\perp|_V}$ is non-degenerate, where $\mu^\perp$ is the orthogonal complement to μ in $\zeta|_Y$.

Consider the restriction μ-FLIF $\widetilde{\Phi} = \Phi|_Y$ defined as follows: $\widetilde{\Phi}^0 = \Phi^0|_Y$, $\widetilde{\Phi}^1$ is the projection of Φ^1 along $\mu^\perp$, $\widetilde{\Phi}^2 = \Phi^2|_\mu$, $\widetilde{\lambda} = \lambda$. Note that we have $V(\widetilde{\Phi}) = V$ and $\Sigma(\widetilde{\Phi}) = \Sigma$.

We will assume that the bundle N is orthogonal to TY. Under this assumption we have $\Gamma_\Phi(N) = N$. The next criterion for a FLIF to be well-balanced is immediate from the definition.

Lemma 12. *If $\widetilde{\Phi}$ is prepared then so is Φ. If $\widetilde{\Phi}$ is well balanced and $\Phi^2|_N = \Gamma_\Phi|_N$ then Φ is well balanced as well.*

3.5 Pleating a FLIF

We adjust in this section the pleating construction defined in Sect. 2.2 for submanifolds to make it applicable for framed well balanced FLIFs. et Φ be a well balanced ζ-FLIF. We will use here the following notation from Sect. 2.2:

- $S \subset V_i \subset V \setminus \Sigma, i = 0,\dots,n$, is a closed cooriented codimension 1 submanifold;
- $U = S \times [-\delta,\delta] \supset S = S \times 0$ is a tubular δ-neighborhood of S in V_i;
- $\nu^+ \in \zeta$ is a unit vector field defined over a neighborhood Ω of U in W;
- $g : S \times [-\delta,\delta] \times [-\epsilon,\epsilon] \to \Omega \hookrightarrow W$ is an embedding such that $\frac{\partial g}{\partial u}(s,t,u) = \nu^+(g(s,t,u)),\ (s,t,u) \in S \times [-\delta,\delta] \times [-\epsilon,\epsilon]$, which maps $S \times 0 \times 0$ onto S and $S \times [-\delta,\delta] \times 0$ onto U;
- $\Gamma \subset P := [-1,1] \times [-1,1]$ is an embedded connected curve which near ∂P coincides with the line $\{u = 0\}$;
- $\widetilde{V} \subset W$ is the result of Γ-pleating of V over S in the direction of the vector field ν_+.

We will make the following additional assumptions:

- The splitting $\mathrm{Vert}|_S = \mathrm{Vert}_+|_S \oplus \mathrm{Vert}_-|_S$ is extended to a splitting $\zeta = \zeta_+ \oplus \zeta_-$ over the neighborhood $\Omega \subset W$;
- The vector field ν^+ is a section of either $\zeta_-|_\Omega$ or $\zeta_+|_\Omega$;
- The vector field $\nu_+|_U$ is an eigenvector field for Φ^2;
- $\mathrm{Norm}(\Phi)|_U = \mathrm{Vert}(\Phi)|_U$ and $\Delta_\Phi|_{\mathrm{Vert}|_U} = \mathrm{Id}$.

There exists a diffeotopy $h_s : W \to W$ supported in Ω connecting Id with a diffeomorphism h such that $h(V) = \widetilde{V}$. We denote $\widetilde{U} = \widetilde{U}_\Gamma := h_1(\Gamma)$. Let $\Psi_s : \zeta \to \zeta$, $s \in [0,1]$, be a family of isomorphisms covering h_s which preserve $\mathrm{Vert}_\pm$ and ν^+.

The manifold $\widetilde{U}$ is folded with respect to ζ with the fold $\widetilde{S} = \bigcup\limits_1^{2N} \widetilde{S}_j$ where $\widetilde{S}_j = h_1(S_j)$, where $S_j = S \times t_j$, $-\delta < t_1 < \dots t_{2N} < \delta$. Over $\widetilde{S}$ we have $\widetilde{\tau} = \nu = T\widehat{V} \cap \zeta$.

Consider the push-forward FLIF $\overline{\Phi} := (h_1,\Psi_1)_*\Phi$. Though the manifold $V(\overline{\Phi}) = \widetilde{V}$ is folded with respect to ζ, it is not prepared. We will modify $\overline{\Phi}$ to a prepared FLIF $\widetilde{\Phi} = \mathrm{Pleat}_{S,\nu^+,\Gamma}(\Phi)$ as follows.

Let $\widetilde{c} : \widetilde{U} \to \mathbb{R}$ be a function which on $\partial\widetilde{U} = \partial U$ coincides with the eigenvalue function of the operator Φ^2 for the eigenvector field ν^+, and have the fold $\widetilde{S} := \bigcup\limits_1^{2N} \widetilde{S}_j$ as its regular 0-level. We call component of $\widetilde{U} \setminus \widetilde{S}$ *positive* or *negative* depending on the sign of the function $\widetilde{c}$ on this component. We then define

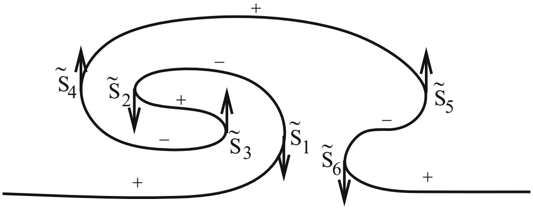


Fig. 11 Γ-pleating of a well balanced FLIF

- $\widetilde{\Phi}^1 = \overline{\Phi}^1$;
- $\widetilde{\Phi}^2|_{\nu^\perp} = \overline{\Phi}^2|_{\nu^\perp}$;
- $\widetilde{\Phi}^2(\nu^+) = \widetilde{c}\,\nu^+$;
- $\lambda^+(\widetilde{\Phi}^2) = \pm\nu^+$, where the sign is chosen in such way that the vector field $\lambda^+(\widetilde{\Phi}^2)$ define an inward coorientation of positive components of $\widetilde{U} \setminus \widetilde{S}$, see Fig. 11.

We say that $\widetilde{\Phi} = \mathrm{Pleat}_{S,\nu^+,\Gamma}(\Phi)$ is obtained from Φ by *Γ-pleating over S in the direction of the vector field ν^+* see Fig. 11.

Lemma 13. *The FLIF $\widetilde{\Phi}$ is well balanced.*

Proof. Consider the $(k+1)$-dimensional manifold

$$Y := g(S \times [-\delta, \delta] \times [-\epsilon, \epsilon]) \subset W\,.$$

Then Y is transverse ζ and $\zeta \cap TY = \nu$. We also note that the orthogonal complement $\nu^\perp$ of $\nu \in \zeta$ is orthogonal to TY, $\widetilde{\Phi}^2|_{\nu^\perp} = \Pi_{\widetilde{\Phi}}|_{\nu^\perp} = \Gamma_{\widetilde{\Phi}}|_{\nu^\perp}$. According to Lemma 12 it is sufficient to check that the restriction $\widehat{\Phi} := \widetilde{\Phi}|_Y$ is well balanced, rel. ∂Y. First, we need to check that this restriction is prepared. By construction, $\widehat{V} = V(\widehat{\Phi})$ is folded with respect to ν and the vector field $\lambda^+(\widehat{\Phi}) = \lambda^+(\widetilde{\Phi})$ defines the index coorientation of $\widetilde{S}$ in $\widetilde{V}$. Next, we need to check that the vector field $n^+(\widehat{\Phi}) = n^+(\widetilde{\Phi}) = \Gamma^{-1}_{\widetilde{\Phi}}(\lambda^+(\widetilde{\Phi})$ defines the characteristic coorientation of the fold. It is sufficient to consider the case when S is the point, and hence $\dim Y = 2$. The general picture is then obtained by taking a direct product with S.

Note that the characteristic co-orientation of the fold $\widetilde{S}_j$ is given by the vector field $\frac{\partial}{\partial t}$ if j is odd, and by $-\frac{\partial}{\partial t}$ if j is even. Consider first the case when j is odd, see Fig. 12. Then if the lower branch of the parabola is positive then the vector field $\Gamma_{\widetilde{\Phi}}(\nu^+)$ defines the same coorientation as the vector field $-\frac{\partial}{\partial t}$. But in this case $\lambda^+ = -\nu^+$, and hence $\Gamma_{\widetilde{\Phi}}(\lambda^+)$ defines the characteristic coorientation of the fold. The other cases can be considered in a similar way. Finally, we use Lemma 9 to conclude that $\widehat{\Phi}$ is well balanced relative the boundary ∂Y. □

In order to extend the Γ-pleating operation to framed well-balanced FLIFs we need to impose additional constraints on the choice of the vector field ν^+ and the curve Γ,

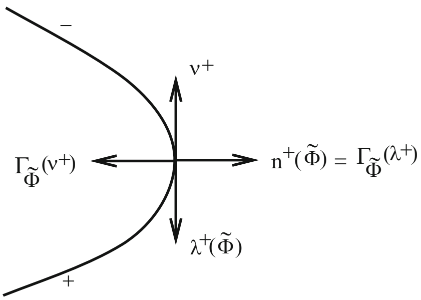

–
ν+
$\Gamma_{\tilde\Phi}(\nu^+)$
$n^+(\tilde\Phi) = \Gamma_{\tilde\Phi}(\lambda^+)$
$\lambda^+(\tilde\Phi)$
+

Fig. 12 Vector $n^+(\widetilde{\Phi})$ determines the characteristic coorientation of the fold

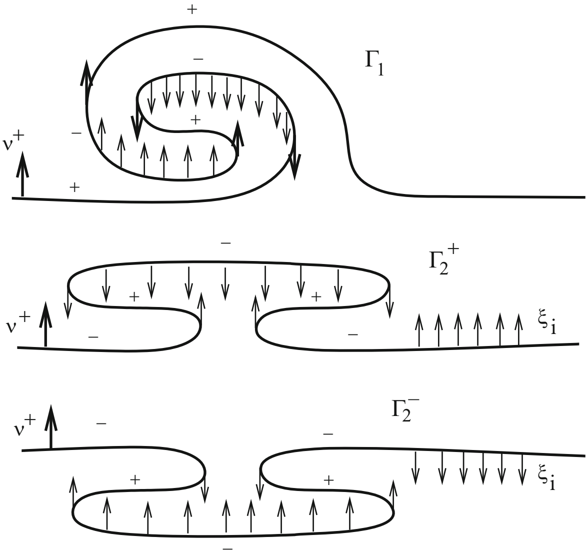


Fig. 13 Framing of a Γ-pleated FLIF

see Fig. 13. For each $j = 1, \ldots 2N$ denote by σ_j the proportionality coefficient in $\lambda^+|_{\widetilde{S}_j} = \sigma_j \nu^+|_{\widetilde{S}_j}$, $\sigma_j = \pm 1$. Then we require that

(α) if $\widetilde{S}_j$ and $\widetilde{S}_{j+1}$, $j = 1, \ldots, 2N - 1$ bound a negative component of $\widetilde{U} \setminus \widetilde{S}$ then $\sigma_j = \sigma_{j+1}$;
(β) if the component bounded by $\widetilde{S}_1$ and $\widetilde{S}_2$ is positive then $\sigma_1 = \sigma_{2N} = \pm 1$ for $\nu^+ = \pm \xi^i$.

Lemma 14 (Pleating a framed FLIF). *If ν^+ and Γ satisfy the above conditions, then given a framed well balanced FLIF (Φ, ξ) the FLIF $\widetilde{\Phi} = \mathrm{Pleat}_{S,\nu^+,\Gamma}$ admits a framing $\widetilde{\xi}$, where the framing $\widetilde{\xi}$ coincides with ξ outside $\widetilde{U}$.*

Proof. The proof is illustrated on Fig. 13. In the case $\nu^+ \in \mathrm{Vert}_+$ the pleating construction adds a 1-dimensional negative eigenspace to Vert_- restricted to negative components $\widetilde{U} \setminus \widetilde{S}$. Condition (α) then allows us to frame this 1-dimensional space either with $\xi^{i+1} := \sigma_j \nu^+$. Similarly, if $\nu^+ \in \mathrm{Vert}_-$ (or, equivalently when the component bounded by $\widetilde{S}_1$ and $\widetilde{S}_2$ is positive) then the pleating construction removes the negative eigenspace generated by ν^+ on positive components. The remaining negative components bounded $\widetilde{S}_{2j}$ and $\widetilde{S}_{2j+1}$, $j = 1, \dots, N-1$, can be framed with $\widetilde{\xi} := (\xi_1, \dots, \sigma_{2j}\xi_i)$. Condition (β) ensures that the existing framing in the complement of U satisfies the necessary boundary conditions on $\widetilde{S}_1$ and $\widetilde{S}_{2N}$. □

Lemma 15. *Given any framed well balanced FLIF (Φ, ξ), one of the curves $\Gamma_1, \Gamma_2^{\pm}$ shown on Fig. 6 can always be used as the curve Γ to produce a framed well balanced FLIF $(\widetilde{\Phi}, \widetilde{\xi})$ by a Γ-pleating.*

Proof. Indeed, as it follows from Lemma 14, the curve Γ_1 can always be used if $\nu^+|_S \in \mathrm{Vert}_+$, while if $\nu^+|_S \in \mathrm{Vert}_-$ then the curve $\Gamma_2^{\pm}$ can be used in the case $\nu^+ = \pm\xi^i$, see Fig. 13. □

The next proposition is a corollary of Lemma 2 and the results discussed in the current section.

Lemma 16 (Pleated isotopy of framed well balanced FLIFs). *Let ζ_s, $s \in [0,1]$, be a family of n-dimensional distributions on W, and (Φ, ξ) a framed well-balanced ζ_0-FLIF with $V(\Phi) = V \subset W$. Then there exist*

- *A framed well balanced ζ_0-$FLIF$ $\widetilde{\Phi}$ obtained from Φ by a sequence of pleatings, and*
- *A C^0-small isotopy $h_s : V \to W$, $s \in [0,1]$ such that h_0 is the inclusion $V \hookrightarrow W$ and $\widetilde{V}_s := h_s(V(\widetilde{\Phi}))$ is folded with respect to ζ_s along $\widetilde{\Sigma}_s := h_s(\Sigma(\widetilde{\Phi}))$.*

If Φ is holonomic over $\mathcal{O}p\,A$ then one can arrange that $\widetilde{\Phi} = \Phi$ on $\mathcal{O}p\,A$ and that the homotopy h_s is fixed over $\mathcal{O}p\,A$.

Proof. According to Lemma 2 there exists a manifold $\widetilde{V}$ for which the isotopy with the required properties does exist. This manifold can be constructed beginning from V by a sequence of Γ_0^+-pleatings along the boundaries of balls embedded into $V \setminus \Sigma$, in the direction of vector fields which extend to these balls. The latter property allows us to deform these vector fields into vector fields contained in Vert_+ or Vert_- (we need to use Vert_- only if $\dim \mathrm{Vert}_+ = 0$). Moreover, when using $\nu^+ \in \mathrm{Vert}_-|_{V_i}$ and when $i = \dim \mathrm{Vert}_-|_{V_i} > 1$ we can deform it further into the last vector ξ^i of the framing. In the case $i = 1$ we can deform ν^+ into $\pm\xi^i$, but we cannot, in general, control the sign. Note that we need to use this case only if $n = 1$. As it was explained in Remark 1, we can replace at our choice each Γ_0^+-pleating in the statement of Lemma 2 by any of the Γ-pleatings with $\Gamma = \Gamma_1, \Gamma_2^{\pm}$. But according to Lemma 15 one can always use one of these curves to pleat in the class of framed well balanced FLIFs. It remains to observe that if Φ is holonomic over $\mathcal{O}p\,A$ then all the constructions which we used in the proof can be made relative to $\mathcal{O}p\,A$. □

3.6 Stabilization

Let Φ be a ζ-FLIF. Suppose that we are given a connected domain $U \subset V \setminus \Sigma$ with smooth boundary such that the bundles $\mathrm{Vert}_{\pm}|_U$ are trivial. Let C be an exterior collar of $\partial U \subset V \setminus \Sigma$. We set $U' := U \cup C$.

Let us assume that U is contained in V^i. If $i < n$ we choose a section θ^+ of the bundle Vert_+ over U' and we define a *negative stabilization* of Φ over U as a FLIF $\widetilde{\Phi} = \mathrm{Stab}^-_{U,\theta^+}(\Phi)$ such that

- $\widetilde{\Phi}^1 = \Phi^1$;
- $\widetilde{\Phi}^2 = \Phi^2$ over $V \setminus U'$;
- $\Sigma(\widetilde{\Phi}) = \Sigma(\Phi) \cup \partial U$; $\mathrm{Int}\,U \subset V^{i+1}(\widetilde{\Psi})$;
- $\mathrm{Vert}_-(\widetilde{\Phi})|_{\mathrm{Int}\,U} = \mathrm{Span}(\mathrm{Vert}_-(\Phi)|_{\mathrm{Int}\,U}, \theta^+)$;
- $\lambda^+(\widetilde{\Phi})|_{\partial U} = \theta^+$.

We will omit a reference to θ in the notation and write simply $\mathrm{Stab}^-_U(\Phi)$ when this choice will be irrelevant.

Note that in order to construct $\widetilde{\Phi}^2$ on U' which ensures these property we need to adjust the background metric on ζ to make θ^+ an eigenvector field for Φ^2 corresponding the eigenvalue $+1$. The vector field θ^+ remains the eigenvector field for $\widetilde{\Phi}^2$ but the eigenvalue function is changed to $c : U' \to [-1, 1]$, where c is negative on U, equal to 1 near $\partial U'$ and has ∂U as its regular 0-level.

If the FLIF Φ is framed by $\xi = (\xi^1, \dots, \xi^i)$ then $\widetilde{\Phi}$ can be canonically framed by $\widetilde{\xi}$ such that $\widetilde{\xi} = \xi$ over $V \setminus U$ and $\mathrm{Vert}_-(\widetilde{\Phi})|_{\mathrm{Int}\,U}$ is framed by $\widetilde{\xi} := (\xi^1 \dots., \xi^i, \theta^+)$ and we define

$$\mathrm{Stab}^-_U(\Phi, \xi) := (\mathrm{Stab}^-_{U,\,\xi_i}(\Phi), \widetilde{\xi}),$$

In the case when $U \subset V_i$ and $i > 0$ we can similarly define a *positive stabilization* of Φ over U as a FLIF $\widetilde{\Phi} = \mathrm{Stab}^+_{U,\,\theta}(\Phi)$, where θ is a section of Vert_- over U' such that

- $\widetilde{\Phi}^1 = \Phi^1$;
- $\widetilde{\Phi}^2 = \Phi^2$ over $V \setminus U'$;
- $\Sigma(\widetilde{\Phi}) = \Sigma(\Phi) \cup \partial U$; $\mathrm{Int}\,U \subset V^{i-1}(\widetilde{\Psi})$;
- $\mathrm{Vert}_+(\widetilde{\Phi})|_{\mathrm{Int}\,U} = \mathrm{Span}(\mathrm{Vert}_+(\Phi)|_{\mathrm{Int}}\,U, \theta^+)$;
- $\lambda^+(\widetilde{\Phi})|_{\partial U} = \theta^+$.

If Φ is framed by a framing $\xi = (\xi^1, \dots, \xi^i)$ then we will always choose $\theta^+ = \xi^i|_U$ and define a positive stabilization by the formula

$$\mathrm{Stab}^{+}_{U}(\Phi, \xi) := (\mathrm{Stab}^{+}_{U,\,\xi_i}(\Phi), \widetilde{\xi}),$$

where $\widetilde{\xi}|_{\mathrm{Int}\,U} = (\xi^1, \dots, \xi^{i-1})$.

Lemma 17 (Balancing via stabilization). *Any FLIF can be stabilized to a balanced one. If Φ is balanced and $\chi(U) = 0$ then* $\mathrm{Stab}^{\pm}_{U}(\Phi)$ *is balanced as well. The statement holds also in the relative form.*

Proof. The obstruction for existence of a fixed over $A \subset V$ homotopy between two monomorphisms $\Psi_1, \Psi_2 : \mathrm{Norm} \to TW|_V$ is an n-dimensional cohomology class $\delta(\Psi_1, \Psi_2; V, A) \in H^k(V, A; \pi_k(V_n(\mathbb{R}^{n+k})))$, or more precisely a cohomology class with coefficients in the local system $\pi_k(V_n(T_vW)), v \in V$. Note that $\pi_k(V_n(\mathbb{R}^{n+k})) = \mathbb{Z}$ if k is even or $n = 1$ and $\mathbb{Z}/2$ otherwise. It is straightforward to see that

$$\delta(\Delta(\Phi), \Delta(\mathrm{Stab}^{\pm}_{U}(\Phi); U, \partial U) = \begin{cases} \chi(U)\Theta, & k \text{ is even;} \\ \pm\chi(U)\Theta, & k \text{ is odd,} \end{cases}$$

for an appropriate choice of a generator Θ of $H^k(U, \partial U; \pi_k(V_n(\mathbb{R}^{n+k})))$. Hence, stabilization over a domain with vanishing Euler characteristic does not change the obstruction class $\delta(\Gamma(\Phi), \Delta(\Phi))$ and with the exception of the case $k = n = 1$ this obstruction class can be changed in an arbitrary way by an appropriate choice of U. Indeed, if $k > 1$ then one can take as U either the union of l copies of n-balls or a regular neighborhood of an embedded bouquet of l circles (comp. a similar argument in [7]). If $k = 1$ and $n > 1$ then the sign issue is irrelevant because the obstruction is $\mathbb{Z}/2$-valued. If $k = n = 1$ then one may need two successive stabilizations in order to balance a FLIF. Indeed, the domain U in this case is a union of some number l of intervals, and hence $\chi(U) = l$. Thus the positive stabilization increases the obstruction class by l, while the negative one decreases it by l. Suppose, for determinacy, we want to stabilize over a domain in V_0. If we need to change the obstruction class by $-l$ then we just negatively stabilize over the union of l intervals. If we need to change it by $+l$ we first negatively stabilize over one interval I and then positively stabilize over the union of $l + 1$ disjoint intervals in I. □

3.7 From Balanced to Well Balanced FLIFs

Lemma 18 (From balanced to well balanced). *Let (Φ, ξ) be a balanced framed ζ-FLIF which is holonomic over a neighborhood of a closed subset $A \subset W$. Then there exists a framed well-balanced FLIF (Φ', ξ') which coincides with Φ over $\mathcal{O}p\, A$. In addition, $V(\Phi')$ is obtained from $V(\Phi)$ via a C^0- small, fixed on $\mathcal{O}p\, A$ isotopy.*

Proof. There exists a family of monomorphisms Ψ_s : Vert $\to TW$, $s \in [0,1]$, connecting Vert $\xrightarrow{\Delta_\Phi} {}^{\Phi}$Vert $\hookrightarrow \tau$ and the inclusion j : Vert $\hookrightarrow \tau$. The homotopy can be chosen fixed over $\mathcal{O}p\,A$. The family Ψ_s can be extended to a family of monomorphisms $\zeta \to TW$. We will keep the notation Ψ_s for this extension. Denote $\zeta_s := \Psi_s(\zeta)$, $s \in [0,1]$. Thus $\zeta_1 = \zeta$ and ζ_0 is an extension to W of the bundle Norm$^{\Phi}$. Lemma 11 then guarantees that the push-forward ζ_0-FLIF (Id, $\Psi_0)_*(\Phi,\xi)$ is well balanced. According to Lemma 16 there exists a well balanced framed ζ_0-FLIF $(\widehat{\Phi},\widehat{\xi})$ where $\widehat{V} = V(\widehat{\Phi})$ is obtained from V by a C^0-small isotopy which i8s fixed outside a neighborhood of V and over a neighborhood of A, and a C^0-small supported in $(\mathcal{O}p\,\widehat{V}) \setminus A$ isotopy g_s starting with $g_0 = \mathrm{Id}$ such that for each $s \in [0,1]$ the manifold $\widehat{V}_s := g_s(\widehat{V})$ is folded with respect to ζ_s along $\widehat{\Sigma}_s = g_s(\widehat{\Sigma})$. There exists a family of bundle isomorphisms $\Theta_s : \zeta_0 \to \zeta_s$ covering the diffeotopy h_s and such that $\Theta_0 = \mathrm{Id}$ and $\Theta_s = dg_s$ over the line bundle $TV|_{\widehat{\Sigma}} \cap \zeta_0$. The homotopy Θ_s can be chosen fixed over $\mathcal{O}p\,A$. Then, according to Lemma 10, the push-forward ζ-FLIF $(g_1,\Theta_1)_*(\widehat{\Phi},\widehat{\xi})$ is well balanced relative A. □

3.8 Formal Extension

Theorem 2 (Formal extension theorem). *Any framed ζ-FLIF (Φ,ξ) on $\mathcal{O}p\,A \subset W$ extends to a framed ζ-FLIF $(\widetilde{\Phi},\widetilde{\xi})$ on the whole manifold W.*

The proof is essentially Igusa's argument in [10] (see pp. 438–442).

We begin with the following lemma which will be used as an induction step in the proof.

Lemma 19 (Decreasing the negative index). *Let $j = 1,\dots,n$. Suppose W is a cobordism between $\partial_- W$ and $\partial_+ W$, and for a framed FLIF (Φ,ξ) on W one has $V^i = \varnothing$ for $i > j$. Then there exists a framed FLIF $(\widetilde{\Phi},\widetilde{\xi})$ such that*

- *$\Phi = \widetilde{\Phi}$ on $\mathcal{O}p\,(\partial_- W)$;*
- *$V^i(\widetilde{\Phi}) \cap \partial_+ W = \varnothing$ for $i \geq j$.*

Proof. of Lemma 19. To prove the claim we recall that the j-dimensional bundle Vert$_-$ over V^j is trivialized by the framing $\xi = (\xi^1,\dots,\xi^j)$, and $\xi^j|_{\Sigma^{j-1}} = \lambda^+$. We can extend the vector field ξ^j to a neighborhood G of V^j in $V^{j-1} \cup V^j \cup \Sigma^{j-1}$ as a unit vector field in V_+^{j-1}. Let X^j be the self-adjoint linear operator Vert$_+ \to$ Vert$_+$ defined on the neighborhood G which orthogonally projects Vert to the line bundle spanned by ξ^j. Choose neighborhoods $H_- \supset \partial_- W$ and $H_+ \supset \partial_+ W$ in W with disjoint closures and consider a cut-off function $\theta : V \to \mathbb{R}_+$ which is equal to 0 on $(V \cap H_-) \cup (V \setminus G)$ and equal to 1 on $V^j \cap H_+$. Set $\widetilde{\Phi}^2 := \Phi^2 + C\theta X^j$. Then for a sufficiently large $C > 0$ the self-adjoint operator $\widetilde{\Phi}^2$ coincides with Φ^2 on $V \cap H_-$, has negative index $\leq j$ everywhere, and $< j$ on $V^j \cap H_+$, see Fig. 14. The kernel of $\widetilde{\Phi}^2$ on $\Sigma^{j-1}(\widetilde{\Phi})$ is generated by ξ^j, and hence there is a canonical way to define

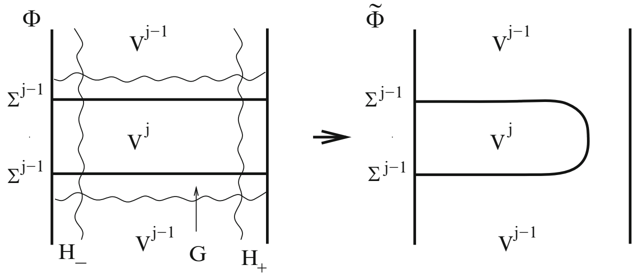


Fig. 14 Decreasing the negative index

the vector field $\lambda^+(\widetilde{\Phi})|_{\Sigma^{j-1}}$. Note that $\widetilde{\Phi} = \Phi$ in the complement $V \setminus V^{j-1}(\Phi) \cup V^j(\Phi) \cup \Sigma^{j-1}(\Phi)$ and at each point $v \in V^{j-1}(\Phi) \cup V^j(\Phi) \cup \Sigma^{j-1}(\Phi)$ the negative eigenspace $\mathrm{Vert}_-(\widetilde{\Phi})$ coincides either with $\mathrm{Vert}_-(\Phi)$, or with the span of the vectors $\xi^1, \dots, \xi^{j-1}$. Hence, the framing ξ of Φ determines a framing $\widetilde{\xi}$ of $\widetilde{\Phi}$. □

Proof. of Theorem 2. Let $\Phi = (\Phi^0, \Phi^1, \Phi^2, \lambda^+)$. Without loss of generality we can assume that (Φ, ξ) is defined on an $(n+k)$-dimensional domain $C \subset W, \mathrm{Int}\, C \supset A$, with smooth boundary. Note that if $\Phi^2|_{V \cap \partial C}$ is positive definite, then the extension obviously exist. Indeed, we can extend Φ^1 in any generic way to W, and then extend Φ^2 as a positive definite operator on Vert. We will inductively reduce the situation to this case. Let $C' \subset \mathrm{Int}\, C$ be a smaller domain such that $A \subset \mathrm{Int}\, C'$. Let us apply Lemma 19 to the cobordism $W_{(0)} = C \setminus \mathrm{Int}\, C'$ between $\partial_- W_{(0)} = \partial C'$ and $\partial_+ W_{(0)} = \partial C$ and to the restriction $(\Phi, \xi)|_{W_{(0)}}$ in order to modify $(\Phi, \xi)|_{W_{(0)}}$ into a framed FLIF $(\Phi_{(0)}, \xi_{(0)})$ which coincides with (Φ, ξ) near $\partial_- W_{(0)}$ and such that $V^n(\Phi_{(0)}) \cap (\partial_+ W_{(0)}) = \varnothing$. Then for a sufficiently small tubular neighborhood $W_{(1)}$ of $\partial_+ W_{(0)}$ in $W_{(0)}$ we have $W_{(1)} \cap A = \varnothing$ and $W_{(1)} \cap V^n(\Phi_{(0)}) = \varnothing$. We view $W_{(1)}$ as a cobordism between $\partial_- W_{(1)} = \partial W_{(1)} \setminus \partial_+ W_{(0)}$ and $\partial_+ W_{(1)} = \partial_+ W_{(0)}$. Now we again apply Lemma 19 to the cobordism $W_{(1)}$ and $\Phi_{(0)}|_{W_{(1)}}$ and construct a framed FLIF $(\Phi_{(1)}, \xi_{(1)})$ on $W_{(1)}$ which coincides with $(\Phi_{(0)}, \xi_{(0)})$ near $\partial_- W_{(1)}$ and such that $V^i(\Phi_{(1)}) \cap \partial_+ W_{(1)} = \varnothing)$ for $i \geq n-1$. Continuing this process we construct a sequence of nested cobordisms $C \supset W_{(0)} \supset W_{(1)} \supset \cdots \supset W_{(n-1)}$ and a sequence of framed FLIFs $(\Phi_{(j)}, \xi_{(j)})$ on $W_{(j)}$, $j = 0, \dots, n-1$, such that for all $j = 0, \dots, n-1$

- $\partial_+ W_{(j)} = \partial C$;
- $(\Phi_{(j+1)}, \xi_{(j+1)})$ coincides with $(\Phi_{(j)}, \xi_{(j)})$ on $\mathcal{O}p\,(\partial_- W_{(j+1)})$;
- $V^i(\Phi_{(j)}) \cap \partial_+ W_{(j)} = \varnothing$ for $i \geq n-j$.

Let us also set $W_{(n)} = \varnothing$. Hence we can define a framed formal Igusa function $(\widetilde{\Phi}, \widetilde{\xi})$ over C by setting $(\widetilde{\Phi}, \widetilde{\xi}) = (\Phi, \xi)$ on C' and $(\widetilde{\Phi}, \widetilde{\xi}) = (\Phi_{(j)}, \xi_{(j)})$ on $W_{(j)} \setminus$

$W_{(j+1)}$ for $j = 0, \dots, n-1$. Note that the quadratic part $\widetilde{\Phi}^2$ of $\widetilde{\Phi}$ is positive definite on ∂C, and hence the framed formal Igusa function $\widetilde{\Phi}$ can be extended to the whole W.

□

3.9 Integration Near V

Lemma 20 (Local integration of a well balanced FLIF). *Any well balanced $\mathcal{F}$-FLIF Φ can be made holonomic near V after a small perturbation near V. Namely, there exists a homotopy of well balanced FLIFs $\Phi_s, s \in [0,1]$, $s \in [0,1]$, beginning with $\Phi_0 = \Phi$ with the following properties:*

- $V(\Phi_s) = V(\Phi), \Sigma(\Phi_s) = \Sigma(\Phi)$ *for all* $s \in [0,1]$;
- (Φ_s^2, λ_s^+) *is* C^0*-close to* (Φ^2, λ^+) *for all* $s \in [0,1]$;
- Φ_1 *is holonomic on* $\mathcal{O}p\,V$.

If for a closed subset $A \subset W$ the FLIF Φ is already holonomic over $\mathcal{O}p\,A \subset W$ then the homotopy can be chosen fixed over $\mathcal{O}p\,A$.

Proof. According to Lemma 1 there exist local coordinates (σ, t, z, y) in a neighborhood of Σ in W, where $\sigma \in \Sigma,\ x, z \in \mathbb{R}$ and $y \in \mathrm{Ver}|_\Sigma$ such that the manifold V is given by the equations $z = x^2, y = 0$ and the foliation $\mathcal{F}$ is given by the fibers of the projection $(\sigma, x, z, y) \to (\sigma, z)$. The vector field $\frac{\partial}{\partial x}$ generates the line bundle $\lambda = TV|_\Sigma \cap \mathrm{Vert}$ and we can additionally arrange that $\frac{\partial}{\partial x}|_\Sigma = \lambda^+$. By a small C^0-small perturbation of the operator Φ^2 (without changing it along Σ) we can arrange the vector field $\frac{\partial}{\partial x}$ serves an eigenvector field for Φ^2 in a neighborhood of Σ. We will keep the notation λ for the extended line field $\frac{\partial}{\partial x}$. Then the operator $\Phi^2 : \mathrm{Vert} = \mathrm{Ver} \oplus \lambda \to \mathrm{Ver} \oplus \lambda$ can be written as $A \oplus c$, where A is a non-degenerate self-adjoint operator and c is an operator acting on the line bundle λ by multiplication by a function $c = c(\sigma, x)$ on $\mathcal{O}p\,\Sigma \subset V$ such that for all $\sigma \in \Sigma$ we have $c(\sigma, 0) = 0, d(\sigma) := \frac{\partial c}{\partial x}(\sigma, 0) > 0$.

Define a function φ on $\mathcal{O}p\,\Sigma \supset W$ given by the formula

$$\varphi(\sigma, x, z, y) = \frac{d(\sigma)}{6}(x^3 - 3zx) + \frac{1}{2}\langle Ay, y\rangle. \tag{3}$$

Then $V(\varphi) = V \cap \mathcal{O}p\,\Sigma$ and the operator $d^2_{\mathcal{F}}\varphi : \mathrm{Ver} \oplus \lambda \to \mathrm{Ver} \oplus \lambda$ is equal to $A \oplus \widehat{c}$, where the operator $\widehat{c}$ acts on λ by multiplication by the function $d(\sigma)x$. Hence the operator functions $d^2\varphi$ and Φ^2 coincides with the first jet along Σ, and therefore, one can adjust Φ^2 by a C^0- small homotopy to make Φ^2 equal to $d^2\varphi$ over $\mathcal{O}p\,\Sigma \subset W$. To extend φ to a neighborhood $\mathcal{O}p\,V \subset W$ we observe that the neighborhood of V in W is diffeomorphic to the neighborhood of the zero section in the total space of the bundle $\mathrm{Vert}|_{V \setminus U}$. In the corresponding coordinates we define

$\varphi(v, y) := \frac{1}{2}\langle \Phi^2(v)y, y\rangle$, $v \in V$, $y \in \mathrm{Vert}_v$. On the boundary of the neighborhood of Σ where we already constructed another function, the two functions differ in terms of order $o(||y||^2)$. Hence they can be glued together without affecting $d^2_{\mathcal{F}}\varphi$, and thus we get a leafwise Igusa function φ with $d^2_{\mathcal{F}}\varphi = \Phi^2$. It remains to extend $\nabla_{\mathcal{F}}\varphi$ as a non-zero section of the bundle $T\mathcal{F}$ to the whole W. According to Lemmas 4 and 7 we have $\Gamma_\varphi = \Pi_\varphi = d^2_\varphi = \Phi^2$. Then the well balancing condition for Φ implies that Γ_φ is homotopic (rel. $\mathcal{O}p\,A$) to Γ_Φ as isomorphisms Norm → Vert. But this implies that there is a homotopy (rel. $\mathcal{O}p\,A$) of sections $\Phi^1_s : W \to \mathrm{Vert}$, $s \in [0, 1]$, connecting $\Phi^1_0 = \Phi^1$ and $\Phi^1_1 = \nabla_{\mathcal{F}}\varphi$ and such that the zero set remains regular and unchanged.

□

4 Proof of Extension Theorem 1

STEP 1. FORMAL EXTENSION. We begin with a leafwise framed Igusa function (φ_A, ξ_A). Using Theorem 2 we extend it to a FLIF (Φ, ξ) on W.
All consequent steps are done without changing anything on $\mathcal{O}p\,A$.

STEP 2. STABILIZATION. Using Lemma 17 we make (Φ, ξ) balanced.

STEP 3. FROM BALANCED TO WELL BALANCED. Using Lemma 18 we further improve (Φ, ξ) making it well balanced.

STEP 4. LOCAL INTEGRATION NEAR V. Using Lemma 20 we deform (Φ, ξ) without changing $V(\Phi)$ to make it holonomic near V.

STEP 5. HOLONOMIC EXTENSION TO W. Now on $W \setminus \mathcal{O}p\,V$ we are in a position to apply Wrinkling Theorem 1.6B from [3] (see also [4], p. 335) to extend the constructed $\varphi_{A\cup V}$ as a leafwise wrinkled map $\varphi : (W, \mathcal{F}) \to \mathbb{R}$. The wrinkles of φ of any index have the canonical framing and thus this completes the proof of Theorem 1.

Acknowledgements We are grateful to D. Kazhdan and V. Hinich for their encouragement to write this paper and to S. Galatius for enlightening discussions.

References

1. V.I. Arnold, Wave front evolution and equivariant Morse lemma. Commun. Pure Appl. Math. **29**, 557–582 (1976)
2. Y. Eliashberg, Surgery of singularities of smooth maps. Izv. Akad. Nauk SSSR Ser. Mat. **36**, 1321–1347 (1972)
3. Y. Eliashberg, N. Mishachev, Wrinkling of smooth mappings and its applications – I. Invent. Math. **130**, 345–369 (1997)
4. Y. Eliashberg, N. Mishachev, Wrinkling of smooth mappings – III. Foliation of codimension greater than one. Topol. Method Nonlinear Anal. **11**, 321–350 (1998)

5. Y. Eliashberg, N. Mishachev, Wrinkling of smooth mappings – II. Wrinkling of embeddings and K.Igusa's theorem. Topology **39**, 711–732 (2000)
6. Y. Eliashberg, N. Mishachev, Wrinkled embeddings. Contemp. Math. **498**, 207–232 (2009)
7. Y. Eliashberg, S. Galatius, N. Mishachev, Madsen-Weiss for geometrically minded topologists. Geom. Topol. **15**, 411–472 (2011)
8. M. Gromov, *Partial Differential Relations* (Springer, Berlin/New York, 1986)
9. K. Igusa, Higher singularities are unnecessary. Ann. Math. **119**, 1–58 (1984)
10. K. Igusa, The space of framed functions. Trans. Am. Math. Soc. **301**(2), 431–477 (1987)
11. J. Lurie, On the classification of topological field theories. arXiv:0905.0465.
12. S. Smale, The classification of immersions of spheres in Euclidean spaces. Ann. Math. **69**(2), 327–344 (1959)

Quantum Gravity via Manifold Positivity

Michael H. Freedman

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract The macroscopic dimensions of space-time should not be input but rather output of a general model for physics. Here, dimensionality arises from a recently discovered mathematical bifurcation: "positive versus indefinite manifold pairings." It is used to build actions on a "formal chain" of combinatorial space-times of arbitrary dimension. The context for such actions is 2-field theory where Feynman integrals are not over classical, but previously quantized configurations. A topologically enforced singularity of the action can terminate the dimension at four and, in fact, the final fourth dimension is Lorentzian due to light-like vectors in the four dimensional manifold pairing. Our starting point is the action of "causal dynamical triangulations" but in a dimension-agnostic setting. Curiously, some hint of extra compact dimensions emerges from our action.

1 Introduction

Can one hope to reconstruct the universe from mathematics? What about its most prominent feature, its (at least coarse) 3+1-dimensionality? It is illuminating that in most formalisms, stable bound states do not easily arise in other dimensions [8, 18], so even a very weak "anthropic principle" would force 3+1 dimensionality. But to avoid completely the taint of circular reasoning, it would be desirable to construct a dimension-agnostic Lagrangian which can then be calculated to concentrate on

M.H. Freedman (✉)
Microsoft Corporation, CNSI Bldg. Rm. 2243, University of California 93106, USA
e-mail: michaelf@microsoft.com

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_6, © Springer-Verlag Berlin Heidelberg 2012

"realistic" 3+1-dimensional spaces. This paper is an initial step in this direction. In this spirit let us think about building up manifolds (space) of increasing dimension starting with the empty set by using the simplest possible operations: "cobounding" and "doubling along the boundary" (mirror double). The former is adjoint to integration and the latter generalizes $z \to z\bar{z}$ on complex numbers.

Thinking of the empty set as having dimension $= -1$,[1] locate a compact 0D manifold X^0, i.e. a finite set of points, and write $\partial^{-1}(\varnothing) = X^0$. (Yes, the boundary of a finite point set is the empty set.) Next let Y^0 be the union of X^0 together with a mirror image copy. Now find an X^1 with $\partial(X^1) = Y^0 \sqcup Y^{0'}$; X^1 is a cobordism from Y^0 to some arbitrary compact 0-manifold $Y^{0'}$. Double X^1 along its boundary to make Y^1 (a collection of circles) and find a surface X^2 satisfying $\partial(X^2) = Y^1 \sqcup Y^{1'}$, where $Y^{1'}$ is an arbitrary compact 1-manifold. Alternately doubling and cobounding produces manifolds X^d and Y^d of increasingly higher dimension d, which we picture as links in a chain X of manifolds $X^0, X^1, \ldots$. The idea is that for an appropriate action, explained below, this process will almost surely get stuck at X^4 or more precisely on some measure on the set of possible X^4's which constitutes a nonperturbative quantum gravity. At each step, choice of coboundary X^d is random but NOT uniform over cobounding manifolds, and is modeled after the procedure of "causal dynamical triangulations" (CDT) [1–3] which has been successful in producing phases in which most metrics fluctuate around those with flat space-like leaves and globally are somewhat deSitter-like.

We have been ambiguous about the signature in which these CDT-like constructions will be made. Actually, one beauty of CDT is that there is a well defined Wick rotation so one may pass back and forth between statistical and quantum mechanical interpretations at will. We also will use a topologically flexible version of CDT [1] which grow not just product collars but general manifolds of zero relative Euler characteristic. In fact, while we will arrange to concentrate the measure on manifolds, X and Y are a priori permitted to be singular.

It is hoped that the process sketched above can produce a superposition concentrated near solutions of Einstein's equations on smooth space-times (or in the Euclidean case a probability measure concentrated near manifolds whose metric is proportional to the Ricci tensor—i.e. "Einstein.") This hope is borrowed from the CDT community; our contribution is to treat the process CDT as recursive in dimension and describe a natural action for which the process almost surely terminates with X^4.

In geometry, it is natural to enhance manifolds to local products with small additional dimensions which can collapse without curvature blow-up [13]. Examples of this include Seifert fibered spaces as enhanced surfaces, Nil-bundles, and more generally manifolds with F-structures. Enhancement with Calabi-Yau directions appears similar, since the basic example of C-Y's are resolution of toroidal orbifolds.

[1] In topology it is natural to associate a negative integer as the dimension of the empty set and setting this to be -1 avoids delaying the nontrivial steps of the construction.

As explained below, there does appear to be some scope for our construction producing small toroidal directions; but unfortunately these are not adequate for standard model physics. It would be interesting to propose a variant which would generate additional compactified dimensions which concentrate in a useful locality of the infamous "landscape."

Using ideas of Connes, it should be possible to give a supersymmetric version of our action, but we will not treat that here.

Let us now discuss the main ingredients for the action S; see (6) for a fuller formulation. We will work with combinatorial d-manifolds X^d built up as in [1, 2] from layers of Lorentzian simplices with space-like edges having length$^2 = a$ and time-like edges having length$^2 = -\alpha a$. In the CDT literature, α is a constant, but one can be more flexible and regard it as a random variable drawn from some distribution. In the simplest model, X^d is built with a fixed space-like foliation but this should be relaxed [1] to allow certain topology changing singularities at constant time levels. Using Regge calculus, scalar curvature R can be defined and integrated on each X^d. Also, the boundary ∂X^d has a distribution valued second fundamental form whose norm squared should be included in S. We also permit X^d itself to be singular, i.e. not a manifold with boundary. This requires extending the definition of R to singular contexts. We do not have a specific proposal here, but note that it may be desirable to supress singular spaces within the path integral by choosing the extension so that they are assigned a large action. However, singularities—at least of the Lorentzian structure of X—should not be completely supressed. They are required to make contact with the smooth (actually P.L.) theory of manifold pairings. Processes that proceed through such singularities are useful as they "forget" details of the causal structure.

Letting G govern the strength of gravity and Λ be the bare cosmological constant, we write (schematically):

$$S_d^{\mathrm{Reg}}(X^d) = -\frac{1}{G}\int_{X^d} R + \delta \int_{\partial X^d} \|2^{\mathrm{nd}}\|^2 + 2\Lambda \mathrm{vol}(X^d) \tag{1}$$

(When we get to details we will actually double X to Y and use $S_d^{\mathrm{Euc}}(\mathrm{Y})$ and no boundary term.) The overall action S will include terms $S_d^{\mathrm{Reg}}(X^d), d = 0, 1, 2, \ldots,$ a fugacity for metric fluctuations, a volume, and a kinetic term.

Each X^d may not be a single piecewise Lorentzian "combinatorial" manifold, but a superposition. This means that the "histories" X over which we integrate to produce a partition function: $Z = \int_{\{X\}} \mathcal{D}X\ e^{-iS(X)}$[2] are not classical but already quantum mechanical objects. (This situation has previously been considered in cosmology [12, 17] under the name "third quantization.") Given a fixed combinatorial $d-1$ manifold Y^{d-1}, X^d may be a single manifold with $\partial X^d = Y^d \sqcup Y^{d'}$, or in the case $Y^{d'} = \varnothing$, X^d is permitted to be a linear combination of combinatorial d-manifolds X_i^d with boundaries equal to Y^{d-1} and normalized

[2] Actually we will work with a Euclidean, Wick rotated version of S.

Fig. 1 Commutative diagram for equivalent kets

$$\begin{tikzcd}
X^{d} \arrow[dd, "f"'] & \\
\arrow[r, phantom, "\circlearrowright"] & Y^{d-1} \arrow[ul, hook'] \arrow[dl, hook] \\
X^{d'} &
\end{tikzcd}$$

coefficients $c_1, \ldots, c_n \in \mathbb{C}$. Then X^d means

$$X^d = \sum_{i=1}^{n} c_i X_i^d, \qquad \sum_{i=1}^{n} |c_i|^2 = 1.$$

We actually permit the case where the sum is infinite and the coefficients L^2-convergent, but less is known mathematically about pairing L^2-combinations.

Finally we come to the pairing $\langle X^d, X^d \rangle$. In [4, 11, 14] the universal manifold pairings were defined and analyzed. Fixing a single closed $d-1$ manifold Y^{d-1}, define $\mathcal{M}_{Y^{d-1}}$ to be the $\mathbb{C}$ vector space of finite[3] linear combinations of the cobounding manifolds $\{X^d\}$ with $\partial X^d = Y^{d-1}$. $\mathcal{M}_{Y^{d-1}}$ becomes a Banach space by declaring $\{X^d, \partial X^d = Y^{d-1}\}$ orthonormal.

The manifolds X have usually been considered up to diffeomorphism or P.L. equivalence (rel boundary), meaning bounding X^d and $X^{d'}$ are the same ket if there is a diffeomorphism $f : X^d \to X^{d'}$ extending the identity $\partial X^d = \partial X^{d'} = Y^{d-1}$, i.e. if there exists an f making Fig. 1 commute.

We will also consider a finer equivalence where Y, X, and X' are metric and f required to be an isometry. $\mathcal{M}_Y^{\text{iso}}$ will denote finite combinations of isometry classes.

Gluing along the common boundary Y_{d-1} yields [4, 11] sesquilinear pairings:

$$\mathcal{M}_{Y^{d-1}} \times \mathcal{M}_{Y^{d-1}} \xrightarrow{\langle\ ,\ \rangle} \mathcal{M}_{\varnothing}. \tag{2}$$

The main result is that for $d > 3$ there are, for certain closed manifolds Y^{d-1}, light-like vectors $v \neq 0$ such that $\langle v, v \rangle = 0$, whereas for $d \leq 3$, $\langle v, v \rangle \neq 0$ for all Y^{d-1} and all $v \neq 0$. We say the low dimensional pairings are *positive*. (Later, we add a $\widehat{\ }$ to the notation for L^2-completions and L^2-pairings.)

To be more precise about the (2-)action S, we need to introduce a little more notation and come to terms with the fact that $X^d \in \mathcal{M}_{Y^{d-1}}$ may be a "superposition" of bounding manifolds—not a classical cobounding manifold. This 'superposition *inside* the "path integral" means that we must work in the context of higher order field theories [12, 17]. This concept is explained in more detail in Sect. 3 and Appendix B . But now let us interrupt our exposition of the technical set up to give in Sect. 2 some historical perspective on how 4-dimensional spaces have been, up until now, regarded as special. Finally, Sect. 4 discusses implications

[3] See Appendix A for both finite and L^2 sums and pairings.

and short-comings of our approach. Appendix A is on pairing Hilbert spaces of manifolds and, and Appendix B is on a formalism for higher quantum field theories. I would like to thank I. Klitch, J. Milnor, C. Nayak, and X. Qi for stimulating discussions on the topic of this paper.

2 4-D Manifolds are Different

$\mathbb{R}^n$ admits a unique smooth (also P.L.) structure for $n \neq 4$ and by DeMichelis and Freedman [5] continuum many smooth (P.L.) structures when $n = 4$. What is going on? The revolution in understanding 4-dimensional manifolds circa 1980 lead to three quite distinct perspectives on the question, "what is special about $D = 4$?"[4] The three answers may be summarized as:

1. Topological: 4-2-2=0,
2. Geometric: $so(4) \simeq so(3) \oplus so(3)$ is reducible,
3. Analytic: $L^{2,2}$ + C.B. Morrey condition $\Rightarrow$ Hölder continuity.

All three answers are essential to the theory of exotic $\mathbb{R}^4$'s.

In smooth and piecewise linear topology, general position is a powerful tool. It states that after perturbation two submanifolds of dimension p and q will meet in a submanifold of dimension $d - p - q$, where d is the ambient dimension. The reader may easily check this fact for affine subspaces of $\mathbb{R}^d$ and this is essentially the whole proof since "submanifold" is a local notion. Algebraic topology is dominated by chain complexes: sequences of modules and boundary maps—the latter encoding intersection points. It turns out that the key player [22] in cancelling oppositely signed intersection points is the Whitney disk, a 2-dimensional disk. How a Whitney disk will cross itself or another Whitney disk is governed by the general position formula:

$$\dim(\text{double pts(Whitney disk)}) = d - 2 - 2. \tag{3}$$

For $d \geq 5$, Whitney disks are imbedded, allowing cancellation; in these dimensions, the algebra of chain complexes fully describes topology [16]: "algebra = topology." Dimension 4 is a borderline case: Whitney disks have isolated point intersections. In this case, there is a useful topological [9]—but not *smooth*—technique for achieving cancellation and linking topology to algebra (Fig. 2).

This topological (but not smooth) Whitney trick allows homological algebra to successfully describe much of 4D topology—but not *smooth* topology—permitting the proliferation of smooth structures.

The Lie algebra of the orthogonal group is simple except for $so(4) \simeq so(3) \oplus so(3)$. Since curvature is a (Lie algebra valued) 2-form within $\Lambda^2(T^*; \mathrm{ad}\mathcal{G})$,

[4] Today a similar situation exists in dimension three. Three-manifolds admit rather disjoint understandings: hyperbolic geometry and Chern-Simons theory linked only weakly by the "volume conjecture."

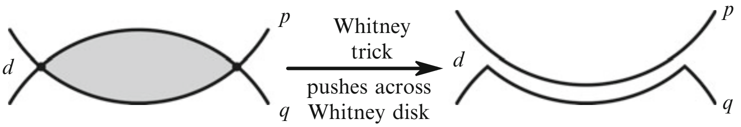


Fig. 2 The Whitney trick

the local identification of 2-forms with skew-symmetric matrices ($so(n)$) allow curvature—only in dimension 4—to be decomposed according to the eigenvalues of the Hodge $*$ operator into positive and negative parts, $\Lambda^2 = \Lambda^+ \oplus \Lambda^-$. The result is that the famous anti-self dual Yang-Mills equations and Serberg-Witten equations can only be formulated in dimension four. One may say that these equations lead to rather unique theories of *smooth* four dimensional spaces (including the exotic structures on $\mathbb{R}^4$,) but perhaps cannot explain how this space emerges in the first place.

Finally, and most technical, is the analytic answer to the question "what is special about 4-space?" This answer dates back to Uhlenbeck's [21] and Taubes' [19] work on the "bubbling" phenomenon and the existence of solutions to the self dual equation. In elliptic PDE, the equation itself is made to speak about the regularity of a weak solution f—this is the famous "boot strap." If the equation is second order, the ellipticity condition says that a weak solution in L^2 is also in $L^{2,2}$, i.e. the function and its first two distributional derivatives are in L^2. The ordinary Sobolev imbedding theorem says that

$$L^{2,2} \subset L^{0,q} \quad \text{for} \quad \frac{1}{q} > \frac{1}{2} - \frac{2-0}{d}. \tag{4}$$

For dimension $d \leq 3$, such solutions are uniformly continuous and the corresponding moduli spaces compact. For $d \geq 5$, $L^{2,2}$ formally is too weak to prove regularity. $d = 4$ is borderline. If one adds the "Morrey" condition that the $L^{2,2}$ norm of f decays at least as fast as r^α for some $\alpha > 0$ on balls of radius r, then f is not only in $L^{0,\infty}$, but is Hölder continuous. The Uhlenbeck-Taubes bubbling phenomenon happens at those isolated points where the Morrey condition is unobtainable. "Bubbling" is responsible for the noncompact product ends in the moduli spaces of anti-self-dual connections. Again, perspective three addresses how analysis works on a smooth 4-space, but does not suggest where or how such a space emerges.

In the last 5 years, the dimensional dichotomy, already discussed above—positive/indefinite manifold pairings—has emerged. Our idea is to build the manifold pairing into an action defined on candidate spaces (actually chains of spaces) and then use this action to construct a quantum gravity. There are two wrinkles which need to be appreciated from the start. First, the manifold pairing approach is dimension-agnostic, so the object that receives a weighting (Euclidean case) or action (Lorentzian case) is not a single 4-manifold but a "chain" starting with the empty set and proceeding upwards in dimension. Because of the nature of the paring and the form of the action, the chain almost surely terminates in dimension 4; this is derived and not assumed. Note that we use the term "chain" rather than "history," because we do not want to confuse the recursive dimension raising processes with the usual notion of time which is an aspect of the final 4-manifold, and not the way it "emerged" from the empty set. The chain is partially ordered and this order may be conceived as a fleeting "pre-time" or as a second independent direction of evolution.

The second wrinkle is that manifold pairings (or more precisely their associated quadratic forms) are defined not on a classical manifold M, but a superposition $\sum a_i M_i$. Chains are formal objects, and we will sometimes refer to them as such to emphasize that point. This means that the "path integral" is over superpositions. Partition functions in a quantum field theory (QFT) are calculated by integrating over classical objects, e.g. Brownian paths or connections on a bundle. However, for us the integral will already be over linearized[5] objects analogous to a vector in a Hilbert space whose kets are Brownian paths or connections. Such constructions are not unknown in quantum gravity [12, 17] and have been referred to as "third quantization" and "n^{th} quantization." We will introduce here only the aspects of this formalism which are presently required.

3 Chains, Action, Hamiltonian

We now describe the form of a "2-action" for quantum gravity in the context of a "2-quantum field theory" (2QFT). The essential feature of a 2QFT is a double layer of quantization. This means studying a wave function of wave functions or, via a Wick rotation to a Euclidean action, constructing a measure whose density is $e^{-S^E(\psi)}$. But instead of being a classical state, ψ is a normalized superposition $\psi = \sum a_i \psi_i$ so that $S^E(\psi)$ may be small or vanish due to interference effects from components of ψ. This has the consequence in 2-field theory that superpositions, which cancel rather than being unobserved (low amplitude), are instead likely to be observed because their action is small. Most of the formalism of 2-field theory is relegated to Appendix B ; here we proceed in a concrete ground-up fashion.

The Hilbert space $\mathcal{A}$ in which we will work has as its "kets" *formal chains* which start at the empty set $\varnothing$, and grow through a process borrowed from CDT, but now in a dimension-agnostic form. Prominent in the construction is "mirror double" which is a generalization of the norm^2 of a complex number $|z|^2 = z\bar{z}$. Here $Z\overline{Z}$ will have the meaning of gluing Z along a space-like boundary with its mirror image: $Z \to Z \cup \overline{Z} := Z\overline{Z}$. On a geometric level, leaving aside formal combinations, our CDT-like growth process starting with $d = 0$ consists of two cycling steps:

[5] We use "linearize" not to mean "to approximate by a linear system," but rather "to replace a set by the complex vector space it spans," e.g. as in the passage from a category to a linear category.

1. (Euclidean, dim $d-1$, manifold Y^{d-1})$\xrightarrow{\text{CDT}}$(Lorentzian, dim d, manifold X^d),
2. (Lorentzian, dim d, manifold X^d)$\xrightarrow{\text{mirror double}}$(Euclidean, dim d, manifold Y^d).

We have used the term manifold loosely. X is permitted singularities in both its causal (Lorentzian) structure and even its manifold structure. Similarly, Y may also be singular. After step 2, Y^d is now allowed to fluctuate to $\tilde{Y}^d$ breaking exact mirror symmetry, then one cycles back to step 1 (with $\tilde{Y}^d$ replacing Y^{d-1}). $\tilde{X}^d$ now doubles to Y^{d+1}. A chain contains $\varnothing, X^0, X^1, X^2, \ldots$ either terminating with $X^{d'}$ for some $d' \geq 1$, or continuing indefinitely.

To define a *formal chain*, we introduce formal combinations to the process. So

$$\varnothing \to X^0 = \sum_i a_i^0 X_i^0, \qquad \sum_i |a_i^0|^2 = 1.$$

Since there is no boundary $\partial X_i^0 = \varnothing$ to glue along, mirror double is simply disjoint union with the orientation reversed point set. The next arrow goes

$$\to \sum_{i,j} a_i^0 \overline{a_j^0} X_i^0 \overline{X_j^0} := \sum_{i,j} a_i^0 \overline{a_j^0} Y_{i,j}^0 := Y^0.$$

We now collect terms according to isometry type (in this case number of (+,–)-points) and write $Y^0 = \sum_{l_0} b_{l_0}^0 Y_{l_0}^0$. Next we would normally permit a topology (actually P.L. structure) preserving fluctuation $Y^0 = Y^{0,k=0} \to Y^{0,k=1} \to \cdots \to Y^{0,k^0} := \tilde{Y}^0$ to

$$\tilde{Y}^0 = \sum_{l_0} b_{l_0}^0 \tilde{Y}_{l_0}^0,$$

but in dimension zero there are no topology preserving fluctuations, so $Y^0 = \tilde{Y}^0$. The next arrow is

$$b_{l_0}^0 Y_{l_0}^0 \to X_{l_0}^1 = \sum_i a_{i,l_0}^1 X_{i,l_0}^1, \qquad a_{i,l_0}^1 = \alpha_{i,l_0} b_{l_0}^0, \ \sum_i |\alpha_{i,l_0}|^2 = 1 \text{ for all } l_0.$$

The terms X_{i,l_0}^1 are combinatorial Lorentzian 1-manifolds whose simplices have (length^2) = $-a\alpha$:

Because the X are Lorentzian, in this dimension the $X_{l_0}^1$ must be P.L. homeomorphic to $Y_{l_0}^0 \times I$ union possible additional circles (Fig. 3). The only restriction to obtain a non-singular Lorentzian extension is on the Euler characteristic $\mathcal{X}(X_{l_{d-1}}^d) = \mathcal{X}(Y_{l_{d-1}}^{d-1})$.

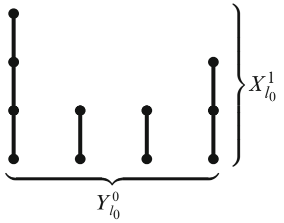


Fig. 3 The next arrow

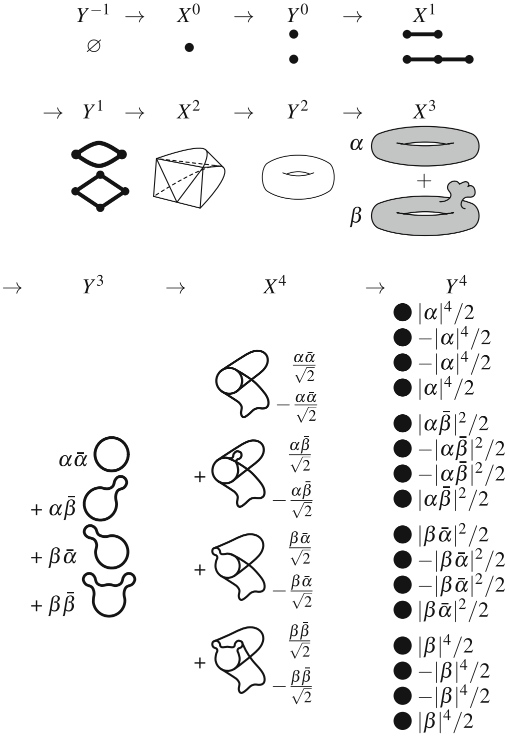
The process now continues this cycle:

$$\begin{aligned}
\varnothing \xrightarrow{\text{grow}} X^0 = \textstyle\sum_i a_i^0 X_i^0 &\xrightarrow{\text{double}} \textstyle\sum_{i,j} a_i^0 \overline{a_j^0} X_i^0 \overline{X_j^0} = Y^0 = \sum_{l_0} b_{l_0}^0 Y_{l_0}^0 \\
&\xrightarrow{\text{fluctuate}} \tilde{Y}^0 = \textstyle\sum_{l_0} b_{l_0}^0 Y_{l_0}^0 \\
\xrightarrow{\text{grow}} X^1 = \textstyle\sum_{i,l_0} a_{i,l_0}^1 X_{i,l_0}^1 &\xrightarrow{\text{double}} \textstyle\sum_{i,j,l_0} a_{i,l_0}^1 \overline{a_{j,l_0}^1} X_i^1 \overline{X_j^1} = Y^1 = \sum_{l_1} b_{l_1}^1 Y_{l_1}^1 \\
&\xrightarrow{\text{fluctuate}} \tilde{Y}^1 = \textstyle\sum_{l_1} b_{l_1}^{k_1,1} Y_{l_1}^{k_1,1} \\
&\quad\vdots \\
\xrightarrow{\text{grow}} X^4 = \textstyle\sum_{i,l_3} a_{i,l_3}^4 X_{i,l_3}^4 &\xrightarrow{\text{double}} \textstyle\sum_{i,j,l_3} a_{i,l_3}^4 \overline{a_{j,l_3}^4} X_i^4 \overline{X_j^4} = Y^4 = \sum_{l_4} b_{l_4}^4 Y_{l_4}^4 \\
&\xrightarrow{\text{fluctuate}} \tilde{Y}^4 = \textstyle\sum_{l_4} b_{l_4}^{k_4,4} Y_{l_4,k^4}^{k_4,4} \\
&\quad\vdots
\end{aligned}$$

where $a_{i,l_3}^4 = \alpha_i^{k_3,3} b_{l_3}^{k_3,3}$, $\sum_i |\alpha_i^{k_3,3}|^2 = 1$. Such a process is called a *formal chain.*

There is a general principle: If $\partial X_{l_{d-1}}^d \neq Y_{l_{d-1}}^{d-1}$, i.e. there is a non-empty "upper boundary," then $X_{l_{d-1}}^d$ cannot be a superposition for $d > 1$, since there will be no canonical way to identify boundary conditions (except when the boundary has dimension $= 0$) of the states supposedly in superposition. This severely limits superpositions within formal chains.

Thus a formal chain has sites or "vertices" which are formal spaces of increasing dimension and links which can be labeled by: "grow", "double", or "fluctuate." The word "formal" means "normalized complex-linear combination." We argue that the amplitude will concentrate on the special case: "formal non-singular Lorentzian manifolds" but a priori one should permit the growth process Euclidean-$d \rightarrow$ Lorentzian-$(d + 1)$ to add $d + 1$ simplices haphazardly. We permit the $(d + 1)$-Lorentzian simplicies to be fitted together without regard to Lorentzian or even manifold structure. The term in our action which we will denote $\int -R$, where R is Regge scalar curvature, should be extended in some (unspecified) fashion to penalize singularities of topology and Lorentzian structure. The role of singular



Y^{-1} → X^0 → Y^0 → X^1
$\varnothing$
→ Y^1 → X^2 → Y^2 → X^3
α
+
β
→ Y^3 → X^4 → Y^4
$\alpha\bar{\alpha}$
$+\,\alpha\bar{\beta}$
$+\,\beta\bar{\alpha}$
$+\,\beta\bar{\beta}$
$\frac{\alpha\bar{\alpha}}{\sqrt{2}}$
$-\frac{\alpha\bar{\alpha}}{\sqrt{2}}$
+
$\frac{\alpha\bar{\beta}}{\sqrt{2}}$
$-\frac{\alpha\bar{\beta}}{\sqrt{2}}$
+
$\frac{\beta\bar{\alpha}}{\sqrt{2}}$
$-\frac{\beta\bar{\alpha}}{\sqrt{2}}$
+
$\frac{\beta\bar{\beta}}{\sqrt{2}}$
$-\frac{\beta\bar{\beta}}{\sqrt{2}}$
$|\alpha|^4/2$
$-|\alpha|^4/2$
$-|\alpha|^4/2$
$|\alpha|^4/2$
$|\alpha\bar{\beta}|^2/2$
$-|\alpha\bar{\beta}|^2/2$
$-|\alpha\bar{\beta}|^2/2$
$|\alpha\bar{\beta}|^2/2$
$|\beta\bar{\alpha}|^2/2$
$-|\beta\bar{\alpha}|^2/2$
$-|\beta\bar{\alpha}|^2/2$
$|\beta\bar{\alpha}|^2/2$
$|\beta|^4/2$
$-|\beta|^4/2$
$-|\beta|^4/2$
$|\beta|^4/2$

Fig. 4 A formal chain terminating with $Y^4 = 0$

structures will be explained shortly. The action will favor cases in which the $d+1$ simplices are organized into a non-singular Lorentzian manifold with an Einstein metric.

It is permissible to think of every link in the chain as reversible so that given one chain c, it implies many related chains c' which simply walk (e.g. randomly) up and down c. Such c' will have larger action than c and be correspondingly supressed.

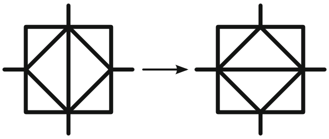
To summarize:

- "Growth": Euclidean-$(d-1)$ → Lorentzian-d adds Lorentzian d simplices.
- "Double": Lorentzian-d → Euclidean-d (Wick rotation).
- "Fluctuate": Euclidean-d → Euclidean-d alters the local combinatorial geometry.

Fluctuation must be allowed in order to make contact with topological pairing. Once fluctuation is permitted on the Euclidean space, it is perhaps natural to introduce it as well as an additional "link" on Lorentzian spaces. For simplicity we have not done this. The action which we describe next is a kind of Einstein-Regge action computed up and down the chain. Figure 4 gives a schematic depiction of a formal chain terminating with $Y^4 = 0$.

The Euclidean sites in a formal chain will concentrate at elements of the L^2 Hilbert space $\widehat{\mathcal{M}}_{\varnothing}^{d,\mathrm{Euc}}$ spanned by formal $\mathbb{C}$-combinations of simplicial Euclidean

Fig. 5 Geometric Pochner move at center in 2D



metric triangulated d-manifolds ($\widehat{\ }$ signifies L^2-completion, $d =$ dimension, Euc $=$ Euclidean, Lor $=$ Lorentzian, and $\varnothing$ indicates the empty set), although formally they are permitted to be more singular (see Fig. 6). Similarly, X^{d+1} concentrates in $\widehat{\mathcal{M}}^{d,\mathrm{Lor}}_{Y^d}$.

The spaces X^d "grow" on Y^{d-1} by adding Lorentzian simplices. We do not assume $X^d_{i,l_{d-1}}$ are manifolds, but the action should favor this case. The Y^d are made of "spatial doubles": glued copies of Y^d and $\overline{Y^d}$ across the space-like simplices. The fluctuations of Euclidean simplicial structure ($Y^d \to \tilde{Y}^d$) are assigned a fugacity f which depends on the geometric Pochner move or Euclidean geometry change occurring at each step $k-1 \to k$, and is to be extended linearly over superpositions, weighting by $|\text{amplitude}|^2$ (Fig. 5).

Since reflection inverts the Lorentzian light cones, the components of the double Y^d only have a canonical Euclidean (simplicial) structure obtained by Wick rotation of Lorentzian simplices to Euclidean geometry. The unscaled action term S_d on Y^d is of the form

$$S_d(Y^d) = \int_{Y^d} -\frac{R}{G} + 2\Lambda_d \, d\mathrm{vol} \tag{5}$$

where the integral of scalar curvature R is interpreted combinatorially [2] according to Regge calculus. No boundary term arises since Y^d is doubled. G is Newton's constant manifesting the strength of gravity, and Λ is a bare cosmological constant. Integrals over superpositions are to be extended linearly weighting by $|\text{amplitude}|^2$. The total action has the form

$$\begin{aligned} S(Y) = \textstyle\sum_{d=0}^{\infty} c_d &\left[S_d(Y^d_{l_d,k}) + \textstyle\sum_{k=0}^{k_d} f_d(Y^{k-1,d} \to Y^{k,d}) \right] \\ &+ \textstyle\sum_{d=0}^{\infty} \sum_{k=0}^{k_d} g_d |Y^{k,d}|^2 + \text{kinetic term} \end{aligned} \tag{6}$$

We define the *volume* term $|Y^{k,d}|^2 := \sum |b^{k,d}_{l_d}|^2$. Clearly, S depends on constants G, Λ_d, c_d, f_d, and g_d and the interesting regime appears to be for all d, $c_d \cdot f_d \ll g_d$. There are further hidden parameters in each dimension d. As in [1], the time-like edges of all Lorentzian simplices should have $(\text{length}^2) = -\alpha_d a$, where as Euclidean edges have $(\text{length}^2) = a$. We wish to take the constants, or random variable, α_k well within the "C-phase" of [3], where the CDT growth process produces roughly deSitter-like space-times. For large values of c_n, the least action principle suggests that the measure $e^{-S(Y)}$ will concentrate on formal chains which terminate,

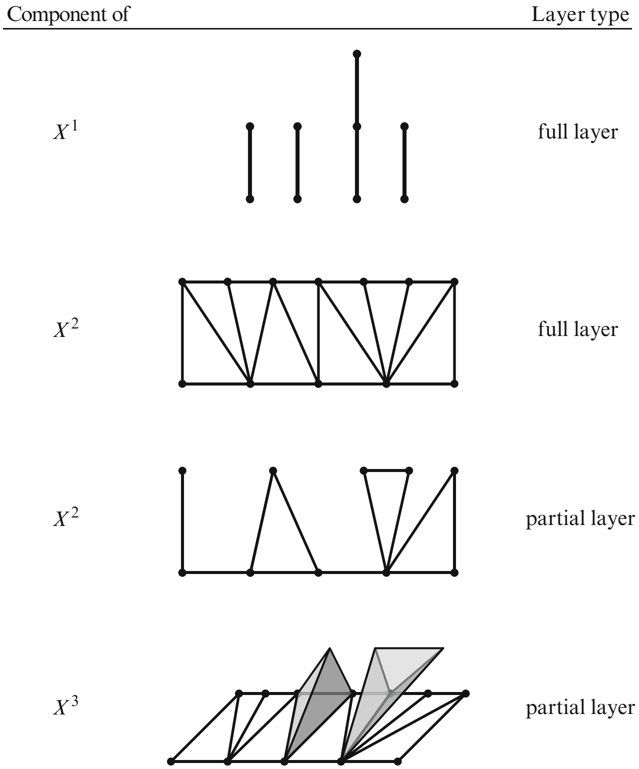


Fig. 6 Aspect ratios for $d = 1, 2, 3$

i.e. achieve all $b_{l_d}^{k,d} \equiv 0$, in the lowest possible dimension d, which is believed to be $d = 4$. (See Appendix A for the open mathematical point.) So we expect formal chains to terminate almost surely with a linear combination of X^4's.

If a classical chain c has amplitude b_i in formal chains $\overline{c}_i$ with normalized amplitudes a_i (computed from the action S), then c has probability $\sum_i a_i^2 b_i^2$. This and other aspects of the 2-field theory formalism are described in Appendix B .

The CDT growth process adds foliated layers of d dimensional Lorentz simplices to an initial $(d-1)$ dimensional Euclidean slice with aspect ratio length^2·(time-like edges)/length^2·(space-like edges) $= -\alpha_d$. In Fig. 6, this is illustrated for $d = 1, 2, 3$ where for $d = 2$ both a partial and full layer is illustrated and for $d = 3$ only a partial layer is shown. After Wick rotation to a Euclidean metric on $Y_{l_d}^{k_d,d}$ the simplices will have a distribution of Euclidean lengths so the condition for α_{d+1} to be in phase C above will be different from the numerically determined range in [3].

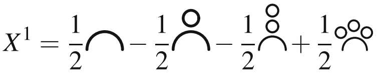
The definition for S_d (the Einstein Hilbert action with cosmological constant in dimension d) should be constructed to give a highly negative result (R near $-\infty$)

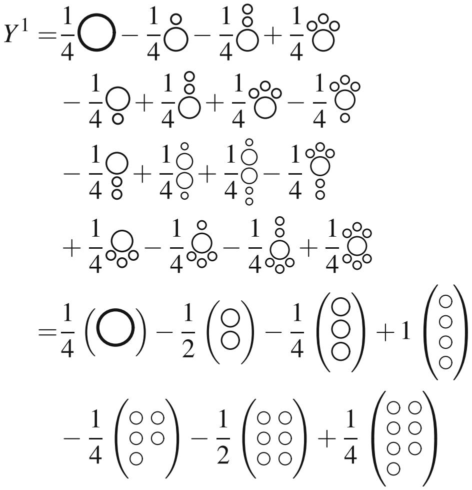
for Y^d, a double of X^d, unless X^d is a non-singular Lorentzian d-manifold with all spatial boundary. Thus the action blows up unless the constituent components of each X^d are Lorentzian manifolds and so the growth process concentrates on formal chains of manifolds—not singular spaces—of increasing dimension.

We have not been explicit on this point until now but as in [1], our CDT-like growth process should not be confined to building product collars, but rather it should allow singularities in the spatial foliations consistent with the creation of arbitrary d dimensional cobordisms $(X^d; X_+^{d-1}, X_-^{d-1})$, consistent with the single restriction on Euler characteristics: $\mathcal{X}(X^d) = \mathcal{X}(X_\pm^{d-1})$. This condition guarantees a relative reduction of the tangent bundle of X^d to $SO(d-1,1)$ and thus a $(d-1,1)$ signature pseudo-Riemannian metric. Fugacities for various foliation singularities must be regulated to obtain the benefits of CDTs, i.e. emergent geometry on the components of X^d with Hausdorff dimension $\approx d$.

Topological variability in the CDT growth through the dimensions allow the second, topological terms $g_d|Y^d|^2$ to vary. Recall that X^d may be a superposition. A sufficiently large constant g_d in S will, for $d \leq 3$, punish superpositions where "the collection of terms into isometry classes" subsequent to mirror doubling and fluctuation is reinforcing, and encourage cases where the collection involves cancellation. This can be seen in Example 1 below. It is a topological theorem (at least for finite superpositions) that not until $d = 4$ is reached can cancellation be complete.

Example 1. Topological (not Lorentzian) X^1 paired with itself.





Terms have been collected according to the topological (actually smooth) type, in this case increasing $|Y^1|^2$.

Note 1. In absence of cancellation, $|Y^1|^2$ would be equal to $|X^1|^2$. In this example, $|Y^1|^2 = \frac{7}{4}$.

We must now explain a rather unexpected possibility on which this paper rests. There are closed 3-dimensional manifolds, of the form Y^3, i.e. a double which bound two distinct 4-manifolds X_1^4 and X_2^4 with $\partial X_1^4 = \partial X_2^4 = Y^3$ so that if we set X^4 to be the formal 4-manifold $X^4 = X_1^4 - X_2^4$, then

$$Y^4 = \langle X_1^4 - X_2^4, X_1^4 - X_2^4 \rangle = X_1^4 \overline{X_1^4} - X_1^4 \overline{X_2^4} - X_2^4 \overline{X_1^4} + X_2^4 \overline{X_2^4} = 0 \in \mathcal{M}_\varnothing^4.$$

This happens because the four closed 4-manifolds appearing in the final sum are all diffeomorphic (equivalently P.L. homeomorphic.) (Similar examples are constructed in [11].) For some Y^3 and with little additional work (see Appendix B), we can ensure Y^3 is a double and $\mathcal{X}(X_1^4) = \mathcal{X}(X_2^4) = \mathcal{X}(Y^3) = 0$. The final column of Fig. 4 illustrates such a cancellation ($Y^4 = 0$).

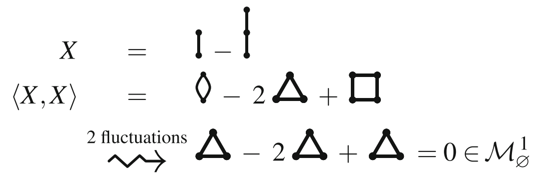
This kind of cancellation occurs in gluing manifolds of dim $d \geq 4$ [11, 14] but does not occur in gluing manifolds of dimension $d \leq 3$ [4]. Appendix B discusses what is known beyond the case of finite combinations considered loc. cit. in the context of completed pairings $\langle\ ,\ \rangle^\wedge$ on L^2 sequences of amplitude labeled manifolds $X^\wedge = \sum_i a_i X_i, \sum_i |a_i|^2 = 1$.

Because collecting terms may show $Y_{k_4}^4 = 0 \in \mathcal{M}_\varnothing^4$, the chain Y may terminate in dimension 4 (or possibly higher) with X^4. Given that the constants g_d are large, terminating chains are energetically favorable. If g_d is sufficiently large, we expect energy to dominate entropy and effectively *all* formal chains terminate at dimension 4. Thus, even at finite temperature β, a basic topological dichotomy may control the macroscopic dimension of space within this class of models.

We now describe the "kinetic" interaction included in (6). While formal smooth 4D spaces may cancel to zero when paired, cancellation never can occur in any dimension if the spaces come equipped with fixed triangulations and if the notion of isomorphism is restricted to a simplicial piecewise linear bijective map (or isometry). (To prove this, apply the discussion of "graph-pairings" within [4] to the dual graph to the codimension 1 simplicies of the fixed triangulations.) However, the fugacity f for geometric fluctuation $Y_{l_d}^{k-1,d} \to Y_{l_d}^{k,d}$ softens the pairing and permits cancellation and termination in dimension 4. But, these fluctuations are too much of a good thing as they allow cancellation in lower dimensions as well between terms that, while differing combinatorially, have identical topology and coefficients of opposite sign (see Example 2.) The job of the kinetic term is to prevent, cancellation ($|\tilde{Y}^d| = 0$) for $d < 4$. When $d = 4$, a new topological phenomenon arises and enforces cancellation for a new reason.

Here is an example of a potential—fluctuation induced—cancellation in dimension $d = 1$ in the combinatorial category.

Example 2.


X = −
$\langle X, X\rangle$ = − 2 +
2 fluctuations
− 2 + $= 0 \in \mathcal{M}^1_\varnothing$

This process has the potential to stop the growth of space in dimension one but may be thwarted by the "kinetic term" in S. If component combinatorial spaces $Y^{k,d}_{l_d}$ and $Y^{k+1,d}_{l_d}$ of a chain Y differ by a geometric Pochner move (or elementary metric change), we call them "nearest neighbors." The kinetic term, similar to $-h_d\sigma^x$ in lattice spin models, penalizes disparity in amplitudes $b^{k,d}_{l_d}$ for $Y^{k,d}_{l_d}$ and $b^{k+1,d}_{l_d}$ for $Y^{k+1,d}_{l_d}$ within the formal chain Y by adding

$$-2h_d \left| b^{k,d}_{l_d} - b^{k+1,d}_{l_d} \right|^2 \tag{7}$$

to the action for all nearest neighbor pairs. The strongest kinetic term would be a hard gauge-like constraint requiring terms differing by Pochner moves to have equal phases. The purpose of (earlier) permitting a non-zero amplitude for singular Lorentzian spaces is to allow the kinetic term to act across these and thus stiffen the phase not merely across spaces X with equivalent causal structures, but also between spaces X_1 and X_2 that are just relatively diffeomorphic, but have unrelated causal structures. Without this, we would encounter even in dimensions 2 and 3 nontrivial light-like vectors like $X_1 - X_2$, since the terms $X_i\overline{X}_j$ are all diffeomorphic for $1 \leq i, j \leq 2$.

There is an interesting statistical mechanics problem implicit in the kinetic term. It concerns the stiffness of the space of formal chains under linkages via nearest neighbor components (as above). In a nutshell, if one considers a graph G whose vertices are formal chains and whose edges are induced by (weighted) nearest neighbor occurrences, one asks if the first eigenvalue λ_1 of the graph Laplacian is positive, i.e. is G gapped or gapless? Is there a region in the large space of model parameters (but constrained by the necessity to lie in "C-phase" in all dimensions $d = 1, 2, 3, 4$) for which $\lambda_1(G) > 0$? In this situation, by setting h_d large enough, the various strands of the chain Y with differing combinatorial geometry but agreeing topology will be so stiffly bound together in phase that any cancellation of the type shown in Example 2 would be energetically unfavorable. The formal chain would be forced to develop up to dimension four where $\tilde{Y}^4$ can cancel out for topological reasons.

If it turns out there is no suitable regime in which $\lambda_1(G) > 0$, there are two possible solutions: (1) to make phase coherence of nearest neighbors a hard (gauge-like) constraint, or less drastically, (2) use a non-local kinetic term to stiffen G so that $\lambda_1(G) > 0$. Non-local means quite different but P.L. homeomorphic geometries directly interact. In condensed matter physics, the preference for local interactions is driven by the ubiquity of charge screening. In constructing a 2-action for a 2QFT, it was more an aesthetic choice to seek, first, a local interaction (among formal chains) sufficient to produce a satisfactory $3+1$ dimensional phase.

4 Conclusions

Physics may well be capable of generating in real time any mathematical tools it requires. This was famously the view of Richard Feynman. Another view is that new mathematical ideas, in this instance manifold pairings, may suggest new approaches to physical problems. Both view may be more or less valid at different times.

This paper begins an exploration of how the positivity of the low dimensional universal manifold pairing might yield a model for quantum gravity. The idea is to write a (2-)action S which picks out something like $(3+1)$-deSitter space from all possible pseudo metric spaces, ideally with no assumptions about regularity, dimension, or long scale structure. We have tried to keep the ingredients abstract and the action S simple. The results are (only) mildly encouraging. Enough has been seen to believe that manifold pairings can play a role in quantizing gravity, but it is quite open how best to formulate that role. This paper is a first attempt.

We began with the CDT approach of building P.L. Lorentzian cobordisms in Euclidean layers but started back at the empty set $\varnothing$, rather than an initial 3-sphere. To this we add the idea of superpositions of cobordism, metrical fluctuations, and a doubling operation $z \to z\bar{z}$ modeled on norm square of a complex number.

Superposition of cobordisms (thought of as paths) is essential to connect with the idea of manifold pairings. This means that the usual formalism for integrating over paths is not the correct analog, but rather one should integrate over superpositions of paths, i.e. *linearized paths*. This puts us in the realm of 2-field theory (see Appendix B), which we regard as a bonus. It seems natural that multiple layers of quantization would be encountered in the trip back to highest energy.

But overall, we are not completely happy with the notion of a formal chain and the action schema S we have written. Both the formal chain and the action S should be simpler—more fundamental. Perhaps general partial orders can stand in for Euclidean and Lorentzian (locally) flux simplicial structures (which appear already to assume too much.) Perhaps the fugacity for Euclidean fluctuation can be expressed as a perturbative consequence of *growth*. The goal would be to begin with an elementary combinatorial structure, the complex numbers, and a rather succinct 2-action S and extract space time. Preferably, the dimension $3+1$ would be singled out from all possible $p+q$, whereas we *assume* the form $p+1$. Also, it would be nice to see compactified dimensions emerge.

Actually, we do see a hint within S (Example 2) that space-time might not be simple deSitter-like but could in some parameter regimes contain small "compact" dimensions. Extra circle factors are the easiest to understand. There is less action

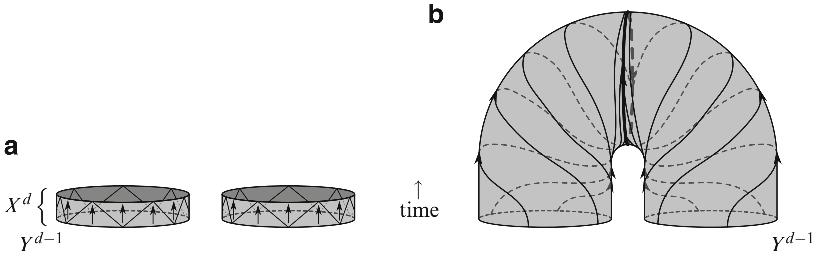


Fig. 7 A product (**a**) and a non-product (**b**) Lorentzian cobordism

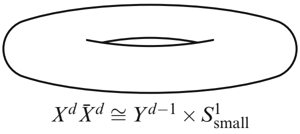


Fig. 8 Doubling Fig. 7a above yields a Euclidean manifold with a small *circle* factor

involved growing a small collar X^d than a cobounding manifold $X^{d'}$ with empty "upper" boundary (see Fig. 7).

The cost in action in building compact directions on the way to building a light-like vector v may be small enough that it wins for entropic reasons for some parameter settings of S. Although compact tori do not, apparently, lead to the field content required by standard model physics, this way of generating compact tori may be a useful start (Fig. 8).

The proposal in this paper may have a falsifiable prediction. In earlier drafts, we hoped to show that S^3 has no light-like vector v in its pairing $\langle\ ,\ \hat{\rangle}_{S^3}$. If this mathematical fact were established, it would seem to exclude S^3 as the spatial topology near the "big bang". We recently discovered light-like vectors in $\langle\ ,\ \rangle_{S^3}$ (see Appendix A), but the manifold constituents of v have much more homology than Mazur-like examples (again, see Appendix A) such as $M\#S^1 \times S^3$, based on a nontrivial homology sphere $\Sigma := \partial M$. Thus it is still possible that a non-trivial homology spheres Σ may be favored by the action over S^3.

A Appendix A: Manifold Pairings

Consider closed oriented d-manifolds $\mathcal{M}^d$ of class P.L. (or Diff.) Define $\mathcal{M}_{\varnothing}^{\text{P.L.}\,d}$ to be the $\mathbb{C}$-vector space consisting of finite linear combinations of P.L. homeomorphisms (alternatively homeomorphism or diffeomorphism) classes of $\mathcal{M}^d$. (Henceforth we treat only the P.L. case.) If S^{d-1} is closed of dimension $d-1$, define $\mathcal{M}_S^d$ to be the $\mathbb{C}$-vector space of finite linear combinations of cobounding M, $\partial M = S$, taken up to the equivalence relation of P.L. homeomorphism rel identity$_S$.

$$\begin{tikzcd} M_1 \arrow[dd, "f"'] & \\ & S \arrow[ul, hook', "\text{inc.}"'] \arrow[dl, hook, "\text{inc.}"] \\ M_2 & \end{tikzcd}$$

$|M_1\rangle = |M_2\rangle$ if and only if there is a P.L. homeomorphism f making this diagram commute

For each S there is a sesquilinear pairing:

$$\mathcal{M}^d_S \times \mathcal{M}^d_S \xrightarrow{\langle\ ,\ \rangle_S} \mathcal{M}^d_\varnothing,$$

$$\left(\sum_i a_i M_i, \sum_j b_j N_j\right) \mapsto \textstyle\sum_{i,j} a_i \overline{b_j} M_i \overline{N_j}$$

where $\bar{}$ means complex conjugation or orientation reversal according to context.

The literature [4, 11, 14] on $\langle\ ,\ \rangle_S$ may be summarized by:

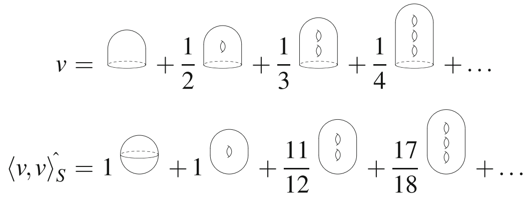
Theorem 1. *If $d \leq 3$, then for all S, $\langle\ ,\ \rangle_S$ is positive, meaning $\langle v, v\rangle_S = 0$ implies $v = 0$. For every $d \geq 4$, there is some S such that for some $v \neq 0 \in \mathcal{M}^d_S$, $\langle v, v\rangle_S = 0$. Such v are called light-like and such pairings indefinite. For $d = 4$, there are homology spheres Σ^3 for which $\langle\ ,\ \rangle_{\Sigma^3}$ is indefinite.*

In forming superpositions, L^2 rather than finite combinations of manifolds would be the more natural setting, so let us see which facts extend formally and which require work. Let a $\hat{}$ denote L^2-completion and let us add hats to the pairing and extend its natural $|\ |^2$ evaluation to $\mathbb{R}$.

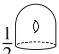
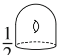
$$\begin{tikzcd} \mathcal{M}^d_S \times \mathcal{M}^d_S \arrow[r, "\langle\ ,\ \rangle_S"] & \mathcal{M}^d_\varnothing \arrow[r, "|\ |^2"] & \mathbb{R} \\ \arrow[r, phantom, "\wr"] & \sum_i c_i Y_i \arrow[r, mapsto] & \sum_i c_i \overline{c_i} w(Y_i) \\ \mathcal{M}^{\wedge d}_S \times \mathcal{M}^{\wedge d}_S \arrow[r, "\langle\ ,\ \rangle^{\wedge}_S"] & \mathcal{M}^{\wedge d}_\varnothing \cup \infty \arrow[r, "|\ |^2"] & \mathbb{R} \cup \infty \end{tikzcd}$$

where w is a weight function on $\{Y_i\}$, which for convenience we take to be $w(Y_i) = 1$.

Notice that $\langle\ ,\ \rangle^{\wedge}_S$ may not land in square summable sequences—hence the symbol ∞. For example, let $S = S^1$, the circle, and let



a series easily estimated *not* to be square summable.

Further notice that failure to converge is due to sufficient constructive interference. In the above example, $\langle v, v\rangle$ contains two terms topologically a torus, $\langle$, $\frac{1}{2}$$\rangle$ and $\langle\frac{1}{2}$, $\rangle$; the coefficients collect to $\frac{1}{2} + \frac{1}{2} = 1$, whose norm squared is 1. Without collection, the contribution to norm squared would be $\frac{1}{2}^2 + \frac{1}{2}^2 = \frac{1}{2}$. In fact, it is immediate that if *no* terms in the pairing can be collected (i.e. none are P.L. homeomorphic), then:

$$
\begin{aligned}
|\langle v, w\rangle|^2 &= (|\langle v||^2)(||w\rangle|^2) \\
&:= \left(\sum_i a_i \overline{a_i}\right)\left(\sum_j b_j \overline{b_j}\right),
\end{aligned}
$$

where $\langle v| = \sum a_i M_i$ and $|w\rangle = \sum b_j M_j$.

Oppositely, destructive interference reduces $|\langle v, w\rangle|^2$. In [11], we found for certain integral homology 3-spheres Σ that there were cobounding pairs of homotopy 4-balls A and B, $\partial A = \Sigma = \partial B$, so that the following 4-closed manifolds were all (oriented) P.L. homeomorphic to the 4-sphere S^4:

$$A\bar{A} \cong A\bar{B} \cong B\bar{A} \cong B\bar{B} \cong S^4$$

Certainly this means $v = |A\rangle - |B\rangle$ is a light-like vector for $\langle\ ,\ \rangle_\Sigma$.

$$\langle A - B, A - B\rangle = A\bar{A} - A\bar{B} - B\bar{A} + B\bar{B} = 0 \in \mathcal{M}^4_{\varnothing}.$$

We are not troubled by infinite values for $|\langle\ ,\ \rangle|^2$ since these will be accorded infinite energy by the action and in our formalism will never be observed. What we would like to know is that Theorem 1 for $d \leq 3$ remains valid after completion. Presently, we know this only for $d \leq 2$. For $d = 3$, we

Conjecture 1 *For all compact 2-dimensional surfaces S, the quadratic function on L^2 completions, $|\langle\ ,\ \rangle^{\wedge}_S|^2 : \mathcal{M}^{\wedge}_{S^2} \to \mathbb{R} \cup \infty$ has no kernel (i.e. $|\langle v, v\rangle^{\wedge}_S|^2 = 0$ implies $v = 0$).*

Discussion In the original (uncompleted) setting, positivity was proved by producing a (remarkably intricate) ordering of d-manifolds ($d \leq 3$) {P.L. homeo. types of closed d-manifold} $:= \{d\}$ to an ordered set $\mathcal{O}_d : \{d\} \overset{o}{\to} \mathcal{O}_d$ obeying what is called the *topological Cauchy-Schwartz inequality*: for all A, B with $A \neq B$ and $\partial A = \partial B = S$,

$$o(A\bar{B}) < \max\{o(A\bar{A}), o(B\bar{B})\}.$$

It is an immediate consequence that, for finite vectors $v_f = \sum_{i=1}^n a_i M_i$, $o(M_i \bar{M}_j)$ is maximized *only* on the diagonal by terms of the form $a_k \bar{a}_k M_k \bar{M}_k$. Since $a_k \bar{a}_k > 0$, these terms cannot cancel when the terms are collected (by P.L. homeomorphism type), thus $\langle v_f, v_f\rangle \neq 0$, and thus $|\langle v_f, v_f\rangle|^2 \neq 0$.

The argument breaks down for more general L^2-convergent sums v because $o(M_i\bar{M}_j)$ may not achieve a maximum at all. If the complexity function [4] can be altered to have an ascending chain condition (a.c.c.), all ascending chains have finite length, then the positivity theorem above would automatically extend to the L^2-completed pairing:

$$\hat{\mathcal{M}}^d_S \times \hat{\mathcal{M}}^d_S \xrightarrow{\widehat{\langle\,,\,\rangle_S}} \hat{\mathcal{M}}^d_\varnothing \to \mathbb{R}^+ \cup \infty, \qquad d = 0, 1, 2, \dots$$

This is easily done for $d \neq 3$. For example, when $d = 0$, 1, and 2, replace the complexity "number of connected components" by "-number of connected components" and when $d = 0$, |Euler characteristic| by Euler characteristic. So we have:

Theorem 2. *For $d = 0, 1,$ and 2, $\langle\ ,\ \rangle^\wedge$ is positive, i.e. $\langle v, v\rangle^\wedge = 0$ implies $v = 0$.*

When $d = 3$ our order contains real quantities such as partition functions of graph TQFT [15] and finite group TQFT [6], which do not lend themselves to an a.c.c. However, the single most important term in the $d = 3$ complexity function is -hyperbolic volume. Since the volumes of compact (or even finite volume) hyperbolic manifolds for a well ordered subset of $\mathbb{R}$, the a.c.c. holds and we have:

Theorem 3. *The L^2-completed hyperbolic manifold pairing*

$$\hat{\mathcal{M}}^3_{hyp,S} \times \hat{\mathcal{M}}^3_{hyp,S} \to \hat{\mathcal{M}}^3_{hyp,\varnothing} \to \mathbb{R} \cup \infty$$

is positive. The subscript "hyp" means the ket 3-manifolds M are compact hyperbolic and with totally geodesic boundary = S if $\partial M \neq 0$. The gluings defining the pairing are only homeomorphisms, not necessarily isometries.

Remark 1. It is a consequence of Thurston [20] that the geometric hypothesis M is *actually* a topological one: M should be irreducible, boundary irreducible, atoroidal, acylindrical, and with incompressible boundary. Furthermore, if M and M' obey these hypotheses with $\partial M = S = \partial M'$, then $M \cup_S M'$ admits a (unique) hyperbolic metric.

Remark 2. When the finite group TQFT term plays no role, the conjecture can be proved. This happens when the surface $S = S^2$ or a disjoint union of 2-spheres, or more generally when the kernel $(\pi_1(S) \to \pi_1(M_i))$ is fixed over all M_i with nonzero coefficients.

Finally, we prove two theorems about 3-manifolds S for which the $d = 4$ pairing is know to contain light-like vectors.

1. The 3-sphere S^3 has a light-like vector in its pairing $\langle\ ,\ \rangle_{S^3}$.

   Proof. According to [7], the anti-self-dual Donaldson invariants (ASDD) of closed 4-manifolds $\mathcal{M}$ with $b_2^\dagger \geq 3$ are stable with respect to complex blow up, i.e. connected sum with orientation reversed complex projective spaces: $\mathcal{M} \to \mathcal{M}\sharp\overline{CP}^2_S$. Since all orientation preserving automorphisms of S^3 are isotopic to id_{S^3}, $\mathcal{M}\overset{\mathit{diff}}{\cong}\mathcal{M}'$ if and only if $(\mathcal{M}_-, S^3)\overset{\mathit{diff}}{\cong}(\mathcal{M}'_-, S^3)$, the punctured manifolds with boundary are diffeomorphic. Let $\mathcal{M}$ be as above and $\mathcal{M}'$ be a smooth closed manifold s-cobordant to $\mathcal{M}$, but distinguished from $\mathcal{M}$ by an ASDD. $\mathcal{M}$ may be taken to be a $\mathbb{K}3$ surface and $\mathcal{M}'$ its logarithmic transform. Let $\sharp_n(\sharp_{-n})$ denote connected sum with n copies of $\overline{CP}^2$ (n copies of CP^2). Define $v_n \in \mathcal{H}_{S^3}$ as:

$$v_n = \mathcal{M}_-\sharp_n - \mathcal{M}'_-\sharp_n.$$

Then $\bar{v}_n = \overline{\mathcal{M}}_-\sharp_{-n} - \overline{\mathcal{M}}'_-\sharp_{-n}$, and so

$$\langle v_n, v_n\rangle = \mathcal{M}\sharp\mathcal{M}\sharp_n\sharp_{-n} - \mathcal{M}\sharp\mathcal{M}'\sharp_n\sharp_{-n} - \mathcal{M}'\sharp\mathcal{M}\sharp_n\sharp_{-n} + \mathcal{M}'\sharp\mathcal{M}'\sharp_n\sharp_{-n}. \quad (8)$$

The four manifolds $\mathcal{M}\sharp\mathcal{M}, \mathcal{M}\sharp\mathcal{M}', \mathcal{M}'\sharp\mathcal{M}$, and $\mathcal{M}'\sharp\mathcal{M}'$ are all s-cobordant. Note that $\sharp_n\sharp_{-n}$ is equivalent to connected sum of $\overline{CP}^2$, CP^2 and $(n-1)$ copies of $S^2\times S^2$, and that [10] s-cobordism becomes products after a finite stabilization by $S^2 \times S^2 \times I$. The result is that for n large, the 4 manifolds in (8) are all diffeomorphic. Since $v_n \neq 0$ for all n (by stability of the Donaldson invariants), for n large, v_n is a light-like vector.

2. If some 3-manifold M contains light-like vectors in its pairing, then any 3-manifold of the form $M\#N$ will as well, provided N admits P.L. (or smooth) imbedding $N \subset S^4$ into the 4-sphere.

Proof. Stabilize the terms of a null vector for M as follows: $v = \sum a_i W_i$ to $v' = \sum a_i(W_i\natural P)$ where we have taken the boundary connected sum with one of the closed complementary components of $N \subset S^4$: $S^4 = P\cup_N Q$. We observe that the composition:

$$W_i \hookrightarrow W_i\natural P \hookrightarrow W_i\natural P \cup_{N\setminus B^3} Q \cong W_i,$$

where $\natural$ denotes boundary connected sum, is simply addition of a product collar (an equivalence so

$$\begin{aligned} W_i\natural P \equiv W_j\natural P &\Rightarrow W_i\natural P \cup_{N\setminus B^3} Q \equiv W_j\natural P \cup_{N\setminus B^3} Q \\ &\Rightarrow W_i \equiv W_j. \end{aligned}$$

Thus $v \neq 0$ implies $v' \neq 0$. But all terms in $\langle v, v\rangle$ are each modified by connected sum with $P\bar{P}$, the double of P, to yield the corresponding term in $\langle v', v'\rangle$. Thus, term by term, we see $\langle v, v\rangle = 0$ implies $\langle v', v'\rangle = 0$.

In the construction of formal chains, we encounter 3-manifolds of the form $Y = X\bar{X}$ where $\partial X = T^2$, the 2-torus, with $\langle\ ,\ \rangle_{X\bar{X}}$ having light-like vectors. Let us understand why this is so. Using the above remark (twice) we can build such Y starting with the Mazur homology 3-sphere M for which light-like vectors were previously found [11]. An example of a $Y = X\bar{X}$, $\partial X = T^2$, can be manufactured as $X\bar{X} = M\#\bar{M}\#(S^1 \times S^2)$, where $X = (M\setminus B^3) \cup$ 1-handle and M is the Mazur homology 3-sphere. Because of the connected sum decomposition and the well known facts that $\bar{M}$ and $S^1 \times S^2$ is imbedded in S^4, $\langle\ ,\ \rangle_Y$ will have light-like vectors.

3. It is not yet proved that $\langle\ ,\ \rangle_S$ is positive for any smooth (P.L.) 3-manifold S.

B Appendix B: 2-Field Theory (and Higher)

We briefly explore a formalism for concatenated quantization. **Warning** This appendix is schematic, please read skeptically. We intentionally suppress analytic detail to sketch a broad picture. For example, any two linear spaces dense within a third function space are treated interchangeably.

In the main text, we worked in a Hilbert space $\mathcal{H}$ whose kets were formal chains—object which are built from linear combinations of piecewise linear spaces. Formal chains are themselves closed under $\mathbb{C}$-linear combinations (up to normalization) so $\mathcal{H}$ is a "relinearization" of an already linear space. It is not a foreign concept. Consider a typical single particle Hilbert space $\mathcal{H} = L^2(\mathbb{R}^3)$ that is promoted to (bosonic) Fock space $\mathcal{F}$ via a formal exponentiation, $\mathcal{F} = e^{\mathcal{H}}$:

$$\mathcal{F} = \mathbb{C} \oplus \mathcal{H} \oplus \frac{\mathcal{H} \otimes \mathcal{H}}{2!} \oplus \frac{\mathcal{H} \otimes \mathcal{H} \otimes \mathcal{H}}{3!} \oplus \dots \tag{9}$$

(the denominators are to remind us that symmetrization scales the inner products).

Since (9) describes polynomials in $\mathcal{H}$, $\mathcal{F}$ is dense in the linear space of continuous functions (weak topology) on $\mathcal{H}$, func$(\mathcal{H}, C)$. Furthermore, for wave functions (not basis kets) $\psi_i \in \mathcal{H}$, if we relinearize—notationally place kets around ψ_i $(|\psi_i\rangle)$—then the expression $\sum_{i=1}^N a_i |\psi_i\rangle \in \mathbb{C}[\mathcal{H}]$, the linear space of complex combinations of elements of $\mathcal{H}$. Dually, $\sum_{i=1}^N a_i |\psi_i\rangle$ determines a distribution (generalized function) $\sum a_i \delta_{\psi_i}$ on $\mathcal{H}$, which again form a dense set in generalized func$(\mathcal{H}, C)$. Thus, we regard, for example, $\mathbb{C}[\mathcal{H}] \sim \mathcal{F}$ as essentially equivalent since both are dense in func$(\mathcal{H}, C)$, and in this sense, view $\mathcal{F}$ as a relinearization of $\mathcal{H}$.

A chain, without linear combinations, is analogous to a Feynman diagram, which propogates dynamics in Fock space. A formal chain is a propogation at the next level represented by

$$\mathcal{A} \sim e^{\mathcal{F}} \subset \mathbb{C}^{\mathcal{F}} = \text{func}(\mathcal{F}, \mathbb{C}).$$

This is the signature of 2-field theory. In general, n-field theory has a Hilbert space at the $n-1$ level above Fock space

$$\mathcal{H} = \underbrace{e^{e^{\cdot^{\cdot^{\cdot^{e^{\mathcal{F}}}}}}}}_{n-1 e\text{'s}},$$

where unless otherwise noted, parentheses are inserted from top to bottom (e.g. $3^{3^3} = 3^{27}$). Using only dense linear subspaces within functions one may avoid the apparent explosion of cardinality. By passing to appropriate dense subspaces, we can keep all Hilbert spaces separable.

To lay the hierarchical structure bare, we work here with a model case, somewhat simpler than formal chains, in which higher Hilbert spaces are promoted from scalar fields $\phi \in L^2(\mathbb{R}^3, \mathbb{R})$ which, extending our policy of ignoring *all* analytical distinctions, we may simply write as functions:

$$L^2(\mathbb{R}^3, \mathbb{C}) \sim \mathbb{C}^{\mathbb{R}^3} \qquad \text{and} \qquad \mathrm{Fock}(L^2(\mathbb{R}^3, \mathbb{C})) \sim e^{\mathbb{C}^{\mathbb{R}^3}} \sim \mathbb{C}^{\mathbb{C}^{\mathbb{R}^3}}$$

For example, in 2-QFT, operators will act on 2-Fock:

$$\mathbb{C}^{\mathbb{C}^{\mathbb{C}^{\mathbb{R}^3}}} \qquad \left(\sim e^{e^{\mathbb{R}^3}}\right)$$

the linear space spanned by wave functionals of multiparticle wave functions ψ, $\Psi = \sum b_i |\psi_i\rangle$, i.e. *non-linear* functionals of multiparticle wave functionals.

If quantum field theory (QFT) computes some unitary fuzziness around classical trajectories, then it is the purpose of 2-QFT to compute some fuzziness around the unitary evolution of a QFT (which is itself unitary but only at a higher level.) To illustrate the scope of the idea, we will briefly touch on the "easier" and "harder" case of n quantum mechanics and n-string field theory. Regarding the terminology, n-QFT with its stratified structure is reminiscent of n-categories; we have kept the notation parallel. Finally, note the index n could also run over the ordinals but we have no use for that here.

Possible applications (besides to the body of this paper) include: (1) investigate models at high energy in which unitarity is only emergent, and (2) construct evective hierarchical description of strongly interacting low energy physics.

The constituents of an n-QFT are named in Table 1.

Observables are not really constituents wholly within quantum theory, but a bridge to the classical world, and so will be defined on the familiar level. Observables may include field strength (curvature), charge, and momentum.

Using this very crude notation, let's describe the Hilbert space for quantum mechanics, field theory, quantum field theory, string quantum field theory, nonlinear sigma models, and gauge field theory. The Hilbert space for QM is $\mathbb{C}^{\mathbb{R}}$, or more precisely $L^2(\mathbb{R})$ or $L^2(\mathbb{R}^n) = \otimes_n L^2(\mathbb{R})$. Now dropping all analytic detail, the space for field theory (FT) is $\mathbb{R}^{\mathbb{R}^3}$ for, say, a real field $\phi \in \mathbb{R}^{\mathbb{R}^3}$. The Hilbert space for QFT is Fock space $\mathbb{C}^{\mathbb{R}^{\mathbb{R}^3}}$, with wave functional $\psi = \sum a_i |\phi_i\rangle, \sum |a_i|^2 = 1$. The Hilbert space for 2-QFT is $\mathbb{C}^{\mathbb{C}^{\mathbb{R}^{\mathbb{R}^3}}}$, with $\Psi = \sum a_i |\psi_i\rangle, \sum |a_i|^2 = 1$.

To get string-QFT from QFT, you fiddle around at the "*top*" of the tower:

$$\begin{array}{ccc} \mathbb{C}^{\mathbb{R}^{\mathbb{R}^3}} & \rightsquigarrow & \mathbb{C}^{\mathbb{R}^{M^{S^1}}} \\ \text{Ordinary Fock space of a} & & \text{Stringy Fock space of a} \\ \text{real scalar field.} & & \text{real scalar field.} \end{array}$$

Table 1 The constituents of an n-QFT

n-Hilbert space	
n-Fock space	
$n\text{-}c^{+}_{\mathbb{k}}, n\text{-}c_{\mathbb{k}}$	$n+1$st quantized operators (to be consistent with the terminology of second quantization)
$n\text{-}H$	n-Hamiltonian
$n\text{-}U$	Unitary evolution at level n
$n\text{-}\mathcal{L}$	n-Lagrangian
$n\text{-}S$	n-action
	But there are *only* ordinary 1-observables

M is an 11-manifold, S^1 a circle which sweeps out a world sheet Σ in time. Both examples can be promoted to the 2-level simply by placing a "$\mathbb{C}$" at the lower left of the stack.

Of course, QFT's come in minor variations:

(a) $\mathbb{C}^{\mathbb{R}^{\mathbb{R}^3}} \rightsquigarrow \mathbb{C}^{X^{\mathbb{R}^3}}$ X, a manifold, is a "non-linear sigma model"

(b) $\mathbb{C}^{\mathbb{R}^{\mathbb{R}^3}} \rightsquigarrow \mathbb{C}^{\text{sections of a } G\text{-principle bundle over } \mathbb{R}^3}$ is a gauge field theory

Case (a) replacing $\mathbb{R}$ by X promotes a real scalar to a nonlinear sigma model. Case (b) functions are replaced by sections to yield a gauge field theory.

2-field theory adds a $\mathbb{C}$ at the bottom of the tower, so wave functionals are of the form $\psi \in \mathbb{C}^{\mathbb{R}^{\mathbb{R}^3}}$ and 2-wave functionals are of the form $\not{\psi} \in \mathbb{C}^{\mathbb{C}^{\mathbb{R}^{\mathbb{R}^3}}}$. 3-field theory treats 3-wave functionals $\not{\not{\psi}} \in \mathbb{C}^{\mathbb{C}^{\mathbb{C}^{\mathbb{R}^{\mathbb{R}^3}}}}$, and so on.

The usual passage[6] between H and $\mathcal{L}$, the "path integral formulation of QFT," is based on the ability to *restrict* fields on $\mathbb{R}^4$ to $\mathbb{R}^3 \times t$. Let's see how this works set theoretically. On adding a functional level, *inclusion* and *restriction* alternate.

$$\begin{array}{rcll}
\mathbb{R}^3 \times t & \hookrightarrow & \mathbb{R}^4 & \text{(inclusion of spaces)} \\
\mathbb{R}^{\mathbb{R}^3 \times t} & \leftarrow & \mathbb{R}^{\mathbb{R}^4} & \text{(restriction of fields)} \\
\mathbb{C}^{\mathbb{R}^{\mathbb{R}^3 \times t}} & \hookrightarrow & \mathbb{C}^{\mathbb{R}^{\mathbb{R}^4}} & \text{(inclusion of 2-fields)} \\
\mathbb{C}^{\mathbb{C}^{\mathbb{R}^{\mathbb{R}^3 \times t}}} & \leftarrow & \mathbb{C}^{\mathbb{C}^{\mathbb{R}^{\mathbb{R}^4}}} & \text{(restriction of 3-fields)}
\end{array}$$

It is important to be able to restrict fields to time slices, but you will notice that the restriction maps exist naturally only for k-fields, k odd. However, for k even, it is possible to pass to the linear duals $V \leftrightarrow V^*$, and ignore the analytic issue of the dual being a much larger space.

All books on QFT derive the evolution U from the Hamiltonian H as a "path integral" over fields ϕ weighted by $e^{-iS(\phi)}$, S the action of an ordinary Lagrangian, i.e. a 1-Lagrangian. Given, say, a 2-Hamiltonian 2-H, there will be a 2-Lagrangian,

[6] Between Hamiltonian and Lagrangian formalisms.

Table 2 Higher field theory. *Arrows* indicate levels which may be removed by squeezing higher order wave functionals

field	2-field	3-field
$\phi \in \mathbb{R}^{\mathbb{R}^4}$	$\not\phi \in \mathbb{C}^{\mathbb{R}^{\mathbb{R}^4}}$	$\not\not\phi \in \mathbb{C}^{\mathbb{C}^{\mathbb{R}^{\mathbb{R}^4}}}$
wave functional	2-wave functional	3-wave functional
$\psi(\phi) \in \mathbb{C}^{\mathbb{R}^{\mathbb{R}^4}}$ $\psi = \sum a_i \lvert\phi_i\rangle, \sum \lvert a_i\rvert^2 = 1$	$\not\psi(\not\phi) \in \mathbb{C}^{\mathbb{C}^{\mathbb{R}^{\mathbb{R}^4}}}$ $\not\psi = \sum a_i \lvert\not\phi_i\rangle, \sum \lvert a_i\rvert^2 = 1$	$\not\not\psi(\not\not\phi) \in \mathbb{C}^{\mathbb{R}^{\mathbb{R}^4}}$ $\not\not\psi = \sum a_i \lvert\not\not\phi_i\rangle, \sum \lvert a_i\rvert^2 = 1$
Fock(H) = F	2-Fock(H) = 2-F	3-Fock(H) = 3-F
$= \mathbb{C}^{\mathbb{R}^{\mathbb{R}^3}\}H}$ $= e^H$ $= \mathbb{C} \oplus H \oplus (H \oplus_S H) \oplus \ldots$	=Fock(Fock(H)) $= \mathbb{C}^{\mathbb{C}^{\mathbb{R}^{\mathbb{R}^3}}_{\nwarrow}}$ $= \mathbb{C} \oplus F \oplus (F \oplus_S F) \oplus \ldots$	=Fock$^3(H)$ $= \mathbb{C}^{\mathbb{C}^{\mathbb{C}^{\mathbb{R}^{\mathbb{R}^3}}_{\nwarrow}}_{\nwarrow}}$ $= \mathbb{C} \oplus 2\text{-}F \oplus (2\text{-}F \oplus_S 2\text{-}F) \ldots$
$\mathcal{L}^\lambda(\phi)$ and S^λ	2-$\mathcal{L}^{0,\lambda}(\not\phi)$ and 2-$S^{0,\lambda}$	3-$\mathcal{L}^{0,\lambda}$
$\int dx^4 \left((\nabla\phi)^2 - \frac{m^2}{2}\phi^2 - \frac{\lambda}{4!}\phi^4\right)$ for $\lambda = 0$, solve in k-space $\lvert 0\rangle$ and $c_k^\dagger$ generate F	$\int \mathcal{D}\phi \left((\nabla\not\phi)^2_\phi - \frac{m^2}{2}\lvert\not\phi\rvert^2 - \frac{\lambda}{4!}\lvert\not\phi\rvert^4\right)$ using translations in $\mathbb{R}^{\mathbb{R}^4}$, $\mathbf{k} \in (\mathbb{R}^*)^{\mathbb{R}^4}$ for $\lambda = 0$, solve in $\mathbf{k}$-space $\lvert\lvert 0\rangle\rangle$ and 2-$c_{\mathbf{k}}^\dagger$ generate 2-F, $\left[c_{\mathbf{k}_i}^\dagger, c_{\mathbf{k}_j}\right]_\xi = \delta_{ij}$	$\int \mathcal{D}\phi \left((\nabla\not\not\phi)^2_{\not\phi} - \frac{m^2}{2}\lvert\not\not\phi\rvert^2 - \frac{\lambda}{4!}\lvert\not\not\phi\rvert^4\right)$ $\lvert\lvert\lvert 0\rangle\rangle\rangle$ and 3-$c_{\mathbf{k}}^\dagger$ generate 3-F $\left[c_{\mathbf{k}_i}^\dagger, c_{\mathbf{k}_j}\right]_\xi = \delta_{ij}$

2-$\mathcal{L}$, constructed as a "path integral" over 2-fields $\not\phi \in \mathbb{C}^{\mathbb{R}^{\mathbb{R}^4}}$ weighted by $e^{-i(2\text{-}S(\not\phi))}$. Formally, this 2-evolution 2-U is *perfectly* unitary. The 2-evolution naturally "drags along" an *ordinary* 1-level linear evolution but this is *not* unitary and only becomes unitary in a certain *squeezed* limit (see below). Consider Table 2. Here, ∇ is the directional derivative at the next level:

$$\nabla_{\phi'}\not\phi|_\phi = {(\not\phi(\phi - \Delta\phi') - \not\phi(\phi))}\big/{\|\Delta\phi'\|_{L^2}}$$

$$\left(\nabla\not\phi|_\phi\right)^2 = \left(\int_{\|\phi'\|_{L^2}=1} d\phi' \,\|\nabla_{\phi'}\not\phi|_\phi\|^2\right)^{\frac{1}{2}}.$$

Parallel formulae give $\nabla\not\not\phi|_{\not\phi}$, and so on. We may also introduce a gradient ∇_x with fewer parameters (coming from a lower level). In the "squeezed context" explained below, ∇_ϕ may be replaced by ∇_x. Introduce the "small gradient" ∇_x based on $x \in \mathbb{R}^4$ (not $\mathbb{R}^{\mathbb{R}^4}$) translation. That is, define $\phi_{\Delta x}(x) := \phi(x - \Delta x)$, then define $\nabla_x\not\phi|_\phi = {(\not\phi(\phi_{\Delta x}) - \not\phi(\phi))}\big/{\Delta x}$, where $x \in \mathbb{R}^3$ or $x \in \mathbb{R}^4$, depending on context. Define a family of 2-actions for $c > 0$ by

$$2\text{-}\mathcal{L}^{c,\lambda} = |\nabla \not{\phi}|_{\phi}|^2 - \frac{m^2}{2}|\not{\phi}|^2 - \frac{\lambda}{4!}|\not{\phi}|^4 - c(\langle|\not{\phi}|^2\rangle - \langle\not{\phi}\rangle^2),$$
$$\text{where the last term is } c\left(\int \mathcal{D}\phi|\not{\phi}(\phi)|^2 - \left|\int \mathcal{D}\phi\not{\phi}(\phi)\right|^2\right).$$

As $c \to \infty$, the 2-physics of 2-$\mathcal{L}^{c,\lambda}$ is expected to concentrate on 2-fields, or "rules," $\not{\phi}$ which are nearly Dirac, i.e. $\not{\phi} \approx \delta\phi$, for some ϕ.

In the $c \to \infty$ limit, only $x \in \mathbb{R}^4$ translations have bounded energy among general variations, so ∇ is expected to reduce to ∇_x. This effectively deletes the $\mathbb{C}$ with the arrow next to it in Table 2. Thus, $c \to \infty$ "squeezes" 2-QFT back to ordinary QFT with $\frac{1}{c}$ the small parameter.

Similarly, let us define a 3-action

$$3\text{-}\mathcal{L}^{c,\lambda} = |\nabla \not{\not{\phi}}|_{\not{\phi}}|^2 - \frac{m^2}{2}|\not{\not{\phi}}|^2 - \frac{\lambda}{4!}|\not{\not{\phi}}|^4 - \text{squeezing term},$$

where the squeezing term—conceptually—is given by

$$\text{const}\min_{\phi_0} \underbrace{\int \mathcal{D}\not{\phi}|\not{\not{\phi}}(\not{\phi}) - \not{\phi}(\phi_0)|^2}_{f(\phi)}.$$

Setwise, evaluation includes {fields}$\subset \mathbb{C}^{\mathbb{C}^{\{\text{fields}\}}}$ by $\not{\not{\phi}}_{\phi}(\not{\phi}) := \not{\phi}(\phi)$. An analytically more convenient squeeze term is given by

$$\beta' \int \mathcal{D}\phi\; e^{-\beta f(\phi)},$$

where $\beta', \beta \gg 0$. As with 2-fields, we now expect that as $\beta \to \infty$, the "physics" of 3-fields will squeeze down to evaluation of 3-fields of the form $\not{\not{\phi}}_{\phi}(\not{\phi}) = \not{\phi}(\phi)$, i.e. a 1-field ϕ. It is also expected that $\int \mathcal{D}\not{\phi}|\nabla\not{\not{\phi}}|_{\not{\phi}}|^2 \rightsquigarrow \int dx^4|\nabla\phi|^2$, similarly for the mass and interaction terms.

Since 3 is odd, 3-fields naturally *restrict* to "time slices":

$$\mathbb{C}^{\mathbb{C}^{\mathbb{R}^3\times t}} \xleftarrow{\text{restriction}} \mathbb{C}^{\mathbb{C}^{\mathbb{R}^4}}.$$

The path integral allows the formal derivation of a unitary evolution 3-U starting from a Hermitian 3-Hamiltonian 3-H. This can also be accomplished at the 2-level by passing to linear duals:

$$\left(\mathbb{C}^{\mathbb{R}^3\times t}\right)^* \xleftarrow{\text{restriction}} \left(\mathbb{C}^{\mathbb{R}^4}\right)^*$$

Two final points should be explained: how the evolution at level n drags along a linear but not-quite-unitary evolution at all levels $m < n$, and what observables in n-QFT are. For both of these, we must define the "ket erasure" maps α_n.

"Erase kets and extend linearly" defines a linear map:

$$\overbrace{\mathbb{C}^{\cdot^{\cdot^{\cdot^{\mathbb{C}^{\mathbb{R}^3}}}}}}^{n} \xrightarrow{\alpha_n} \overbrace{\mathbb{C}^{\cdot^{\cdot^{\cdot^{\mathbb{C}^{\mathbb{R}^3}}}}}}^{n-1},$$

$$\sum a_i|\phi_i\rangle \xrightarrow{\alpha_n} \sum \tilde{a}_i\phi_i^n, \text{ where } \tilde{a}_i = \begin{cases} \bar{a}_i & \text{n odd} \\ a_i & \text{n even.} \end{cases}$$

There is also the familiar evaluation map e_{n-2}, given by

$$\overbrace{\mathbb{C}^{\cdot^{\cdot^{\cdot^{\mathbb{C}^{\mathbb{R}^3}}}}}}^{n-2} \xrightarrow{e_{n-2}} \overbrace{\mathbb{C}^{\cdot^{\cdot^{\cdot^{\mathbb{C}^{\mathbb{R}^3}}}}}}^{n},$$

$$e_{n-2}\phi_0^{n-2}(\phi^{n-1}) = \phi^{n-1}(\phi_0^{n-2})$$

Formally, $\alpha_{n-1} \circ \alpha_n \circ e_{n-2} = \mathrm{id}_{n-2}$, up to an infinite constant.

Proof. If $\not\not\phi(\not\phi) = \not\phi(\phi_0)$, then $\not\not\phi = \sum_i \not\phi_i(\phi_0)|\not\phi_i\rangle$, and so

$$\alpha_2\not\not\phi = \sum_i \bar{\not\phi}_i(\phi_0)\not\phi_i = \sum_i b_{i0}\not\phi_i,$$

where we have written $\not\phi_i = \sum_j b_{ij}|\phi_j\rangle$. Then

$$\begin{aligned} \alpha_1\alpha_2\not\phi &= \sum\nolimits_{i,j} b_{i0}\bar{b}_{ij}\phi_j \\ &= \sum\nolimits_i b_{i0}\bar{b}_{i0}\phi_0 + \sum\nolimits_{i,j\neq 0} b_{i0}\bar{b}_{ij}\phi_j \\ &= \infty(\phi_0) + \sum\nolimits_{j\neq 0} 0\phi_j, \end{aligned}$$

where zero on the last line comes from the symmetry of the sum.

Measurement will merely be by a Hermitian operator $\mathcal{O}$ on ordinary Fock space $F = \mathbb{C}^{\mathbb{R}^3}$. The protocol is "reduce, then observe": $\psi^n \xrightarrow{\alpha_n} \psi^{n-1} \to \cdots \to \psi^1$, and observe λ_i of $\mathcal{O}$ with probability $|a_i|^2$, where $\psi^1 = \sum a_i\psi_i^1$, where $\{\psi_i^1\}$ is an eigen-basis for $\mathcal{O}$. Suppose n is odd (and if not, pass to the dual). Then the successive evaluation maps promote ψ^1 back to level n where n-U evolves the promoted wave function until the next measurement by some $\mathcal{O}'$ also acting on ordinary Fock space $F = \mathbb{C}^{\mathbb{R}^{\mathbb{R}^3}}$. If the level n-evolution is sufficiently squeezed, then n-U evolves very nearly within evaluation subspace $F \subset n$-F and exact unitarity on n-F implies that a nearly exact unitarity will be observed on F.

Final notes and examples: The level 2 creation operators, $2\text{-}c^{\dagger}_{\mathbb{k}}$, create a set of states of varying particle numbers, e.g. the set may contain a scalar, a singleton of momentum k, linear combinations of pairs $(k' \otimes_S k'')$, and so on. In other words, $2\text{-}c_{\mathbb{k}}$ creates an arbitrary element of Fock space. $3\text{-}c_{\mathbb{k}}$ creates sets of sets of states, i.e. an arbitrary element in 2-Fock space, and so on.

Unitarity of the U is derived from the Lagrangian $\mathcal{L}$: $S = \int \mathcal{L}$ reverses sign (via complex conjugation) with reversal of orientation of slab $X \times [0, 1]$:

$$U_{ij} = \int_0^1 e^{iS} = \overline{\int_1^0 e^{iS}} = \overline{U}^{-1}_{ji}.$$

This argument is formally identical at level n.

Although this appendix has focused on n-QFT, one may promote the discussion to 2-string field theory or, in the other direction, cut the discussion down to n-quantum mechanics n-QM. By linearizing the top of the tower, we can produce 2-string FT:

$$\int_{\text{all 2D field theories}} S \longrightarrow \int_{\text{all 2D 2-field theories}} 2\text{-}S$$

Note 2. Among 2D field theories are nonlinear sigma-models of the form: (a string action, S)$\in \mathbb{R}^{M^{\Sigma}}$. Similarly, among 2D 2-field theories are function on nonlinear sigma-modules of the form: (a 2-string action, $2\text{-}S$)$\in \mathbb{C}^{\mathbb{R}^{M^{\Sigma}}}$. Presumably, these may be important in evaluating the integral perturbatively but are not exhaustive.

Now for 2-QM: consider a wave function $\psi \in H = \mathbb{C}^{\mathbb{R}^{\text{pt.}}}$ and a 2-wave function $\Psi \in 2\text{-}\mathcal{H} = \mathbb{C}^{\mathbb{C}^{\mathbb{R}}}$. To get a picture of how 2-QM can work, consider as a model for part of $2\text{-}\mathcal{H}$ consisting of $\Psi \in 2\text{-}\mathcal{H}$, made from just two Dirac functions,

$$\Psi = \frac{\sqrt{2}}{2}|\psi_1\rangle + \frac{\sqrt{2}}{2}|\psi_2\rangle,$$

where we think of ψ_i as the amplitude for particle i in position x_i.

Choose a 2-Hamiltonian analogous to an ordinary Hamiltonian for a "molecule" moving in potential:

$$2\text{-}H = \frac{1}{2}p^2_{\delta x_1} + \frac{1}{2}p^2_{\delta x_2} + V(x_1 - x_2) + \frac{x_1^2}{2} + \frac{x_2^2}{2} + \frac{\lambda}{4!}x_1^4 + \frac{\lambda}{4!}x_2^4,$$

where $p_{\delta x_i} = i\partial_{x_i}$ acts inside kets, so for example, $p_{\delta x_1+\delta x_2}\psi = \frac{\sqrt{2}}{2}|i\partial_{x_1}\psi_1\rangle + \frac{\sqrt{2}}{2}|i\partial_{x_2}\psi_2\rangle$.

Passing to a center of mass coordinate $\frac{x_1+x_2}{2}$, in the case where $\lambda = 0$, we have that

$$2\text{-}H = \frac{1}{2}p^2_{\delta x_1+\delta x_2} + \left(\frac{x_1+x_2}{2}\right)^2 + \frac{1}{2}p^2_{\delta x_1-\delta x_2} + \left(\frac{x_1-x_2}{2}\right)^2 + V(x_1-x_2),$$

so the center of mass is still SHO, and the evolution is actually unitary at the 1-level.

If $\lambda \neq 0$, the center of mass wave function at the 1-level is induced by ket erasure:

$$\phi(c) = \frac{\int dx_1[\phi_1(x_1) + \phi_2(2c - x_1)]}{\text{norm}}$$

does not evolve unitarily. I would like to thank Israel Klitch for suggesting this example.

Formal manipulations in 2-QFT, e.g. of (perturbed) Gaussian integrals, at higher levels will produce analogs of many familiar calculational features such as 2-ghosts, 2-Hubbard Stratonovich, and 2-perturbative expansions.

References

1. J. Ambjørn, R. Loll, Non-perturbative Lorentzian quantum gravity, causality and topology change. Nucl. Phys. B **536**(1–2), 407–434 (1998)
2. J. Ambjørn, J. Jurkiewicz, R. Loll, Reconstructing the universe. Phys. Rev. D **72**(6), 064014 (2005)
3. J. Ambjørn, J. Jurkiewicz, R. Loll, Quantum gravity as sum over spacetimes. arXiv:0906.3947v2 (2009)
4. D. Calegari, M. Freedman, K. Walker, Positivity in the universal pairing in 3 dimensions. J. Am. Math. Soc. **23**, 107–188 (2010)
5. S. DeMichelis, M. Freedman, Uncountably many exotic $\mathbb{R}^4$'s in standard 4-space. J. Differ. Geom. **35**, 219–254 (1992)
6. R. Dijkgraaf, E. Witten, Topological gauge theories and group cohomology. Commun. Math. Phys. **129**(2), 393–429 (1990)
7. S. Donaldson, P. Kronheimer, *The Geometry of Four-Manifolds*. Oxford Mathematical Monographs (Oxford University Press, New York, 1997)
8. P. Ehrenfest, Welche Rolle spielt die Dreidimensionalität des Raumes in den Grundgesetzen der Physik? Ann. Phys. **61**, 440 (1920)
9. M. Freedman, The topology of four dimensional manifolds. J. Differ. Geom. **17**, 357–453 (1982)
10. M. Freedman, F. Quinn, *Topology of 4-Manifolds* (Princeton University Press, Princeton, 1990)
11. M. Freedman, A. Kitaev, C. Nayak, J.K. Slingerland, K. Walker, Z. Wang, Universal manifold pairings and positivity. Geom. Topol. **9**, 2303–2317 (2005)
12. S. Giddings, A. Strominger, Baby universe, third quantization and the cosmological constant. Nucl. Phys. B **321**(2), 481–508 (1989)

13. M. Gromov, Groups of polynomial growth and expanding maps. Sci. Publ. Math. **53**, 53–73 (1981)
14. M. Kreck, P. Teichner, Positive topological field theories and manifolds of dimension > 4. J. Topol. **1**, 663–670 (2008)
15. R. Penrose, Applications of negative dimensional tensors, in *Combinatorial Mathematics and Its Applications*, ed. by D. Welsh (Academic, London, 1972)
16. S. Smale, Generalized Poincare's conjecture in dimensions greater than 4. Ann. Math. **74**(2), 391–406 (1961)
17. M. Srednicki, Infinite quantization. UCSB 88–07. The paper is still in preprint
18. F. Tangherlini, Schwarzchild field in n dimensions and the dimensionality of space problem. Nuovo Cimento **27**(10), 636–651 (1963)
19. C. Taubes, Self-dial Yang-Mills connections over non-self-dual 4-manifolds. J. Differ. Geom. **17**, 139–170 (1982)
20. W. Thurston, Hyperbolic structures on 3-manifolds I: deformation of acylindrical manifolds. Ann. Math. **124**(2), 203–246 (1986)
21. K. Uhlenbeck, Connections with L_p bounds on curvature. Commun. Math. Phys. **83**, 31–42 (1982)
22. H. Whitney, The self-intersections of smooth n-manifolds in $2n$-space. Ann. Math. **45**(2), 220–246 (1944)

Parabolic Explosions in Families of Complex Polynomials

Estela A. Gavosto and Małgorzata Stawiska

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract We present a new algebro-algebraic approach to the parabolic explosion of orbits for polynomials of a fixed degree $d \geq 2$, $P(z) = z^d + a_{d-1}z^{d-1} + \ldots + \lambda z$, $a_{d-1}, \ldots, a_2, \lambda \in \mathbb{C}$ where 0 is a multiple fixed point of $P_{\mathbf{a}}^{\circ q}$ for some $\mathbf{a} = (a_{d-1}, \ldots, a_2, \lambda_0)$ with $\lambda_0^q = 1$, $\lambda_0^k \neq 1$ for $k = 1, \ldots, q-1$. We show using methods based on Puiseux series that for an open dense set of perturbed maps with $\lambda = \lambda_0 \exp(2\pi i u)$, 0 becomes a simple fixed point and a number of periodic orbits of period q appear which are holomorphic in u^q. We also prove that the unwrapping coordinates for perturbations of an analytic map with a parabolic periodic point converge uniformly to the unwrapping coordinate for the map itself.

1 Introduction

There are many dynamical phenomena occuring in the one-parameter family of complex polynomials $P_\lambda(z) = \lambda z + z^2$. Throughout this paper we will assume that $\lambda^q = 1$, $\lambda^k \neq 1$ for $k = 1, \ldots, q-1$. Observe that 0 is a multiple fixed point of $P_\lambda^{\circ q}$ if $\lambda^q = 1$, $\lambda \neq 1$. Then, in the family of maps under consideration, so-called "parabolic explosion" takes place, which heuristically can be described as follows: when the parameter λ varies between $\lambda = e^{2\pi i p/q}$ and $\lambda' = e^{2\pi i(p/q+u)}$, with p, q coprime integers and u in a sufficiently small neighborhood of 0 in $\mathbb{C}$, the

E.A. Gavosto (✉)
Department of Mathematics, University of Kansas, Lawrence, KS 66045, USA
e-mail: gavosto@math.ku.edu

M. Stawiska
Mathematical Reviews, 416 Fourth St., Ann Arbor, MI 48103, USA
e-mail: stawiska@umich.edu

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_7, © Springer-Verlag Berlin Heidelberg 2012

fixed point 0 of $P_{\lambda'}^{\circ q}$ becomes simple and a new periodic q-cycle for $P_{\lambda'}$ appears near 0. An important property is that this cycle can be followed holomorphically in the variable $u^{1/q}$. The discussion of parabolic explosion (sometimes referred to as "implosion", because of change of properties of Julia sets associated to perturbed maps) was initiated by Douady in [11], in connection with the question of continuity of Julia sets depending on the defining polynomials. As far as other important problems of holomorphic dynamics are concerned, parabolic explosion and its control was essential for the proof of existence of quadratic Julia sets with positive measure by Buff and Chéritat [4, 5]. It also played a major role in the proof of the fact (due to Shishikura, [23]) that Hausdorff dimension of the boundary of the Mandelbrot set equals two. (For relations to other problems, see e.g. [2, 3].)

For one-parameter families of analytic maps of a real variable with real coefficients, $f(x) = (-1 + \varepsilon)x +$ higher order terms, ε small, (assuming some non-degeneracy conditions) an analogous phenomenon is known as a flip or period-doubling bifurcation (see [13], Theorem 3.5.1). It was proved in [6] (and later, in a simpler way, in [27]) that period-tripling (or higher multiplicity) bifurcations cannot occur for $\mathcal{C}^1$-families of self-maps of a compact interval or a circle. On the other hand, the "period-multiplying" bifurcations of complex quadratic polynomials were analyzed numerically already in [10, 15], where also certain universality properties were observed. The study of the parabolic explosion from the point of view of bifurcation theory was carried out in detail in [19, 22].

In this paper we take an algebro-geometric standpoint to look at bifurcations associated with polynomials of a fixed degree $d \geq 2$ such that 0 is one of fixed points for all polynomials in the (multi-parameter) family. More precisely, we consider $P(z) = z^d + a_{d-1}z^{d-1} + \ldots + \lambda z$, $a_{d-1}, \ldots, a_2, \lambda \in \mathbb{C}$ (this representation for P can be achieved by applying a translation of the complex plane). As before, we assume that 0 is a multiple fixed point of $P_{\mathbf{a}}^{\circ q}$ for some $\mathbf{a} = (a_{d-1}, \ldots, a_2, \lambda_0)$ with $\lambda_0^q = 1$, $\lambda_0^k \neq 1$ for $k = 1, \ldots, q-1$ (i.e., λ_0 a primitive q-th root of unity), and vary coefficients of $P^{\circ q}$. In Sect. 3 we prove our main result, namely that for an open dense set of perturbed maps with $\lambda = \lambda_0 \exp(2\pi i u)$, 0 becomes a simple fixed point and a number of periodic orbits of period q appear which are holomorphic in $u^{1/q}$.

In the proof we apply methods based on Puiseux series (see [25] for the treatment of quadratic polynomials). These methods are introduced and discussed in Sect. 2. Additionally, in Sect. 4, we analyze some facts related to the construction of Fatou coordinates for perturbed maps. We discover and make transparent a relation between different unwrapping coordinates, one used in [21] and the other (up to minor variations) in [5, 19, 22–24]. We then use this relation to prove that the unwrapping coordinates for perturbations of an analytic map with a parabolic periodic point converge uniformly to the unwrapping coordinate for the map itself. Our result (Theorem 6) unifies the two existing approaches to Fatou coordinates for perturbed maps within a natural algebraic framework and also offers a major shortcut to estimates involved in the construction of these coordinates.

2 Iterated Polynomials and Their Roots

Our study of parabolic explosion in a one-parameter family of polynomials relies on the normal form for P_λ^q near $z = 0$ due to [12]. Let us recall the general statement:

Theorem 1 ([12], Proposition 9.6). *Let $f : \mathbb{C} \mapsto \mathbb{C}$ be a polynomial of degree d, α be a periodic point for f of order k such that $\lambda = (f^{\circ k})'(\alpha) = e^{2\pi i p/q}$, with p and q coprime integers. Then the multiplicity of α as a fixed point of $f^{\circ qk}$ is of the form $\nu q + 1$, where $\nu \in \{1, \ldots, d-1\}$.*

Corollary 1. *Let $k = 1$ and $\alpha = 0$. Under the assumptions of Theorem 1, the following expansion holds in a neighborhood of* 0*:*

$$f^{\circ q}(z) = z + Az^{\nu q+1} + \mathcal{O}(z^{\nu q+2}),$$

with an $A \in \mathbb{C}^*$.

Example 1. (a) ($\nu = 2$) Let $P(z) = z^3 - iz^2 - z$. The normal form for $P^{\circ 2}$ is $P^{\circ 2}(z) = z + 4z^5 + \mathcal{O}(z^6)$

(b) ($\nu = 3$; cf. [8]) Let $P(z) = z^4 - z$. The normal form for $P^{\circ 2}$ is $P^{\circ 2}(z) = z - 4z^7 + \mathcal{O}(z^{10})$.

Recall also the following function, which was introduced in [9], Proposition 2.2: Let $\lambda = e^{2\pi i\theta}$, where $\theta = \theta_0 + u = p/q + u$, and $P_\lambda = P_{\mathbf{a},\lambda}$. Define $F(u,z) := (P_\lambda^{\circ q}(z) - z)/z$ for $z \neq 0$ and $F(u,0) = (\partial(P_\lambda^{\circ q}(z))/\partial z)\mid_{z=0} -1 = e^{2\pi iuq} - 1$. Then F is a function holomorphic in z and u, with the following Taylor expansion near $(0,0)$:

$$F(u,z) = 2\pi iqu + \mathcal{O}(uz) + Az^{\nu q} + \mathcal{O}(z^{\nu q+1}), \tag{1}$$

with $A \neq 0$ as in 1. The nonzero periodic points of P_λ of period q are solutions of the equation $F(u,z) = 0$. In what follows we will express z as $z(u)$ which is a series in nonnegative fractional powers of u (i.e., a Puiseux series). We will also describe the birth of periodic cycles.

Fix a natural number $d \geq 2$ and consider the family of polynomials $P_{\mathbf{a}',\lambda}(z) = z^d + a_{d-1}z^{d-1} + \ldots + \lambda z$, $a'_{d-1}, \ldots, a'_2, \lambda \in \mathbb{C}$, which in particular contains the subfamily $P_{\mathbf{a},\lambda_0}$ such that $\lambda_0 = e^{2\pi ip/q}$ with p, q coprime integers. In order to show the occurence of parabolic explosion in this family, we will proceed as follows: First we will fix $\mathbf{a} = (a_2, \ldots, a_{d-1}) \in \mathbb{C}^{d-2}$ and vary only $\lambda = e^{2\pi i(p/q+u)}$ with u in a sufficiently small neighborhood of 0 in $\mathbb{C}$. We will prove that the multiple fixed point 0 of $P_{\lambda_0,\mathbf{a}}^{\circ q}(z)$ gives rise to a simple fixed point $z = 0$ and ν periodic q-cycles near zero for $P_{\lambda,\mathbf{a}}$. The number ν is the same as in Theorem 1. Our argument will apply Puiseux theorem, using an approach that started in [25]. In particular, it will follow that all the periodic points are simple and are represented by holomorphic functions in u^q in a neighborhood of 0. To deal with an arbitrary $P_{\mathbf{a}',\lambda,}$, $\mathbf{a}' = \mathbf{a} + (u_2, \ldots, u_{d-1})$, where u_j are complex numbers in a neighborhood of 0, we will apply a theorem from [14]. Here and below, $\mathbb{C}[[u]]$ denotes the ring of formal power series in u with complex coefficients.

The following lemma gives factorization in $\mathbb{C}[[u,z]]$ of the function F introduced above:

Lemma 1. *(cf. [25], Theorem 1 for F associated with P of deg P =2.)*
If Q is a polynomial in z with coefficients holomorphic in u, irreducible in the ring $\mathbb{C}[[u]][z]$, and Y is a unit, then

$$F(u,z) = Q(u,z)Y(u,z),$$

Proof. We have $zF(0,z) = P^{\circ q}(z) - z = z^{\nu q+1}(z-c_1)\dots(z-c_n)$, where $c_1,\dots,c_n \neq 0$. By Hensel's lemma ([16], Theorem 1.16), $zF(u,z) = Q_0Q_1\dots Q_n$ in $\mathbb{C}[[u]][z]$, with $Q_0(0,z) = z^{\nu q+1}, Q_j(0,z) = z - c_j$, deg $Q_0 = \nu q + 1$, deg $Q_j = 1, j = 1,\dots,n$. Moreover, z divides Q_0, hence $F(u,z) = Q(u,z)Y(u,z)$ with $Q(u,z) = Q_0(u,z)/z$, $Y = Q_1\dots Q_n$. The polynomial Q is irreducible. Indeed, if $Q = R_1R_2$ with deg $R_1 = k, 0 < k < \nu q$, then $R_1(0,z) = z^k, R_2(0,z) = z^{\nu q-k}$ and $R_1(0,0) = R_2(0,0) = 0$. Hence $Q(u,z) = au^2 + \mathcal{O}(uz) + \mathcal{O}(z^2)$, which contradicts the normal form $F(u,z) = 2\pi qu + \dots$. □

Remark 1. Consider an open polydisk $\Delta = \{(\mathbf{a}',\lambda) \in \mathbb{C}^{d-2}\times\mathbb{C} : |a'_j - a_j| < \delta_j$, $|\lambda - \lambda_0| < \eta\}$, where $\mathbf{a} = (a_2,\dots,a_{d-1})$ is some point in $\mathbb{C}^{d-2}$ and $\eta,\delta_j > 0$ for $j = 2,\dots,d-1$. Let $D_{\mathbf{a}',\lambda}$ denote the discriminant of the polynomial $P^{\circ q}_{\mathbf{a}',\lambda} -$ Id. Since the Weierstrass polynomial Q of F is irreducible, it follows that F is irreducible in $\mathbb{C}[[u,z]]$ and, by Theorem 1.18 in [17], that the discriminant $D(u)$ of Q as a polynomial in z is not identically 0 in the ring $\mathbb{C}[[u]]$. Let Δ_1 denote the projection of Δ onto the plane $\{\mathbf{a}\}\times\mathbb{C}$. Then $0 \in \mathbb{C}$ is the only multiple fixed point of $P^{\circ q}_{\mathbf{a},\lambda}$ for $(\mathbf{a},\lambda) \in \Delta_1$. The neighborhood Q_1 of $(\mathbf{a},\lambda_0)$ determines a neighborhood of 0 in the complex plane corresponding to the variable u.

Knowing that F is irreducible in $\mathbb{C}[[u,z]]$, we can apply Puiseux expansion theorem in the following version:

Theorem 2. *(cf. [16], Corollary 3.13): Let $F(u,z) \in \mathbb{C}[[u,z]]$ with $F(0,0) = 0$ be irreducible and regular in z of order n (i.e., $\partial^n F/\partial z^n(0,0) \neq 0$). Then there exists $\sigma(u^{1/n}) = \sum_{k\geq 1} b_k u^{k/n} \in \mathbb{C}[[u^{1/n}]]$ such that $F(z,\sigma(u^{1/n})) = 0$. Moreover, any $\alpha \in \mathbb{C}[[u^{1/n}]]$ satisfying $F(z,\alpha) = 0$ is such that $\alpha = \sigma(\eta u^{1/n})$ for some n-th root of unity η.*

Corollary 2. *(cf. [16], Corollary 3.12 and proof of Lemma 3.15): Let F be as in Theorem 2 and let $Q(u,z)$ be the Weierstrass polynomial of F. Then $Q(u,z) = \prod_{j=0}(z - \sigma(\eta^j u^{1/n}))$, where η is a primitive n-th root of unity.*

Remark 2. In our situation F is a holomorphic function of the pair (u,z) and all statements above are true with formal series rings replaced by convergent series rings. Most importantly, there exists a holomorphic function $\sigma(t)$ defined on a small open disk $\Delta_\varepsilon = \{t : |t| < \sqrt[n]{\varepsilon}\}$ and an open neighborhood $U' = \Delta_\varepsilon \times \Delta_\delta \subset U$ such that $F(u,z) = 0$ for a pair $(u,z) \in U'$ if and only if $z = \sigma(t), u = t^n$ for some $t \in \Delta_\varepsilon$. Also, $Q(u,z) = \prod_{j=0}^{n-1}(u - \sigma(t\eta^j))$.

Let σ be a Puiseux series and let ζ be a root of unity of order n (not necessarily primitive). A series $\sigma_\zeta(t) := \sum \zeta^k b_k u^{k/n}$ is called a conjugate of σ. Note that in the decomposition of $f(u,z) = P_\lambda^{\circ q}(z) - z$ we have $r = 1$ and the function Y is the product of simple linear factors $z - c_i(u),\ i = 1,\dots,d^q - (\nu q + 1)$, where c_i are holomorphic functions of u in a neighborhood of 0 whose values at 0 are given by the nonzero simple roots of $P_{\lambda_0}^{\circ q}(z) - z = 0$. Let $\zeta = e^{2\pi i/\nu q}$ (in particular, ζ is a primitive root of unity of order νq), so that $\lambda_0 = e^{2\pi i \nu p/\nu q} = \zeta^{\nu p}$.

Theorem 3. *Let $\lambda = \lambda_0 \exp(2\pi i u)$ and P_λ be as above. Then all solutions of the equation*

$$P_\lambda^{\circ q}(z) - z = 0 \tag{2}$$

can be expressed as holomorphic functions of $u^{1/\nu q}$.

Proof. By the Puiseux theorem, the (2) has a solution $s = s(u) = \sum b_i u^{i/\nu q}$. Then $P_\lambda(s(u))$ is also a solution of (2). Note that $P_\lambda(s(u))$ equals the conjugate of s in which $u^{1/\nu q}$ gets replaced by $\lambda_0 u^{1/\nu q}$, as the corresponding analytic functions of the variable t, $P_\lambda(s(t^{\nu q}))$ and $s((\lambda_0 t)^{\nu q})$, have the same derivatives at $t = 0$. Similarly, $P_\lambda^{\circ 2}(s)$ is the conjugate of s by $\lambda_0^2 = \zeta^{2p\nu}$ etc., up to $P_\lambda^{\circ q}(s)$, which is the same as s. In this way we obtain an orbit of period q associated with a solution s of (2) in which all elements are holomorphic functions of $u^{1/\nu q}$. □

Remark 3. Thanks to our assumption about discriminants we know there are νq distinct solutions, so we can repeat the argument starting with a conjugate of s that does not belong to any of previously determined P_λ-orbits. Thus we get ν orbits of period q given by distinct Puiseux series in $u^{1/\nu q}$ in a neighborhood of 0.

3 Main Result

Now we will describe solutions of $P_{\lambda,\mathbf{a}'}^{\circ q} - z = 0$ for an arbitrary $\mathbf{a}' \in \mathbb{C}^{d-2}$ and their properties. Our main result can be stated as follows:

Theorem 4. *Let $P_{\mathbf{a}',\lambda}(z) = z^d + a'_{d-1}z^{d-1} + \dots + \lambda z$ be a family of complex polynomials, with $a'_j = a_j + u_j,\ j = 2,\dots,d-1,\ \lambda = \exp(2\pi i(p/q + u_1)),\ \lambda_0 = \exp(2\pi i p/q)$. Assume that $P_{\mathbf{a}',\lambda}^{\circ q}$ do not have multiple fixed points, except $P_{\mathbf{a},\lambda_0}^{\circ q}$, for which 0 is a multiple fixed point. Then for each $(\mathbf{a}',\lambda)$ close to $(\mathbf{a},\lambda_0)$ there are ν cycles of period q of $P_{\mathbf{a}',\lambda}$ close to 0, $1 \le \nu \le d-1$, whose points, after coordinate changes $u_1 = V_1^{\nu q}, u_j = V_j V_1 p_j, p_j \in \mathbb{N}, j = 2,\dots,d-1$, can be represented in some neighborhood of $0 \in \mathbb{C}^{d-1}$ as absolutely convergent power series in the variables $V_1,\dots,V_{d-1}$. The number ν depends on the polynomial P.*

In the proof we will use a result from [14]. The author of that paper considers a function $f : \mathbb{C}^d \mapsto \mathbb{C}$ which is analytic at the origin with $f(0) = f'_z(0) = 0$. He shows existence and finds the form of small solutions $z = z(u)$ (which means that $z(u) \to 0$ as $\|u\| \to 0$) of the equation

$$f(z, u_1, \dots, u_{d-1}) = 0 \tag{3}$$

in terms of the solutions of equation

$$f(z,0,\dots,0,u_k,0,\dots,0)=0,\ k=1,\dots,d-1. \tag{4}$$

The precise formulation is as follows:

Lemma 2. *(cf. Theorem in [14]) Suppose there exists $k\in\mathbb{N}$ such that the (4) has a simple small solution. Then the (3) has a small solution which (after the coordinate changes $u_k=V_k^r, u_j=V_jV_k^{p_i}$, $j=1,\dots,d-1$, $r\in\mathbb{N}$, $p_i\in\mathbb{N}$) can be represented as an absolutely convergent (in a neighborhood of 0) power series in variables $V_1,\dots,V_{d-1}$ without a free term.*

Proof. (of Theorem 4) In our case $u_1=u, u_j$, $j=1,\dots,d-1$ are small complex numbers, and $f(z,u)=P^{\circ q}_{\lambda,\mathbf{a}'}(z)-z$, so that $f(z,u_1,0,\dots,0)=P^{\circ q}_{\mathbf{a},\lambda}-z$, which by Theorem 3 has νq simple solutions. The existence of small solutions of 3 representable by fractional power series follows now directly from Lemma 2. It remains to identify q-cycles of $P_{\mathbf{a}',\lambda}$ among those solutions. The proof of Lemma 2 starts with fixing a solution of 4 of the form $\sum_{i=1}^{\infty}b_iV_1^i$ (where $u_1=V_1^{\nu q}$), then changing the coordinates $z=\sum_{i=1}^{d-1}b_iV_1^i+V_1^{d-1}y$ and $u_j=V_1^{p_j}V_j$, $j=2,\dots,d-1$ in the Taylor expansion of $f(z,u_1,\dots,u_{d-1})$ and dividing (3) by $V_1^{\nu q+d-2}$. As a result, one gets a function satisfying the assumptions of implicit function theorem, from which one obtains the unique solution y, which is a holomorphic function of $V_1,\dots,V_{d-1}$ near 0. A different initial choice of solution to (4) yields a different $z=\sum_{i=1}^{d-1}b_i'V_1^i+V_1^{d-1}y$, so we get νq small solutions of (3) (4). Similarly to the proof of Theorem 3, by taking the first partial derivative in V_1 at $(0,\dots,0)\in\mathbb{C}^{d-1}$ of z and $P_{\mathbf{a}',\lambda}(z)$ expressed as power series of $V_1,\dots,V_{d-1}$, we see that the $(d-1)$-th partial sum of $P_{\mathbf{a}',\lambda}(z)$ is the same as the $(d-1)$-th partial sum of the λ_0-conjugate of the solution of (4) that yielded z. Thus the periodic orbits of solutions of (4) yield periodic orbits of solutions of (3). □

4 Applications to Fatou Coordinates for Perturbed Maps

The local dynamics near a parabolic fixed point of a holomorphic map can be described by means of a Fatou coordinate. This follows from the Leau-Fatou Flower Theorem, which gives construction of this coordinate as a certain conjugacy map ([7], Theorem 1.2; cf. also [1, 8, 18]).

Theorem 5. *Let $f(z)=z+z^r+\mathcal{O}(z^{r+1})$ with $r>1$. Then there exist $(r-1)$ domains called petals P_j, symmetric with respect to the $r-1$ directions $\arg z=2\pi l/(r-1)$, $l=0,\dots,r-2$ such that $P_k\cap P_j=\emptyset$ for $j\neq k$, $0\in\partial P_j$, each P_j is conformally equivalent to the right half-plane H and $f^{\circ k}(z)\to\infty$ as $k\to\infty$ for all $z\in P_j$, all j. Moreover, for all j, the map $f\mid_{P_j}$ is holomorphically conjugate to the automorphism $z\mapsto z+i$ of H.*

Shishikura in [23, 24] constructed Fatou coordinates for analytic maps in a neighborhood of an $f_0(z) = z + z^2 + \ldots$. (See also [4,19] for alternative calculations and visualization). The paper [23] also sketched an idea for near-Fatou coordinates for f in a neighborhood of $f_0(z) = z + z^{r+1} + \ldots$ with $r > 1$, which was carried out in full detail in [22]. The case $r > 1$ was also analyzed by Oudkerk in [21] by methods relying on ordinary differential equations in the complex plane. All mentioned authors elaborated on the idea from [11], where Fatou coordinates were constructed for $f_\varepsilon(z) = z + z^2 + \varepsilon,\ \varepsilon \in (0, \varepsilon_0)$ with a small real ε_0 by means, among other things, of estimates for orbits of certain flows. In this section we will focus on some facts related to the construction of near- Fatou coordinates and describe them in terms of the periodic points near the origin that are created when the parameter is perturbed as in previous sections. By our main result, we can consider only one-parameter families, i.e., $P_\lambda(z) = \lambda z + z^2$ with $\lambda_0 = e^{2\pi i p/q}$ and $\lambda = e^{2\pi i (p/q+u)}$, with p, q coprime integers and u in a sufficiently small neighborhood of 0 in $\mathbb{C}$. We assume here that $\nu = 1$.

Let $Q_0(u, z) = zQ(u, z)$ be the Weierstrass polynomial for $P_\lambda^{\circ q}(z) - z$ with coefficients analytic in u. In the variable z, we have $Q_0(z) = z(z - \sigma_1)\ldots(z - \sigma_q)$. Changing coordinates as in [23] we can write $P^{\circ q}(z) - z = z(z^q - \sigma^q)h(z)$ with $h(0) \neq 0$, i.e., we can take $\sigma_i = \zeta^{i-1}\sigma,\ i = 1, 2, \ldots, q$, where ζ is the primitive q-th root of unity. Then we define an unwrapping coordinate $w(z) = (1/q\sigma^q)\log((z^q - \sigma^q)/z^q)$ (choosing an appropriate branch of logarithm). We have the following:

Theorem 6. *Up to an additive constant, $w(z)$ is equal to the integral*

$$\int_{z_0}^{z} \frac{1}{Q_0(u, \zeta)} d\zeta.$$

Proof. Observe first that $Q_0'(0) = -\sigma^q$ and $Q_0'(\sigma_i) = q\sigma_i{}^q = q\sigma^q$, since for all i, $\sigma_i^q = \sigma^q$. (The symbol' denotes here differentiation in z.) By Subramaniam and Malm [26], for $Q_0(z) = z(z^q - \sigma^q)$, the integral equals $\frac{1}{Q_0'(0)}\log z + \sum_{i=1}^{q} \frac{1}{Q_0'(\sigma_i)}\log(z - \sigma_i)$ (up to an additive constant). Hence $\int_{z_0}^{z}(1/Q_0(u, \zeta))d\zeta = (1/q\sigma^q)(-\log(z^q) + \prod_{i=1}^{q}\log(z - \sigma_i)) = w(z)$. □

Recall that the construction of Fatou coordinate for a map with a parabolic fixed point starts with applying the unwrapping map $z \mapsto -\frac{1}{q}z^q$. We will now show that this is the limit of our unwrapping coordinates w.

Proposition 1. *$w(z) \to -1/qz^q$ uniformly for $|z| > R > R_0$ (with R_0 big enough) as $\sigma \to 0$.*

Proof. Note that for $q = 1, 2, \ldots$, $z \mapsto z^q$ is a proper holomorphic mapping with a pole at infinity. It is therefore enough to consider $w(\zeta) = (1/\sigma')\log((\zeta - \sigma'/\zeta))$ with substitution $\zeta = z^q,\ \sigma' = \sigma^q$ and prove the proposition only for $q = 1$. The pointwise convergence is due to the fact that $\log'(z) = 1/z$ (this is how one proves the known formula for the complex potential of an electric dipole at $z = 0$ obtained from two charges $1/\sigma$ and $-1/\sigma$ placed respectively at $z = 0$ and $z = \sigma$ as $\sigma \to 0$; cf. [20]). The expansion of $w(z) + 1/z$ in powers of $1/z$ is:

$$(1/\sigma)\log(1 - \sigma/z) = -\sigma/(2z^2) - \sigma^2/(3z^3) - \dots$$

Thus $|w(z) + 1/z| \leq (1/|z|)\frac{|\sigma/z|}{1-|\sigma/z|}$, which is less than $|\sigma|/2$ if e.g. $|\sigma| < 1$ and $R > 2$. □

One of the goals of the study of perturbations of an analytic map with a parabolic fixed point is to establish a coordinate change in which the map becomes close to a translation of the complex plane. In the family of polynomials P_λ the coordinates w have this property, which can be seen from the following:

Proposition 2. *Small perturbations of the polynomial P_{λ_0} with a parabolic fixed point are close to the translation $w \mapsto w + 1$ in $\mathcal{C}^1$-topology.*

Proof. In the unwrapping coordinate $z \mapsto -1/qz^q$ the unperturbed map is close to the translation. By Proposition 1, the coordinate changes w are close to $z \mapsto -1/qz^q$, hence small perturbations of P_λ are also close to $w \mapsto w + 1$. □

This is quite a straightforward proof and it allows one to avoid estimates that were proved separately as Proposition 3.1 in [22]. Moreover, our Theorem 6 also points to a relation between the unwrapping coordinate w and the unwrapping coordinate in [21], which was defined by $\int_{z_0}^{z}(1/(f(\zeta) - \zeta)d\zeta$. Only through our approach does it become clear that the expression in the denominator is replaced just by its Weierstrass polynomial, so the two different kind of coordinates have in fact similar behavior.

Acknowledgements This article was written when the second named author was a Robert D. Adams Visiting Assistant Professor in the Department of Mathematics of the University of Kansas. She thanks for 3 years of being supported in this position. She also thanks the Mathematische Forschunginstitut Oberwolfach for an opportunity to participate in Arbeitsgemeinschaft "Julia sets of positive measure" in April 2008. Particular thanks go to the coordinators, Xavier Buff and Arnaud Chéritat, who inspired her interest in parabolic explosions.

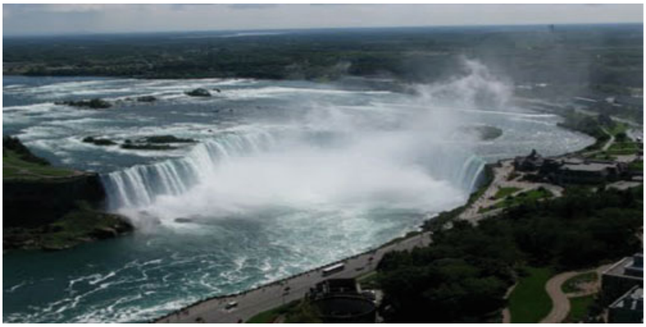
References

1. M. Abate, Discrete holomorphic local dynamical systems. Notes of the CIME course given in Cetraro (Italy) in July 2008, in *Holomorphic Dynamics*, ed. by G. Gentili, J. Guenot, G. Patrizio. Lecture Notes in Math (Springer, Berlin, 2010), pp. 1–55.
2. A. Avila, X. Buff, A. Chéritat, Siegel disks with smooth boundaries, Acta Math. **193**, 1–30 (2004)
3. X. Buff, A. Chéritat, Upper bound for the size of quadratic Siegel disks. Invent. Math. **156**(1), 1–24 (2004)
4. X. Buff, A. Chéritat, *Quadratic Julia Sets with Positive Area*, preprint, arXiv:math/0605514
5. X. Buff, A. Chéritat, *Arbeitsgemeinschaft "Julia sets of positive measure"*, Mathematische Forschunginstitut Oberwolfach, Report No. 17/2008

6. L. Block, D. Hart, The bifurcation of periodic orbits of one-dimensional maps. Ergod. Theory Dyn. Syst. **2**(2), 125–129 (1982)
7. F. Bracci, Local dynamics of holomorphic diffeomorphisms. Boll. Unione Mat. Ital. Sez. B Artic. Ric. Mat. 8 **7**(3), 609–636 (2004)
8. L. Carleson, T.W. Gamelin, *Complex Dynamics*. Universitext: Tracts in Mathematics (Springer, New York, 1993)
9. A. Chéritat, Recherche d'ensembles de Julia de mesure de Lebesgue positive, Thèse, Orsay, décembre 2001, available at: http://www.math.univ-toulouse.fr/~cheritat/publi2.php
10. P. Cvitanoviè, J. Myrheim, Universality for period n-tuplings in complex mappings, Phys. Lett. A **94**(8), 329–333 (1983)
11. A. Douady, Does a Julia set depend continuously on the polynomial? in *Complex Dynamical Systems*. Proceedings of Symposia in Applied Mathematics, Cincinnati, vol. 49 (American Mathematical Society, Providence, 1994), pp. 91–138.
12. A. Douady, J.H. Hubbard, *Étude Dynamique des Polynômes Complexes* (Publications Mathématiques, Orsay), 84–92 (1984); 85–94 (1985), available at: http://www.math.cornell.edu/~hubbard/
13. J. Guckenheimer, P.H. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Applied Mathematical Sciences, vol. 42, (Springer, New York, 1990), Revised and corrected reprint of the 1983 original.
14. V.A. Gromov, Some solutions of a scalar equation with a vector parameter. Sibirsk. Mat. Zh. **32**(5) 179–181, 210 (1991); translation in Siberian Math. J. **32**(5), 882–883 (1991), (1992) (Russian)
15. A.I. Gol'berg, Y.G. Sinaĭ, K.M. Khanin, Universal properties of sequences of period-tripling bifurcations, Uspekhi Mat. Nauk. **38**(1), 159–160 (1983) (Russian)
16. A. Hefez, Irreducible plane curve singularities. *Real and Complex Singularities*. Lecture Notes in Pure and Applied Mathamatics, vol. 232 (Dekker, New York, 2003) 1–120
17. K. Kodaira, *Complex Manifolds and Deformation of Complex Structures*. Classics in Mathematics. (Springer, Berlin, 2005), Translated from the 1981 Japanese original by Kazuo Akao. Reprint of the 1986 English edition.
18. J. Milnor, *Dynamics in one Complex Variable,* 3rd edn. Annals of Mathematics Studies, vol. 160 (Princeton University Press, Princeton, 2006)
19. P. Mardešiè, R. Roussarie, C. Rousseau, Modulus of analytic classification for unfoldings of generic parabolic diffeomorphisms. Mosc. Math. J. **4**(2), 455–502 (2004)
20. T. Needham, *Visual Complex Analysis* (The Clarendon Press/Oxford University Press, New York, 1997)
21. R. Oudkerk, The parabolic implosion for $f_0(z) = z + z^{\nu+1} + \mathcal{O}(z^{\nu+2})$, Ph.D. thesis, Warwick, 1999, available at: http://www.math.sunysb.edu/dynamics/theses/index.html
22. C. Rousseau, C. Christopher, Modulus of analytic classification for the generic unfolding of a codimension 1 resonant diffeomorphism or resonant saddle. Ann. Inst. Fourier (Grenoble) **57**(1), 301–360 (2007)
23. M. Shishikura, The Hausdorff dimension of the boundary of the Mandelbrot set and Julia sets. Ann. of Math. 2 **147**(2), 225–267 (1998)
24. M. Shishikura, Bifurcation of parabolic fixed points, in *The Mandelbrot Set, Theme and Variations*, London Mathematical Society Lecture Note Series, vol. 274 (Cambridge University Press, Cambridge, 2000) pp. 325–363
25. M. Stawiska, Parabolic explosions via Puiseux theorem, in [5], pp. 14–16
26. T.N. Subramaniam, D.E.G. Malm, How to integrate rational functions. Am. Math. Mon. **99**(8), 762–772 (1992)
27. G.Y. Zhang, A simple proof of a theorem of block and hart. Am. Math. Mon. **107**(8), 751 (2000)

Super Stable Kählerian Horseshoe?

M. Gromov



Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract Following the prints of Smale's horseshoe, we trace the problems originated from the interface between hyperbolic stability and the Abel-Jacobi-Albanese construction.

M. Gromov (✉)
IHES, Bur-sur-Yvette, France

CIMS, New York, USA
e-mail: gromov@ihes.fr

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_8, © Springer-Verlag Berlin Heidelberg 2012

1 Abelianization, Super-Stability and Universality

Let V be a compact manifold (or, more generally, a compact locally contractible space), denote by $\tilde{A}$ the real 1-homology $H_1(V;\mathbb{R})$; thus, $\tilde{A}$ is an $\mathbb{R}$-vector space of finite dimension, say of $N = rank(H_1(V;\mathbb{R}))$, and let $A = A(V)$ be the flat 1-homology (Abel-Jacobi-Albanese) torus $A = H_1(V;\mathbb{R})/H_1(V;\mathbb{Z})$.

Strictly speaking, we factorize not by $H_1(V;\mathbb{Z})$ but by $H_1(V;\mathbb{Z})/torsion$, and we denote by h^{Ab} be the canonical (Abel's) homomorphism from the homology of V to that of A, i.e. $h^{Ab} : H_1(V;\mathbb{Z}) \to H_1(V;\mathbb{Z}) = \mathbb{Z}^N$, where h^{Ab} is an *isomorphism* from $H_1(V;\mathbb{Z})/torsion$ onto $H_1(A;\mathbb{Z})$.

Since the universal covering $\tilde{A} = H_1(V;\mathbb{R})$ of A is *contractible*, and since the fundamental group $\pi_1(A) = \mathbb{Z}^N = H_1(V;\mathbb{Z})/torsion$ is *Abelian*, there exists a *unique (Abel's) homotopy class* $[f]^{Ab}$ of continuous maps $f : V \to A$, such that the induced homology homomorphism $[f]^{Ab}_{*1} : H_1(V;\mathbb{Z}) \to H_1(A;\mathbb{Z})$ equals h^{Ab}.

There are two remarkable instances of classes Σ of "geometric structures" with the following property.

For every structure $\sigma \in \Sigma$ on an arbitrary V, there exists a unique Σ-structure on $A = A(V)$, say $\sigma_A = \sigma_A(\sigma)$, and an *essentially unique* σ/σ_A*-compatible map* $f_\sigma : V \to A$ in the class $[f]^{Ab}$, (where, moreover, σ_A "commutes" with the group translations in A in the examples **A** and **B** below).

A. Holomorphic Abelianization Theorem. Let σ be a complex structure on V, such that (V,σ) is *Kähler*. e.g. complex *algebraic*. Then A admits a *unique translation invariant complex structure* σ_A, such that the homotopy class $[f]^{Ab}$ contains a *holomorphic* map $f_\sigma : (V,\sigma) \to (A,\sigma_A)$, where this f_σ is *unique up-to* A*-translations*.

B. Dynamical Superstability/Universality Theorem. Let $\sigma : V \to V$ be a self-homeomorphism, such that the induced homology automorphism

$$\sigma_{*1} : H_1(V;\mathbb{R}) = \mathbb{R}^N \to \mathbb{R}^N = H_1(V;\mathbb{R})$$

is *hyperbolic*, i.e. all (real and complex) eigenvalues λ of σ_{*1} satisfy $|\lambda| \neq 1$. Then the torus $A = A(V)$ admits a *unique (continuous group) automorphism* $\sigma_A : A \to A$, such that the class $[f]^{Ab}$ contains a *continuous* (typically, *non-smooth* even for real analytic σ) *map* $f_\sigma : (V,\sigma) \to (A,\sigma_A)$ *commuting with the two* σ which means the commutativity of the diagram

$$\overset{\sigma}{\circlearrowleft} V \overset{f_\sigma}{\to} A \overset{\sigma_A}{\circlearrowright}, \text{ that is } f_\sigma \circ \sigma = \sigma_A \circ f_\sigma,$$

where this f_σ is *unique, up-to translations by the (finite) subgroup* $fix(\sigma_A) \subset A$ *of the fixed points of* σ_A.

The complex torus (A,σ_A) in **A** is called the *Albanese variety*; it generalizes the classical Jacobian of an algebraic curve.

If $dim_{\mathbb{R}}(V) = 4$ and $N = rank(H_1(V))$ is *even*, then, Kodaira proved **A** *without* the assumption of V being Kähler. (Later on, these V were shown to be Kähler anyway as was was pointed to me by Domingo Toledo.) On the other hand, this fails to be true in the non-Kähler case for $dim_{\mathbb{R}}(V) \geq 6$.

Theorem **B** is due to John Franks [21]; it was preceded by a similar result by Michael Shub [76] for *expanding* (rather than hyperbolic) endomorphisms. The idea goes back to *Smale's horse-shoe* [79] – the first example of a *structurally stable* diffeomorphism with uncountable closure of the set of periodic points. Later on, Smale announced the stability of hyperbolic automorphisms of the 2-torus but his proof remained unpublished. The first accepted proof of the stability of *locally split hyperbolic* diffeomorphisms is due to Anosov [2].

C. Universality Problem. Let Σ be a class of "geometric structures" and Γ be a group, possibly with additional data expressible in the group theoretic terms. Construct a space $\tilde{A} = \tilde{A}(\Gamma, \Sigma)$ with a geometric structure $\sigma_{\tilde{A}} = \sigma_{\tilde{A}}(\Gamma) \in \Sigma$, or a "canonical family" of such $(\tilde{A}, \sigma_{\tilde{A}})$, with the following properties.

- *Γ-Invariance*. The space A is acted upon by Γ where this action is compatible with $\sigma_{\tilde{A}}$.

•• *(Γ, Σ)-Universality*. Let $\tilde{V}$ be a space with a structure $\tilde{\sigma} \in \Sigma$ and with a Γ-action on $\tilde{V}$ which is compatible with $\tilde{\sigma}$ and let $\tilde{f} : \tilde{V} \to \tilde{A}$ be a Γ-equivariant map. Then $\tilde{f}$ is Γ equivariantly homotopic to an essentially unique Γ-equivariant map $\tilde{f}_\sigma : \tilde{V} \to \tilde{A}$ which is compatible with the two Σ-structures: $\tilde{\sigma}$ on $\tilde{V}$ and $\sigma_{\tilde{A}}$ on $\tilde{A}$.

In examples **A** and **B**, the group Γ is isomorphic to $\mathbb{Z}^N = H_1(V)/torsion$ and $\tilde{A} = H_1(V;\mathbb{R}) = \mathbb{Z}^N \otimes \mathbb{R} = \mathbb{R}^N$, while $\tilde{V}$ is the maximal covering of V with a free Abelian Galois group, where this Galois group equals our Γ, (isomorphic to $\mathbb{Z}^N$ in the present case) such that $\tilde{V}/\Gamma = V$ and $\tilde{f}_\sigma : \tilde{V} \to \tilde{A}$ equals a lift of $f_\sigma : V \to A$ to $\tilde{V}$.

In the Albanese case, Σ is a subclass of complex analytic structures. These structures on $\tilde{A}$ are translation invariant. They make a family parametrized by $GL_N(\mathbb{R})/GL_{N/2}(\mathbb{C})$, where N is necessarily even.

In the Franks case, the structures $\sigma_{\tilde{A}}$ on $\tilde{A}$ are hyperbolic automorphisms of $\tilde{A}$, i.e. hyperbolic linear self-maps of $\mathbb{R}^N = \tilde{A}$, while $\tilde{\sigma}$ are lifts of self-homeomorphisms of $V = \tilde{V}/\mathbb{Z}^N$ to $\tilde{V}$.

What we want to is a characterization of groups Γ and of classes Σ, where our problem would be solvable and where the apparent similarity between **A** and **B** would be embodied into a general functorial framework. We expose below the basic ideas underlying **A** and **B** in the hope they would direct one toward such a general theory.

2 Symbols of Shadows

The construction of Franks' map

$$f_\sigma : V \to A = H_1(V;\mathbb{R})/H_1(V;\mathbb{Z})$$

satisfying the equation $f_\sigma \circ \sigma = \sigma_A \circ f_\sigma$ ultimately depends on *Markov (symbolic) shadowing of quasi-orbits of group actions* (see Sects. 2.3 and 2.4), but we start with a slightly different argument due to Shub and Franks with no explicit mentioning of symbols and shadows.

2.1 *Contraction, Expansion, Split Hyperbolicity and Fixed Points*

Recall that $\sigma_A : A \to A$ corresponds to the action of the self-homeomorphism σ of V on $H_1(V;\mathbb{R})$.

Since $\sigma_A : A \to A$ is invertible, the equation $f \circ \sigma = \sigma_A \circ f$ can be rewritten as the fixed point condition in the space F of maps $f : V \to A$,

$$\sigma^\bullet(f) = f \text{ for } \sigma^\bullet(f) =_{def} \sigma_A^{-1} \circ f \circ \sigma.$$

This equation is solved by using the standard (and obvious)

***U**nique **f**ixed **p**oint property* for ***u**niformly **e**ventulally contracting* self-maps of complete metric spaces, where: a self map φ of metric space X is called ***ue** contracting* if there exists a locally bounded function $i(d) = i_\varphi(d)$, $0 < d < \infty$, such that $diam(\varphi^i(U)) \leq \frac{1}{2}diam(\varphi(U))$ for all subset $U \subset X$ and all $i \geq i_\varphi(diam(U))$.

Usually, this property is formulated for contracting maps; the advantage of "virtually" is a low sensitivity to the metric involved.

Namely, define the *expansion* (control) function, $e(d) = e_f(d)$, $d \geq 0$ of a map f between metric spaces, say $f : X \to Y$ by

$$e(d) = \sup_{dist_X(x_1,x_2)\leq d} dist_Y(f(x_1), f(x_2)),$$

and say that an f has *controlled expansion* or, just, that f *is controlled* if

f is *controlled at infinity*, that means $e_f(d)$ is a *locally bounded* function, and f is *uniformely continous*, that is $\lim_{d\to 0} e(d) = 0$.

Clearly, the **ue**-contraction condition is invariant under controlled homeomorphisms between metric spaces.

*Local **UFP***. The unique fixed point property remains valid for *partially defined* (uniformly eventually) contracting maps $\varphi : U \to X$, where $U \subset X$ is a ρ-ball, $\rho > 0$, around some point $x_0 \in X$. Namely, *There exists an* $\varepsilon_0 = \varepsilon_0(i(d), \rho) > 0$, *such that if* x_0 *is* ε-*fixed by* φ, *i.e.* $dist_X(x, \varphi(x_0)) \leq \varepsilon \leq \varepsilon_0$ *then* x_0 *is accompanied by a unique fixed point* $x_\bullet \in U$, *where* $dist_X(x_0, x_\bullet) \leq \delta = \delta(\varepsilon) \to 0$ *for* $\varepsilon \to 0$.

In fact, if $\varepsilon > 0$ is small enough, then the *forward orbit*

$$x_0,\ \varphi(x_0),\ \varphi^2(x_0) = \varphi \circ \varphi(x_0),\ \varphi^3(x_0) = \varphi \circ \varphi^2(x_0), \ldots, \varphi^i(x_0), \ldots$$

is defined for all $i = 1, 2, \ldots$, and the limit $x_\bullet = \lim_{i\to\infty} \varphi^i(x_0) \in U$ is fixed by φ.

Given a set X and a metric space Y, define $DIST = DIST_F = \sup_X dist_Y$ in the space F of maps $f : X \to Y$ by

$$DIST(f_1, f_2) = \sup_{x\in X} dist_Y(f_1(x), f_2(x)).$$

This $DIST$ is *not* a true metric in F since it may be infinite for infinite X and unbounded Y, but it *is* a true metric on every *DIST-finiteness component* F, that is a maximal subset where $DIST < \infty$. Accordingly, the usual "metric language" applies to these components.

For example, we say that F is *complete* if every its finiteness component is complete and observe that if Y is complete, then the space F is complete. Furthermore, given a metric in X, the subspaces of continuous, of uniformly as well as of controlled maps are also complete.

Call an invertible self-map $\varphi : X \to X$ **ue** *expanding* if the reciprocal map $\varphi^{-1} : X \to X$ is **ue** contracting.

Observe that **ue** *expanding* maps φ have (locally as well as globally) the **ufp** property which follows from that for φ^{-1}.

Call a self-map $\varphi : X \to X$ *split hyperbolic* if (X, φ) topologically decomposes into a Cartesian product,

$$(X, \varphi) = (X^+, \varphi^+) \times (X^-, \varphi^-),$$

where the spaces $X^\pm$ admit metrics, say $dist^+$ and $dist^-$, such that the map $\varphi^+ : X^+ \to X^+$ is **ue** expanding while $\varphi^- : X^- \to X^-$ is **ue** contracting.

The **ufp**-property of the **ue** *contracting* self-maps $(\varphi^+)^{-1}$ and φ^- implies that

split hyperbolic maps, of complete metric spaces enjoy the ***ufp****-property: every such map has a unique fixed point, provided the metric spaces* $(X^\pm, dist^\pm)$ *are complete.*

Notice that this fixed point is of somewhat different nature than that for contracting map: if $\varphi : X \to X$ is contracting, then the forward orbit of every point $x \in X$ converges to the fixed point, but in the (nontrivially) split hypebolic case, the forward and backward orbits of almost all points go to infinity.

Another relevant and (equally obvious) property of split hyperbolicity is that it is inherited by spaces $F = X^Y$ of maps of an arbitrary Y to X.

Indeed, if $\varphi : X \to X$ is a **ue** contracting self-map, then the corresponding self-map, say $\varphi_F : F \to F$, on the space F of maps $f : Y \to X$ with the "metric" DIST is also **ue** contracting and the same is true for **ue** expanding maps.

Therefore,

if X is complete, and $\varphi : X \to X$ is either ***ue*** *contracting or* ***ue*** *expanding, then every φ_F-invariant DIST-finiteness component $F_0 \subset F$ has a unique fixed point $f_0 \in F_0$ of the map $\varphi_F : F \to F$*

If φ is split hyperbolic for $X = (X^+, dist^+) \times (X^-, dist^-)$ then, obviously, φ_F is split hyperbolic for $F = (F^+, DIST^+) \times (F^-, DIST^-)$ where $F^{\pm} = (X^{\pm})^Y$ with $DIST^{\pm}$ in $F^{\pm}$ corresponding to $dist^{\pm}$ in $X^{\pm}$. Moreover, this remains true for subspaces of continuous, uniformly continuous and controlled maps in F.

Warning. Usually, the space X comes with its own metric $dist_X$, but we can not claim that every φ_F-invariant DIST-finiteness component of F for *DIST* associated to $dist_X$ has a fixed point, since our hyperbolic splitting $X = X^+ \times X^-$ is *not* necessarily a *metric* splitting.

However,

if the two projections $(X, dist_X) \to (X^{\pm}, dist^{\pm})$, are controlled at infinity, then they preserve the finiteness components of DIST; therefore, every such φ_F-invariant component $F_0 \subset F$ has a ***u****nique* ***f****ixed* ***p****oint of the map $\varphi_F : F \to F$, provided the metric spaces $(X^+, dist^+)$ and $(X^-, dist^-)$ are complete.*

Remark. In what follows, our self-map $F \to F$ for $F = X^Y$ depends on a given self-map $Y \to Y$ as well as on $X \to X$, but F inherits the hyperbolic splitting from X anyway and the above **ufp** property for the self maps of the *DIST*-finiteness components of F remained valid.

Proof of Franks' Super-stability for Self-homomorphisms $\sigma : V \to V$. Let $\tilde{V} \to V$ denote the Abelian covering of V that is induced by Abel's map $V \to A$ from the covering map $H_1(V;\mathbb{R}) = \tilde{A} \to A = H_1(V;\mathbb{R})/H_1(V;\mathbb{Z})$, i.e. the Galois group Γ of this covering equals $H_1(V;\mathbb{Z})/torsion = \mathbb{Z}^N$.

Denote by $\tilde{F}$ the space of continuous maps $\tilde{V} \to \tilde{A}$, let

$$\tilde{\sigma}^{\bullet}(\tilde{f}) = \sigma_{\tilde{A}}^{-1} \circ \tilde{f} \circ \tilde{\sigma},$$

where $\sigma_{\tilde{A}} : \tilde{A} \to \tilde{A}$ is the homology automorphism $\sigma_{*1} : H_1(V;\mathbb{R}) = \tilde{A} \to \tilde{A} = H_1(V;\mathbb{R})$ and $\tilde{\sigma}$ is a lift of $\sigma : V \to V$ to the covering $\tilde{V}$ of V.

We look for an $\tilde{f} \in \tilde{F}$ which satisfies the equation

$$\tilde{\sigma}^{\bullet}(\tilde{f}) = \tilde{f},$$

where, moreover, this $\tilde{f} : \tilde{V} \to \tilde{A}$ must be *equivariant* for the (Galois) actions of $\Gamma = H_1(X)/torsion$ on $\tilde{V}$ and $\tilde{A}$, i.e. $\tilde{f}$ must be a lift of some $f : V \to V$ from our homotopy class $[f]^{Ab}$ of maps $V \to A$.

Start by observing the following obvious.

1. *Quasi-morphism Property.* Since V is compact, the lift $\tilde{f} : \tilde{V} \to \tilde{A}$ of *every continuous* map $f : V \to A$ in the class $[f]^{Ab}$ *almost commutes* with the two lifted σ in the sense that the diagram

$$\overset{\tilde\sigma}{\circlearrowright}\ \tilde V \xrightarrow{\tilde f} \tilde A\ \overset{\sigma_{\tilde A}}{\circlearrowleft}$$

commutes *up-to a bounded error*. This means

$$\mathit{DIST}(\sigma_{\tilde A}\circ \tilde f, \tilde f\circ\tilde\sigma) =_{def} \sup_{\tilde v\in\tilde V} \mathit{dist}_{\tilde A}(\sigma_{\tilde A}\circ \tilde f(\tilde v), \tilde f\circ\tilde\sigma(\tilde x)) < \infty,$$

or, equivalently (for $\sigma_{\tilde A}$ being invertible),

$$\mathit{DIST}(\tilde\sigma^\bullet(\tilde f), \tilde f) = \sup_{\tilde v\in\tilde V} \mathit{dist}_{\tilde A}(\sigma_{\tilde A}^{-1}\circ \tilde f\circ\tilde\sigma(\tilde v), \tilde f(\tilde v)) < \infty.$$

In other words,
$\tilde\sigma^\bullet$ sends the DIST-finiteness component of the lift $\tilde f$ of every $f \in [f]^{Ab}$ into itself.
Such $\tilde f : \tilde V \to \tilde A$ are regarded as *quasi-morphisms* in the category of metric spaces with self-mappings where morphisms are maps where the corresponding diagram is truly commutative.

2. *Split Hyperbolicity.* The hyperbolicity of the action $\sigma_{\tilde A} = \sigma_{*1}$ of $\sigma : V \to V$ on the homology $\tilde A = H_1(V;\mathbb{R})$ amounts to the existence of a unique splitting of $\underset{\sigma_{\tilde A}}{\circlearrowright}\tilde A$ into a Cartesian sum of an *expanding* and a *contracting* self-mappings,

$$(\tilde A, \sigma_{\tilde A}) = (\tilde A^+, \sigma_{\tilde A}^+)\times(\tilde A^-, \sigma_{\tilde A}^-),$$

where $\tilde A^+ (= \mathbb{R}^{N+})$ corresponds to the sum of the eigen-spaces of $\sigma_{\tilde A} = \sigma_{*1}$ regarded as a linear operator on $\mathbb{R}^N = H_1(X;\mathbb{R}) = \tilde A$ with the eigen-values λ where $|\lambda| > 1$ and $\tilde A^- (= \mathbb{R}^{N-})$ represents $|\lambda| < 1$.
Since the hyperbolic splitting $\tilde A^+\times\tilde A^- = \mathbb{R}^{N+}\times\mathbb{R}^{N-}$ of $\tilde A = \mathbb{R}^N$ is a *metric* one, the corresponding hyperbolic splitting for $\tilde\sigma^\bullet : \tilde F \to \tilde F$, where $\tilde F$ is the space $\tilde F, \tilde\sigma^\bullet$ of continuous maps $\tilde f : \tilde X \to \tilde A$, is a "metric" splitting for DIST in $\tilde F$,

$$(\tilde F, \tilde\sigma^\bullet) = (\tilde F^+, \tilde\sigma^{\bullet+})\times(\tilde F^-, \tilde\sigma^{\bullet-})$$

where
the self-map $\tilde\sigma^{+\bullet} : \tilde F^+ \to \tilde F^+$ is *contracting* with respect to $\mathit{DIST}_{\tilde F+}$, that is $\sup_{\tilde x\in\tilde X} \mathit{dist}_{\tilde A+}$, due to the contraction by $\sigma_{\tilde A+}^{-1}$ in the decomposition $\sigma_{\tilde A+}^{-1}\circ\tilde f\circ\tilde\sigma = \tilde\sigma^{+\bullet}(\tilde f)$, while $\tilde\sigma^{\bullet-} : \tilde F^- \to \tilde F^-$ is *expanding* for $\mathit{DIST}_{\tilde F-}$, since $\tilde\sigma$ *invertible.*
In other words
the self-map $\tilde\sigma^\bullet : \tilde F \to: \tilde F$ is split hyperbolic.
Therefore,

the DIST-finiteness component $F_0 \subset F$ of every map $\tilde{f}_0 : \tilde{V} \to \tilde{A}$ lifted from a map $V \to A$ contains a unique fixed point, say $\tilde{f}_\bullet \in F_0$ that is a map $\tilde{f}_\bullet : \tilde{V} \to \tilde{A}$, such that DIST$(\tilde{f}_\bullet, \tilde{f}_0) < \infty$.

To conclude the proof of Franks' theorem let us show that $\tilde{f}_\bullet$ equals a lift of some map $V \to A$, where, observe, the "lifted" maps $\tilde{f} : \tilde{V} \to \tilde{A}$ are exactly the *Γ-equivariant* ones for the group $\Gamma = H_1(V)/torsion = \mathbb{Z}^N$ which acts on $\tilde{V}$ and on $\tilde{A}$.

Define the action of Γ on $\tilde{F}$ by $\tilde{f} \overset{\gamma}{\to} \gamma_{\tilde{F}}(\tilde{f}) = \gamma_{\tilde{A}}^{-1} \circ \tilde{f} \circ \gamma_{\tilde{V}}$, where $\gamma_{\tilde{A}}$ and $\gamma_{\tilde{X}}$ denote the γ-transformation of $\tilde{V}$ and $\tilde{A}$ for all $\gamma \in \Gamma$. Thus,

the fixed point set fix_Γ of this action equals the subset $\tilde{E}_\Gamma \subset \tilde{F}$, of equivariant maps, where, moreover, this $\tilde{E}_\Gamma$ is contained in a single DIST-finiteness component of $\tilde{F}$.

The actions of Γ on $\tilde{F}$ and $\tilde{\sigma}^\bullet : \tilde{F} \to \tilde{F}$ define an action of the normal extension $\Gamma' \supset \Gamma$ generated by translation of Γ on itself together with the automorhism $\sigma_\Gamma : \Gamma \to \Gamma$ induced by $\sigma : X \to X$ on $\Gamma = H_1(X)/torsion.$

Since $\Gamma \subset \Gamma'$ is normalized by $\tilde{\sigma}^\bullet$, the fixed point set $fix_\Gamma \subset \tilde{F}$ is $\tilde{\sigma}^\bullet$-invariant; hence, the *unique* $\tilde{\sigma}$-fixed point $\tilde{f}_\bullet \in \tilde{F}$ is contained in $\tilde{E}_\Gamma = fix_\Gamma$. □

This is not especially surprising since *globally* split hyperbolic self-maps are rather primitive dynamical creatures and the above proof might appear a pure tautology.

On the other hand, Franks' theorem easily implies (see below) that hyperbolic automorphisms σ_A of tori $A = \mathbb{T}^N$ are *structurally C^1-stable.*

This may appear paradoxical, since these σ_A are topologically and measure theoretically *ergodic*: naively intuitively, ergodicity and stability seem incompatible. (The idea that such σ could be structurally stable goes back to Thom and Smale, but a realization of this idea has undergone a few unsuccessful attempts at the proof by several great mathematicians.)

To derive structural stability from super-stability all one needs to show is that if a $\sigma : A \to A$ is *C^1-close* to σ_A then Franks' morphism $f_\sigma : (A, \sigma) \to (A, \sigma_A)$ is *one-to-one*.

But σ_A, and, therefore, *every σ which is sufficiently C^1-close to σ_A*, satisfy the following obvious

Infinitesimal Expansiveness Property. There as an integer $k = k(\sigma_A) \geq 0$ such the norms of the differentials of the iterated maps σ^i, $i = -k, \ldots, -1, 0, 1, \ldots, k$, satisfy

$$\sup_{|i| \leq k} ||D_{\sigma^i}(\tau)|| \geq (1 + \varepsilon)||\tau||$$

for some $\varepsilon = \varepsilon(\sigma_A) > 0$ and all tangent vectors τ of A.

It follows by the implicit function theorem that σ is *locally expansive*: the $sup_\mathbb{Z}$-distance in A between every two distinct σ-orbits $\mathbb{Z} \to A$ is $\geq \varepsilon' > 0$; hence, the map f_σ, being $\mathbb{Z}$-equivariant and C^0-close to the identity map $A \to A$, is, necessarily, locally one-to-one.

Since A is a *closed* manifold and f_σ is homotopic to a *homeomorphism*, that is to σ_A, it *is* globally one-to-one. □

Similar stability theorems for general group actions often go under the heading of "rigidity" see [18] borrowing from the fame of the Mostow-Margulis-Zimmer (super)rigidity theory for semisimple groups, while what we call "*super*-stable" is called "*semi*-stable" by people in dynamics, who, apparently, are disgruntled with non-injectivity, rather than being excited by uniqueness and universality.

2.2 Shadowing Lemmas

The idea underlying the above argument, due to Smale, Anosov, Tate, Shub and Franks, can be vaguely formulated as follows:

if a group (sometimes a semigroup) Σ of self-maps σ of a metric space X "strongly contracts/expands in many directions" then every Σ-quasi-orbit in X is shadowed by an orbit. Moreover, the space of this shadows is "small", i.e. it consists of a single orbit.

Recall that *orbits* of an action of Σ on X are Σ-equivariant maps $o : \Sigma \to X$, say, for the left action of Σ on itself, i.e. $o(\sigma \cdot \sigma') = \sigma_X(o(\sigma'))$, where $\sigma_X : X \to X$ denotes the action of $\sigma \in \Sigma$ on X.

These orbits are the same as the *fixed points* of the $\bullet$-*action* of Σ on the space X^Σ of all maps $q : \Sigma \to X$, defined by

$$\sigma^\bullet(q) = \sigma_X^{-1} \circ q \circ \sigma.$$

The deviation of a general $q \in X^\Sigma$ from being an orbit can be measured, for example, with a given (usually, finite generating) subset $\Theta \subset \Sigma$, by

$$DI_\Theta(q) =_{def} \sup_{\sigma \in \Theta} DIST(\sigma(q), q),$$

where, recall, $DIST$ in X^Σ is defined as $\sup_\Sigma dist_X$. (One can use various "weighted versions" of $DIST$ which lead to somewhat different "quasi"s and/or "shadows".)

Call $q : \Sigma \to X$ a *quasi-orbit* if $DI_\Theta(q) < \infty$; more generally, ε-*orbit* signifies $DI_\Theta(q) < \varepsilon$ (where this $\varepsilon \geq 0$ may be large in the present context).

In other words quasi-orbits of an action of Σ on X are the same as *quasi-fixed points* for the corresponding $\bullet$-action of the group Σ on X^Σ.

An orbit o is said to δ-*shadow* a q if $DIST(o, q) < \delta$, where the plain "shadow" refers to $\delta = \infty$.

The essence of Franks' argument is the following

Shadowing for Split Hyperbolic Actions of $\Gamma = \mathbb{Z}$: Every quasi-orbit of such an action on a complete metric space is shadowed by a unique orbit.

Indeed split hyperbolicity *passes from X to the corresponding $\bullet$-action on* $X^{\mathbb{Z}}$ and every almost fixed point of the resulting (split hyperbolic!) action is accompanied by a unique fixed point in $X^{\mathbb{Z}}$ that is an orbit of the original action in X.

Moreover, split hyperbolicity implies

Uniform Shadowing. Every ε-orbit q is δ-shadowed by an orbit where δ depends on ε but not on q.

The key property of the hyperbolic splitting $\tilde{A} = \tilde{A}^{+} \times \tilde{A}^{-}$ in Franks' argument (where $\tilde{A}$ is a Euclidean space with a linear hyperbolic self-map $\sigma_{\tilde{A}}$) is that this splitting is *metrically controlled*, i.e. the projections $p^{\pm} : \tilde{A} \to \tilde{A}^{\pm}$ are *conrtolled at infinity*, for the (Euclidean) metrics in $\tilde{A}$ and $\tilde{A}_{\pm}$,

$$dist_{\tilde{A}}(\tilde{a}_1, \tilde{a}_2) \le d < \infty \Rightarrow dist_{\tilde{A}^{\pm}}(p^{\pm}(\tilde{a}_1), p^{\pm}(\tilde{a}_2)) \le e(d) < \infty;$$

hence, *quasi-orbits project to quasi-orbits.*

And on the bottom of all this lies **ufp** – the *unique fixed point* property of (uniformly eventually) *contracting* self-maps of *complete* metric spaces that are the spaces of $\sigma^{\pm}_{\tilde{A}}$-quaisi-orbits in $(\tilde{A}^{\pm})^{\mathbb{Z}}$ in Franks' case.

Question. What are the most general *algebraic* properties of a group Σ and *geometric* properties of a metric space X and the action of Σ on X that would guaranty the (unique) fixed point and shadowing property of the action?

For example, which (semi)groups of linear operators in a Banach (e.g. Hilbert) space have the shadowing property?

Ultimately, one looks for criteria that would imply, for example, *Kazhdan's T-property* of Σ:

by definition of T, every *isometric* action of a T-group Σ on a *Hilbert* space has a fixed point, where the non-trivial point is proving T for particular groups Σ, such as $SL_N(\mathbb{Z})$, for $N \ge 3$.

Also Margulis' super rigidity theorem can be viewed as a fixed point theorem for semisimple groups acting on some spaces of "quasi-representations".

In what follows, we only look at "hyperbolic-like" actions with "strong contraction" in certain directions.

Definition of Stable Partitions. A family of self-maps $\{\sigma_i : X \to X\}_{i \in I}$ *eventually contracts* a subset $S \subset X$ if all subsets $S' \subset S$ with $diam_X(S') < \infty$ satisfy

$$diam_X(\sigma_i(S)) \to 0 \text{ for } i \to \infty;$$

this means that, for each $D > 0$, there are at most finitely many $i \in I$, such that $diam_X(\sigma_i(S')) > D$.

Thus, X is *partitioned* into maximal $\{\sigma_i\}$*-stable* subsets, called $\{\sigma_i\}$*-stable slices (leaves)* that are eventually contracted by $\{\sigma_i\}$, where the corresponding quotient space $X/partition$ is denoted $X^{+} = X^{+}_{\{\sigma_i\}}$.

Observe that every uniformly continuous map $\sigma : X \to X$ which *commutes* with all σ_i sends stable subsets to stable ones; thus, σ induces a self-map of X^{+}, say $\sigma^{+} : X^{+} \to X^{+}$. Moreover, this remains true for σ which *eventually commute* with σ_i, i.e.

$$DIST(\sigma \circ \sigma_i, \sigma_i \circ \sigma) \to 0 \text{ for } i \to \infty.$$

An eventual $\{\sigma_i\}$-contraction is called *uniform* if for each $d > 0$ there is a *cofinite* (i.e. with finite complement) subset $I(d) \subset I$, such that every $\{\sigma_i\}$-stable subset $U \subset X$ of diameter $\leq d$ satisfies

$$diam(\sigma_i(U)) \leq \frac{1}{2} diam(U) \text{ for all } i \in I(d).$$

Call $x \in X$ an *eventual fixed point* of σ if

$$dist(\sigma_i(x), x) \to 0 \text{ for } i \to \infty.$$

The set of eventually fixed points is (obviously) *invariant* under every uniformly continuous map $\sigma : X \to X$ which eventually commutes with σ_i.

In fact, this is also true if the set $\{\sigma_i\}$ is *eventually normalized* by σ, i.e. there exists a *proper* map $j : I \to I$, (i.e. the pull-backs of finite set are finite) such that

$$DIST(\sigma \circ \sigma_i, \sigma_{j(i)} \circ \sigma) \to 0 \text{ for } i \to \infty.$$

Let us spice these definitions,with the two (obvious) observations.

Relative FP Property. Let X be a complete metric space and Σ be a semigroup of uniformly continous maps $\sigma : X \to X$ such that Σ admits a family of maps $\sigma_i : X \to X$ which *eventually commute with all* $\sigma \in \Sigma$ and such that

⋆ *The eventual contraction (if any) of the family* $\{\sigma_i\}$ *is uniform;*

⋆⋆ *The induced action of* Σ *on* X^+, *for* $\sigma \mapsto \sigma^+ : X^+ \to X^+$ *has a fixed point* $x^+_\bullet \in X^+$

Then Σ *has a fixed point* $x_\bullet \in X$.

Observe that uniqueness of $x^+_\bullet$ *does not*, in general, imply uniqueness of $x_\bullet$ and that the ordinary **ufp** for contracting maps σ reduces to the above with $\Sigma = \{\sigma_i\} = \{\sigma^i\}_{i>0}$ and X^+ being a single point, where the uniqueness of the fixed point of a *contracting* σ is obvious.

Relative Unique Shadowing Property. Let X, X^- be metric spaces, $p_- : X \to X_-$ be a continuous map controlled at infinity and let a semigroup Σ act on X, where this action preserves the partition of X into the fibers (i.e. pullbacks of points) of p_- and, thus, Σ acts on X_- as well.

Let $\{\sigma_i\}_{i\in I} \subset \Sigma$ be a subset of invertible maps $\sigma_i : X \to X$, such that

- The maps $\sigma_i^{-1} : X \to X$ eventually contract the fibers of p_- and this contraction is uniform;

•• The family $\{\sigma_i^{-1}\}$ eventually commutes with every $\sigma \in \Sigma$.

Let $\Theta \subset \Sigma$ be a generating subset.

If every quasi-orbit q_- *with respect to* Θ *of the action of* Σ *on* X_- *is shadowed by an orbit* o_-, *then also every quasi-orbit* q *in* X *is followed by an orbit* o; *moreover, if* $o_- = o_-(q_-)$ *is unique for every* q_-, *then* $o = o(q)$ *is also unique.*

2.3 Applications of Shadows

Nilpotent Example. Let $\tilde{G}$ be a simply connected nilpotent Lie group and $\sigma_0 : \tilde{G} \to \tilde{G}$ be a *hyperbolic* automorphism which means that the corresponding automorphism $d\sigma_0$ of the Lie algebra $\tilde{g}$ of $\tilde{G}$ is a hyperbolic linear self-map. Then

*every group $\Sigma \ni \sigma_0$ of automorphisms of $\tilde{G}$ commuting with σ_0 has the unique shadowing property. Furthermore, this remains valid (and becomes more obvious) for **semi**groups Σ, provided σ_0 is an expanding map.*

Indeed, since the differential $d\sigma_0$ (of σ_0 at $id \in \tilde{G}$) is a *linear* self-map of the Lie algebra $\tilde{g}$ of $\tilde{G}$ and since the center $\tilde{c} \subset \tilde{g}$ is $d\sigma_0$-invariant, the action of $d\sigma_0$ on $\tilde{c}$ is also hyperbolic.

Let $\tilde{C}^+ \subset \tilde{G}$ be the central subgroup corresponding to the subspace in $\tilde{c}$ with eigenvalues having $|\lambda| > 1$. If all eigenvalues λ of the action of σ on $\tilde{c}$ have $|\lambda| < 1$ and if σ_0 is invertible in Σ (e.g. if Σ is a group), we just replace σ_0 by σ_0^{-1}.

Since the quotient map $\tilde{G} \to \tilde{G}^- = \tilde{G}/\tilde{C}^+$, being a Lie group homomorphism, is controlled at infinity, the relative shadowing property above yields the unique shadowing property in $\tilde{G}$ by induction on $dim(\tilde{G})$,

(One may equally use the induction on the *nilpotency degree*, rather than on dimension; thus, the above applies to *infinite dimensional* nilpotent Lie groups and also to *projective limits* of such groups.)

Remark. The hyperbolic splittings $G = G^+ \times G^-$ are *not*, in general, metrically controlled for "natural" (e.g. left-invariant Riemannian) metrics in G and $G^{\pm}$, since the subgroups $G^{\pm} \subset G$ are not necessarily normal; possibly, the control can be regained with some "unnatural" metrics.

Infra-nilpotent Shub-Franks Super-stability Theorem. Let B be a (possibly non-compact) infra-nil-manifold, that is a quotient of a simply connected nilpotent Lie group $\tilde{G}$ by a group Γ freely and discretely acting on $\tilde{G}$, such that Γ equals an extension of a discrete subgroup $\Gamma_0 \subset \tilde{G}$ by a finite group of automorphisms of $\tilde{G}$ preserving Γ_0.

Let $\sigma_{\tilde{G}} : \tilde{G} \to \tilde{G}$ be an automorphism which descends to a self-map of B with a fixed point $b_\bullet$, say $\sigma_B : B \to B$. Denote by σ_Γ the induced endomorphism of $\Gamma = \pi_1(B, b_\bullet)$.

Let V be a compact locally contractible space with a continuous self-map $\sigma : V \to V$ with a fixed point $v_\bullet$ and let $\sigma_* : \pi_1(V, v_\bullet) \to \pi_1(V, v_\bullet)$ denote the induced endomorphism of the fundamental group.

Let $h : \pi_1(V, v_\bullet) \to \Gamma$ be a homomorphism compatible with the two endomorphisms, i.e. the following diagram commutes

$$\overset{\sigma_*}{\circlearrowright}\ \pi_1(V, v_\bullet) \overset{h}{\to} \Gamma\ \overset{\sigma_\Gamma}{\circlearrowleft},\ \textit{that is } h \circ \sigma_* = \sigma_\Gamma \circ h.$$

In the following two cases there exists a unique continuous σ/σ_B-morphism $f : (V, \sigma) \to (B, \sigma_B)$, such that $f(v_\bullet) = b_\bullet$ and the induced homomorphism $f_* : \pi_1(V, v_\bullet) \to \pi_1(B, b_\bullet) = \Gamma$ equals h.

Shub Case. The self-map $\sigma_{\tilde{G}} : \tilde{G} \to \tilde{G}$ is expanding.

Franks Case. The differential $d\sigma_{\tilde{G}}$ of $\sigma_{\tilde{G}} : \tilde{G} \to \tilde{G}$ at $id \in \tilde{G}$ is hyperbolic and $\sigma : X \to X$ is a homeomorphism.

Proof. Since $\tilde{B} = \tilde{G}$ is contractible, there exists a continuous map $(V, v_\bullet) \to (B, b_\bullet)$ which implements h.

The lift of this map to the respective Γ-coverings $\tilde{V}$ of V and $\tilde{B} = \tilde{G}$ sends $\tilde{\sigma}$-orbits to $\sigma_{\tilde{G}}$-quasi-orbits.

The shadowing orbits of these, which are provided by the nilpotent example, transform this lift to a $\tilde{\sigma}/\sigma_{\tilde{G}}$-morphism.

Since Γ is normalized by $\sigma_{\tilde{G}}$, this morphism is Γ-equivariant. □

Remarks and Questions.

(a) The fixed point $x_\bullet \in X$ of σ is not truly needed as one can always extend σ to a self-map σ' of a path connected $X' \supset X$ *with* a fixed point $x'_\bullet \in X'$.

(b) The contractibility of $\tilde{B}$ and the freedom of the action of Γ are used only for an implementation of $h : \pi_1(X) \to \Gamma$ by an equivariant map $\tilde{f_0} : \tilde{X} \to \tilde{B}$. Granted such an $\tilde{f_0}$, one may drop the freedom of the Γ-action on B. Thus, for example, one obtains intersting self-homeomorphisms of simply connected spaces, such as the sphere $S^2 = \mathbb{T}^2/\pm 1$.

(c) Possibly, infra-nilpotent (B, σ_B) are the only "superstable" compact *topological manifolds*. But, probably, there are interesting "superstable" B with a complicated, e.g. non-locally compact and/or fractal, local topology.

(d) The images of **F**ranks- **A**bel morphisms $(X, \sigma) \to (B, \sigma_B)$ are particular *closed connected σ_B-invariant* subsets in (B, σ_B). What are, for example, these images for the FA maps of closed surfaces X^2 with, say, *pseudo-Anosov* homeomorphisms $X^2 \to X^2$? There are lots of other such closed connected σ_B-invariant subsets, e.g. the "walls" of *Markov partitions.* Can one explicitly describe the "topologically simplest" of them, e.g. locally contractible and/or having equal topological and Hausdorff dimensions, say for hyperbolic automorphisms of the tori $A = \mathbb{T}^N$? Jarek Kwapisz pointed out to me that the description problem for closed invariant subsets for hyperbolic automorphisms of tori was raised by M.Hirsh [45] and that, according to a conjecture attributed to Smale in [3], hyperbolic automorphisms of tori admit no invariant *compact topological submanifolds*, except for unions of subtori.

Can one systematically describe n-dimensional spaces, $n < N$, with self-homeomorphisms, say (X, σ), with a given Franks-Abel Jacobian (A, σ_A), i.e. with $rank(H_1(X)) = N$ and where σ induces a given automorphism of $H_1(X)$?

(e) Some pseudo-Anosov homeomorphisms (X, σ) are obtained from hyperbolic automorphisms σ_0 of 2-tori $\mathbb{T}^2$ by taking ramified coverings X^2 of $\mathbb{T}^2$ with the ramification locus contained in the fixed-point set of σ_0, where the resulting ramified covering $X \to T^2$ equals the corresponding Franks map.

This map is non-injective, but the full Franks-Abel map $X = X^2 \to A = \mathbb{T}^N$, $N = rank(H_1(X^2))$ may seem injective for pseudo-Anosov $\sigma : X^2 \to X^2$ with hyperbolic $\sigma_* : H^1(X^2) \to H^1(X^2)$. In fact, it is shown in [4] that these maps are almost everywhere one-to-one.

However, this I learned from Jarek Kwapisz, certain FA maps may contain horseshoes of double points [3] and, it was recently proven by Jarek Kwapisz with Andy Bouwman, these maps are never injective for genus two surfaces X (where the torus has dimension $4 = 2dim(X)$).

This strikes contrast with the complex analytic Abel-Jacobi maps which *are injective* by Torelli theorem. Yet, one wonders if Franks-Abel maps and invariant subsets of hyperbolic toral automorphisms come as limit sets of "nice complex analytic somethings", similarly to how quasi-circles appear as limit sets of complex analytic actions of discrete groups on the Riemann sphere. (The Formula 2.7 in [4] is indicative of such an "analytic connection" as was pointed out to me by Jarek Kwapisz.)

(f) What are "ramification constructions" for automorphisms of $\mathbb{T}^N$ for $N > 2$, and of infra-nilmanifolds in general?

(Besides unions of codimension two subtori one, probably, can ramify $\mathbb{T}^N$ along "wild" codimension two subsets, such as the images of FA-maps of genus two surfaces in $\mathbb{T}^4$.)

What is a "good natural" class of self-homeomorphisms encompassing Anosov along with $2D$-pseudo-Anosov maps and closed under Cartesian products and ramified coverings?

"Connected Hyperbolicity". Let $\{\Sigma_j\}_{j \in J}$ be a collections of transformation groups of an X and denote by $(E = E_{fix}, J)$ the graph on the vertex set J where the edges $e \in E$ correspond to the pairs of subgroups (Σ_j, Σ_k), $j, k \in J$, such that

The intersection $\Sigma_{jk} = \Sigma_j \cap \Sigma_k$ is *normal* in Σ_j as well as in Σ_k,
The action of Σ_{jk} on X has the **ufp** property.

Similarly, define the graph (E_{shad}, J) with the unique shadowing property in lieu of **ufp**.

It is obvious that

*If the set E_{fix} is non-empty and the graph (E_{fix}, J) is connected, then the action of the group Σ generated by all Σ_j satisfies the **ufp** property.*

Consequently, if the graph (E_{shad}, J) is connected, then Σ has the unique shadowing property.

This applies to certain "sufficiently hyperbolic" homeomorphisms groups, e.g. linear groups Σ with "many" hyperbolic $\sigma \in \Sigma$ but the "connected hyperbolicity", as it stands, has limited applications, since the connectedness often fails for (E_{shad}, J); however, it may be satisfied by a larger graph of "virtual subgroups" in Σ, e.g. corresponding to subgroups in a group $S \supset \Sigma$.

For example, such "virtual connectedness" holds for lattices in semisimple groups of $\mathbb{R}$-rank≥ 2.

Questions.

(A) Let a group Σ act by linear transformations σ of a Banach (e.g. Hilbert) space X, where all (or, at least, "many") $\sigma \neq id$ are split hyperbolic.

Is there a general "virtual connectedness" criterion for the unique shadowing property of such an action in the spirit of Kazhdan's T and/or of the Mostow-Margulis (super)rigidity?

(Mostow's proof of the rigidity of cocompact lattices for $rank_{\mathbb{R}} \geq 2$ depends on connectedness of *Tits' geometries* of the ideal boundaries of the orresponding symmetric spaces, and similar ideas underly "softer" arguments by Kazhdan, Margulis and Zimmer.)

Let Σ be a group of linear transformations of $\mathbb{R}^n$. Then every Σ-quasi-orbit q is shadowed by σ-orbits for hyperbolic $\sigma \in \Sigma$ and Σ acts on the space $H(q)$ of these orbits; yet, it may fail to have a fixed point in there. For example, this typically happens for free non-Abelian groups Σ.

(B) What is the dynamics of the action of Σ on the closure of $H(q)$? When does it admit a *compact* invariant subset?

Suppose Σ is a *word hyperbolic* group and let $\partial_\infty = \partial_\infty(\Sigma)$ denote its *ideal boundary*.

(C) Is there a natural map from $\partial_\infty \times \partial_\infty \setminus diagonal$ to the clousure of $H(q)$?

(D) Are there meaningful examples of the unique shadowing for actions of *non-elementary hyperbolic* groups Σ? Is Kazhdan's T relevant for this?

Let Σ be a group of hyperbolic conformal transformations of the sphere S^n and let q be a quasi-orbit of the corresponding action of Σ on the tangent bundle $T = T(S^n)$ or on some associated bundle, e.g. on $\bigwedge^n T$. Let Q denote the space of Σ-quasiorbits in T.

(E) What are "interesting" invariant subsets (minimal? compact? finite?) of the $\bullet$-action of Σ on Q, e.g. those contained in the closure of the Σ-orbit of a single $q \in Q$? (Recall that the bullet action of a $\sigma \in \Sigma$ on $q : \Sigma \to T$ is given by $\sigma_\bullet(q)(\sigma') = \sigma^{-1} \circ q(\sigma \cdot \sigma') : \sigma' \mapsto T$).

What happens if we enlarge Σ by a group of automorphisms of the bundle $T \to S^n$?

2.4 Combinatorial Reconstruction of Shub-Franks Group Actions

Franks, (Shub) superstability theorem, applied to a hyperbolic automorphism (expanding endomorphism) of compact infra-nil-manifold, $\sigma : B \to B$, shows that the space B and the transformation σ are uniquely determined by purely algebraic data encoded in the automorphism $\overset{\sigma_*}{\circlearrowright} \Gamma = \pi_1(B)$.

Let us describe a *functorial construction* , called *combinatorial reconstruction*

$$\circlearrowright \Gamma \rightsquigarrow \circlearrowright B$$

which is built into the Shub-Franks argument.

Let Γ be a group with a left invariant metric, e.g. coming with a finite generating subset in Γ and let $\Sigma = \Sigma_\Gamma$ be a semigroup of endomorphisms (e.g. group of automophisms) $\sigma = \sigma_\Gamma : \Gamma \to \Gamma$.

Denote by $Q_\varepsilon = Q_\varepsilon(\Gamma, \Sigma, \Theta)$ the space of *ε-orbits* that are maps $q : \Sigma \to \Gamma$, such that

$$DI(q) = DI_\Theta(q) = \sup_{\sigma\in\Theta}\sup_{\sigma'\in\Sigma} dist_\Gamma(\sigma(q(\sigma')), q(\sigma\cdot\sigma')) \leq \varepsilon$$

where Θ is a given generating subset in Σ, which we assume being *finite* in most of what follows.

Endow Q_ε with the topology of point-wise convergence (or, rather, point-wise *stabilization* for Γ discrete) in the space of maps $q : \Sigma \to \Gamma$

Observe that Γ embeds into Q_ε via the orbit map $\gamma \mapsto o = o_\gamma$ for $o_\gamma(\sigma) = \sigma(\gamma)$; thus, Γ acts on Q_ε by $\gamma(q)(\sigma) = o_\gamma(\sigma)\cdot q(\sigma)$ and this action commutes with the action of Σ on Q_ε, that is $\sigma(q(\sigma')) = q(\sigma\sigma')$.

Let

$$\tilde{B}_\varepsilon = \tilde{B}_\varepsilon(\Sigma_\Gamma) = Q_\varepsilon/[DIST_\Gamma < \infty]$$

that is the space of *DIST*-finiteness component of Q_ε (for our $DIST(q_1, q_2) = \sup_\Sigma dist_\Gamma$) with the quotient space topology, and let

$$\tilde{B} = \tilde{B}(\Sigma_\Gamma) = \bigcup_{\varepsilon>0} \tilde{B}_\varepsilon, \text{ and } B = B(\Sigma_\Gamma) = \tilde{B}/\Gamma,$$

where Σ naturally acts on B since the action of Σ on $\tilde{B}$ commutes with the action of Γ.

Say that

Σ_Γ is rigid if there exists a (possibly large, but finite) $\varepsilon = \varepsilon(\Gamma, dist_\Gamma, \Sigma_\Gamma, \Theta) > 0$, such that, for every quasi-orbit q, there is an ε-orbit, say $q_\varepsilon = q_\varepsilon(q)$, such that $DIST_\Gamma(q, q_\varepsilon) < \infty$,

and

Σ_Γ is Divergent if there is a constant $d = d(\Gamma, dist_\Gamma, \Sigma_\Gamma, \Theta, \varepsilon) > 0$, such that the inequality $DIST_\Gamma(q_1, q_2) \leq 2d$ for two ε-orbits, implies that $DIST_\Gamma(q_1, q_2) \leq d$.

This implies that the inequality $DIST_\Gamma(q_1, q_2) \leq d$ is a *transitive* (i.e. equivalence) relation between ε-orbits.

Divergence can be also regarded as a *convexity-type* inequality or a *maximum principle* satisfied by the function

$$di(\sigma) = dist_\Gamma(\sigma(q_1(\sigma)), q_2(\sigma)) :$$

if two ε-orbits, $q_1, q_2 : \Sigma \to \Gamma$, are $2d$-close on a (large) ball in Σ then they are d-close at the center of the ball.

Shadowing $\Rightarrow$ Rigidity. Let $\Gamma = \pi_1(B, b_\bullet)$ for a locally contractible space B and let Σ be induced by a semi-group of self-maps of $(B, b_\bullet)$. If B is *compact*, then, obviously, the (not necessarily unique) *shadowing property of the lifted action of Σ to the universal covering $\tilde{B}$ of B implies rigidity of Σ_Γ.*

Furthermore,

if the shadowing is unique and uniform, the action is divergent as well as rigid.

In particular "connectedly hypebolic" (e.g. cyclic hyperbolic) automorphisms groups Σ of compact infra-nil-manifolds $B = \tilde{G}/\Gamma$ are rigid and divergent. (We did not pay much attention to uniformity of shadowing in these examples, but the proofs of unique shadowing automatically deliver the uniformity as well since hyperbolic actions are *globally expansive*: the sup_Σ distance (i.e. *DIST*) between every two distinct orbits is infinite.

Remarks. (a) The above construction is just a "discrete time" counterpart to the Efremovich-Tikhomirova-Mostow-Margulis description of the ideal boundary of a hyperbolic group via quasi-geodesic rays.

(b) If we drop the rigidity and/or the divergence condition then the resulting space $B = B(\Sigma_\Gamma)$ and $B(\sigma_\Gamma)$ may become non-compact and/or *non-Hausdorff*; yet, such a B may may have non-trivial Hausdorff Σ-equivariant quotient spaces; besides, the geometry of "non-Hausdorffness" of a B may be interesting in its own right.

(c) There are by far (?) more rigid divergent actions then those on infra-nil-manifolds, especially if we allow infinitely generated groups, e.g. $\Gamma = \mathbb{Z}^\infty$, where the construction must incorporate a choice of a suitable metric on such a Γ.

Questions.

(a) Find/classify rigid divergent (semi)groups of automorphisms (endomorphisms) of *finitely generated* (finitely presented?) groups. For example:

Which automorphism groups Σ of $\mathbb{Z}^N$ are rigid?

Are, for instance, "virtually connectedly hyperbolic" Σ rigid? (Much is known in this direction: [18, 20, 53, 60, 73].)

Which automorphisms groups of free groups and of surface groups are rigid?

Are automorphisms of fundamental groups of surfaces induced by pseudo-Anosov maps rigid? (This, probably, follows from [17] and [22].)

Let A be a ramified covering of $\mathbb{T}^N$ with ramification locus being the union of flat codimension two subtori in general position. The fundamental group of such an A often admits hyperbolic-like automorphisms.

Are these ever rigid?

Notice that there are lots of homeomorphisms in the world and typical groups generated by several homeomorphisms are *free*; "intersting" Σ-actions of groups Σ with "many" relations can not be constructed at will – such actions are usually associated with specific geometric structures on corresponding spaces.

Also, there is no systematic way to produce *finitely presented* groups Γ with "large" automorphisms groups Σ. Notice that every Γ acted by Σ is obtained from the free group $F(\Sigma)$ freely generated by $\sigma \in \Sigma$ which is factorized by a

Σ-invariant set $\mathcal{R}$ of relations. If $\mathcal{R}$ is "small", the resulting Γ remains infinitely generated; if $\mathcal{R}$ is too large, Γ becomes trivial. It seems difficult to strike the right balance.

Yet, there is a construction by Rips [71] of "generic" finitely generated (but not finitely presented) groups Γ with a given Σ in their outer automorphism groups.

On the other hand, the quotient groups $F(\Sigma)/\mathcal{R}$ acted upon by Σ look interesting even for infinitely generated Γ.

(b) When does $\tilde{B} \supset \Gamma$ admit a (natural?) group structure extending that of Γ?

(c) The group Γ enters the definition of "rigid" and/or "divergent" only via its (word) metric, but eventually, Γ acts on $\tilde{B}$. Can one incorporate the action of Γ into "rigid/divergent" to start with?

(d) The notion of a quasi-orbit make sense for *partially defined* actions on Γ, such as the obvious partial action of $GL_N(\mathbb{Q})$ on $\mathbb{Z}^N$, e.g. the (contracting) inverse σ^{-1} of an expanding endomorphism $\sigma : \mathbb{Z}^N \to \mathbb{Z}^N$.

Namely, an ε-orbit is a map $q : \Sigma \to \Gamma$ such that for each $\sigma \in \Sigma$ there exists $\gamma = \gamma(q, \sigma) \in \Gamma$ with $dist_\Gamma(\gamma, q(\sigma)) \leq \varepsilon$ such that γ belongs to the domain of definition of all $\theta \in \Theta$, for a given generating subset $\Theta \subset \Sigma$, and such that $dist_\Gamma(\theta(\gamma), q(\theta \cdot \sigma)) \leq \varepsilon$ for all $\theta \in \Theta$.

However, there are no full-fledged orbits and the corresponding action of Γ on Q, is only partially defined.

What kind of spaces B with Σ-dynamics can be constructed for partial actions? (The above contracting σ^{-1} serves as an encouraging example.)

If Σ acts by *injective* homomorphisms defined on subgroups of *finite index* in Γ one defines the (generalized) *semidirect product* $\Gamma \rtimes \Sigma$ that is the set of partial transformations of Γ generated by, say, left translations $\gamma : \Gamma \to \Gamma$ and all partial $\sigma : \Gamma \to \Gamma$.

Probably, the rigidity of the original action, can be adequately addresses in terms of the group $\Gamma^\star = \Gamma \rtimes \Sigma$, as suggested by the construction from the next section. and /or in the spirit of *permutational bimoduli* of Nekrashevych [47, 67], where one works, instead of partial quasi-orbits, with the, say right, action of Σ on the product $\Gamma^* \times \Delta$ and the obvious left action of Γ, for some finite set Δ, where the two actions commute.

(If Δ is the one element set, these correspond, modulo $DIST < \infty$ condition, to the action of Σ on Γ^* by conjugations.)

(e) There is a remarkable class of groups Γ discovered by Grigorchuk, where such a Γ admits subgroups of finite index, say $\Gamma_1, \Gamma_2 \subset \Gamma$, and a *contracting* (in a suitable sence) isomorphism of a Cartesian power of Γ_1 to Γ_2, say $\sigma : \Gamma_1^k \to \Gamma_2$, that is a partially defined contracting "isomorphism" $\sigma : \Gamma^k \to \Gamma$.

(See [47, 67] for a combinatorial (re)construction of a dynamical system associated to this kind of σ.)

Observe that every *k-bracketing* of a finite ordered set I, e.g.

$$\big([(\bullet\bullet)\bullet)](\bullet\bullet)\big)\big(([(\bullet\bullet)(\bullet\bullet)]\bullet)\bullet\big) \text{ where } k = 2 \text{ and } card(I) = 11,$$

provides a partial map $\Gamma^I \to \Gamma$. The totality $\Sigma^\star$ of such maps make what is called an *operad*.

Can one make the full use of the this operad $\Sigma^\star$ generated by σ in Nekrashevych' construction of B with $\Sigma^\star$ "acting on B"?

Is there, in general, a meaningful dynamical theory of operads acting on (compact?) spaces?

(A system of N bracketing on a set I of cardinality N gives one a partial endomorphism $\Gamma^N \to \Gamma^N$ while the full operad comprises a "functorially coherent" family of all such maps for all possible bracketing.)

2.5 Symbolic Dynamics, Markov Coding and Markovian Presentations

Recall the obvious *Bernoulli shift action* of a discrete (semi)group Σ on the space Δ^Σ of maps of Σ to a set Δ, let $\Theta \subset \Sigma$ be a finite subset and take some $M \subset \Delta^\Theta$.

Denote by $C = C_M \subset \Delta^\Sigma$ the pullback of M under the obvious (restriction) map $\Delta^\Sigma \to \Delta^\Delta$ and define the corresponding *Markov (sub)shift* (of finite type) $Q = Q_M \subset \Delta^\Sigma$ as

$$Q = \bigcap_{\sigma \in \Sigma} \sigma(C).$$

Next, observe that $Q \times Q$ is a Markov (sub)shift in $(\Delta \times \Delta)^\Sigma$ and let $R \subset Q \times Q$ be a Markov subshift in this $Q \times Q$.

If this R, regarded as a binary relation on Q, is symmetric and transitive, let $X = Q/R$ be the quotient space of Q by this *equivalence relation* R and observe that the (semi)group Σ naturally acts on this X.

Such an (X, Σ) is called *Markov hyperbolic* (dynamical system), usually for *finite* sets Δ, and the corresponding surjective map $Q \to X = Q/R$ is called (finitery) *Markovian presentation.*

Let us explain why

"rigid Markovian" for a finitely generated (semi)group Σ of automorphisms (endomorphisms) of a finitely generated group Γ implies "rigid hyperbolic" for $(B(\Sigma_\Gamma), \Sigma)$.

Given subsets $\Delta \subset \Gamma$ and $\Theta \subset \Sigma$ take a subset $M \subset \Delta^\Theta$ and define *M-orbits* $q : \Sigma \to \Gamma$ by the condition

$$m(\theta) = (q(\theta \cdot \sigma))^{-1} \cdot \theta(q(\sigma)) \in \Delta \text{ for all } \sigma \in \Sigma \text{ and } \theta \in \Theta$$

and this function $m : \theta \mapsto \delta \in \Delta$ belongs to M for all $\sigma \in \Sigma$. The orbits $o : \Sigma \to \Gamma$ correspond to $\Delta = id \in \Gamma$, i.e. $(o(\theta \cdot \sigma))^{-1} \cdot \theta(o(\sigma)) = id$ and the space $\tilde{Q}_M$ of M-orbits is invariant under the multiplication by orbits, since

$$(o(\sigma \cdot \sigma') \cdot q(\sigma \cdot \sigma'))^{-1} \cdot \sigma(o(\sigma')) \cdot \sigma(q(\sigma')) = (q(\sigma \cdot \sigma'))^{-1} \cdot \sigma(q(\sigma')).$$

(We switched from the notation Q_ε used earlier to $\tilde{Q}_M$ but our ε-orbits are special cases of M-orbits)

Thus every $M \subset \Delta^\Theta$, for given finite subsets $\Theta \subset \Sigma$ and $\Delta \subset \Gamma$, define a Markov Σ-shift $Q_M = \tilde{Q}_M/\Gamma$.

Given an arbitrary finite generating subset $\Theta \subset \Sigma$, a *sufficiently large* finite $\Delta \subset \Gamma$ and $M = \Delta^\Theta$, then the quotient $\tilde{B}_M$ of $\tilde{Q}_M$ by the equivalence relation $\tilde{R} \subset \tilde{B}_M \times \tilde{B}_M$ given by $[DIST_\Gamma < \infty]$ does not depend on Δ: this is our old $\tilde{B}$.

The group Γ (embedded into the space of maps $\Sigma \to \Gamma$ via Σ-orbits as earlier, i.e. by $\gamma \mapsto \sigma(\gamma)$) acts on this R and the quotient $R/\Gamma \subset B_M \times B_M$ is Markov by the Marcovian property of Σ_Γ which, recall, corresponds to the uniformity of shadowing. □

Remarks. (a) "Markov symbolic coding" can be traced to the work by Hadamard (1898) and Morse (1921) on geodesics in hyperbolic surfaces. The above argument is essentially the same as the derivation of Markov property for locally split hyperbolic (Bowen-Anosov) actions of $\mathbb{Z}$ (defined below) on compact spaces from Anosov's *local shadowing lemma* and local expansiveness. This was exploited/refined by Sinai and Bowen in their *Markov partition theory* [5, 77] and then extended to general Markov hyperbolic transformations in [23] (where these were called "finitely presented dynamical systems".)

(b) The major advantage of "Markov hyperbolicity" (originally called just "hyperbolicity" [26]) over "split hyperbolicity" is the applicability of "Markov" to arbitrary (semi)group Σ, not only to $\Sigma = \mathbb{Z}$ and/or $\mathbb{Z}_+$.

However, non-trivial examples of Markov hyperbolic actions of non-cyclic groups Σ in [26] were limited to *word hyperbolic groups* Σ acting on their *ideal boundaries*. (These groups were called "coarse hyperbolic" in [26].)

We shall see in Sect. 3 below further examples that became available due to the recent progress in the *geometric rigidity theory*.

Anosov-Bowen Systems. An action of $\mathbb{Z}$ by uniformly continuous homeomorphisms of a metric space X is called *locally split hyperbolic*, if there exists a $\rho > 0$ such that every ρ-ball in X is contained in a split neighbourhood $U = U^+ \times U^- \subset X$ such that

The fibers of the projection $U \to U^+$ are $\{\sigma^i\}_{i\to+\infty}$-stable, i.e. uniformly eventually contracted by the positive powers of σ, that is the transformation corresponding to $1 \in \mathbb{Z}$,

The fibers of $U \to U^-$ are $\{\sigma^i\}_{i\to-\infty}$-stable,

The projections $U \to U^\pm$ are uniformly continuous with the moduli of continuity independent of U.

Anosov Shadowing Lemma, says that if all metric spaces $U^\pm$ are complete then *there exists an $\varepsilon_0 > 0$, such that every ε-orbit $q : \mathbb{Z} \to X$ with $\varepsilon \le \varepsilon_0$ is $\delta = \delta(\varepsilon)$-shadowed by a unique orbit $o : \mathbb{Z} \to X$ where $\delta = \delta(\varepsilon) \to 0$ for $\varepsilon \to 0$.*

Proof. Let $X^\mathbb{Z}$ be the space of maps with the (possibly infinite) metric $DIST = \sup_\mathbb{Z} dist_X$ and observe that the local split hyperbolicity of an action on X implies that the corresponding $\bullet$-action of $\mathbb{Z}$ on $X^\mathbb{Z}$ is also split hyperbolic.

Thus we need to show that every ε-fixed point of σ is accompanied by a nearby fixed point.

This is done exactly as in the globally split case with a little caveat that the two contracting actions are only *partially* defined and one needs to use the *local* **ufp** property of contracting maps from Sect. 2.

The Markov partition theorem of Sinai-Bowen says, in effect, that

every locally split hyperbolic action of $\mathbb{Z}$ *on a compact space* X *admits a cofinite Markov presentation* $Q \to X$*, i.e. where the cardinalities of the pullbacks of all* $x \in X$ *are bounded by a constant* $< \infty$.

This, as it was shown by [23], remains true for *all Markov hyperbolic* $\mathbb{Z}$-actions.

Question. Which Markov hyperbolic Σ-actions admit cofinite Markov presentations?

This is, apparently, so for the *boundary actions* of the word hyperbolic groups Σ and for many (all?) Markov hyperbolic actions of free groups.

There is nothing of "holomorphic" in all this so far. Of course, endomorphisms $\sigma : \mathbb{T}^N \to \mathbb{T}^N$ analytically extend to holomorphic endomorphisms $\sigma_{\mathbb{C}} : (\mathbb{C}^\times)^N \to (\mathbb{C}^\times)^N$ where $(\mathbb{C}^\times)^N$ are affine *toric* (hence, rational) varieties, where, small rational deformations of $\sigma_{\mathbb{C}}$ may be of some interest.

Questions.

(a) Does $\mathbb{C}P^N$ admit a birational action by the group $SL_{N+2}(\mathbb{Z})$ which does not factor through a finite group?

(b) Does it admits such an action by $SL_{N+1}(\mathbb{Z})$ which is not conjugate to a projective action?
(It is "No" for (b), hence, for (a), see [14].)

Similarly, compact real nil-manifolds G/Γ complexify to $G_{\mathbb{C}}/\Gamma$ where these quotients of complex nilpotent groups $G_{\mathbb{C}}$ are, apparently, algebraic (affine? rational?) as well.

But this is not quite a kind of "holomorphic connection" we are looking for – the first whiff of "Kähler" can be felt, however, in the examples we shall see presently.

3 Inner Rigidity and Markov Coding

A finitely generated subgroup Σ in a finitely generated group Γ is called *inner rigid* if the action of Σ on Γ by inner automorphisms is rigid. In particular, Γ is called inner rigid, if its action on itself by inner automorphisms is rigid.

Let us express this in geometric language and show that actions of discrete rigid groups on homogeneous spaces are Markov.

3.1 Quasi-isometries of Lie Groups and Combinatorial Reconstruction of Homogeneous Spaces

A (possibly discontinuous) map $f : X \to Y$ is called *d-eventually λ-Lipschitz* if $dist_Y(f(x_1), f(x_2)) \leq \lambda \cdot dist_X(x_1, x_2)$ for all those $x_1, x_2 \in X$ where $dist_X(x_1, x_2) \geq d$. for some constant $d = d(f, \lambda) < \infty$.

An eventually Lipschitz map is called a *quasi-isometry* if it is invertible modulo *translations*, where a map $\tau : X \to X$ is called a δ-translation if $supdist(\tau, id) \leq \delta < \infty$.

More specifically, an $f : X \to Y$ is a (d, λ)-*quasi-isometry* if there is a $g : Y \to X$ such that both f and g are d-eventually λ-Lipschitz and the composed self-maps $g \circ f : X \to X$ and $f \circ g : Y \to Y$ are translations. Just "quasi-isometry" means λ-quasi-isometry for some $\lambda < \infty$. (See [70] and references therein for another concept of approximate isometry.)

Quasi-isometric Rigidity and Completeness. A metric space X is called *quasi-isometrically rigid* if there is some $\lambda < \infty$ such that every quasy-isometry $q : X \to X$ lies within bounded distance from a λ-quasi-isometry, say q_λ such that $supdist(q, q_\lambda) < \infty$.

X is called *quasi-isometrically complete* if the group of quasi-isometries modulo translation of X, denoted $qis(X)$, equals the isometry group $iso(X)$.

Clearly, this *completeness implies rigidity* and

co-compact discrete isometry groups Σ of rigid spaces X are inner rigid.

What is non-trivial is that (see [54, 68])

Let X be a Cartesian product of irreducible symmetric spaces of non-compact type and irreducible Euclidean buildings. If X does contains among its factors trees, as well as real hyperbolic and complex hyperbolic spaces then X is quasi-isometrically complete.

Accidentally(?) this the same assumption which ensures the T-property of the group $G = iso(X)$.

Markov Corollary. *Let Σ, Γ be discrete cocompact subgroups in the above group $G = iso(X)$, e.g. $\Sigma = \Gamma$. Then the right action of Σ on the left quotient space G/Γ is Markov hyperbolic.*

Indeed, what one needs besides rigidity is the expansiveness property which is obvious in the present case since G has trivial center and $\sup_{x \in X} dist(g(x), x) = \infty$ for all $g \neq id \in G$.

Remarks and Questions

(a) Probably, "generic" finitely presented groups Γ are q.i. complete, i.e. $qis(\Gamma) = \Gamma$, but $qis(\Gamma)$ may be quite large for certain (which?) groups Γ [16, 84].

(b) Let Σ be a rigid (and divergent) group of automorphisms of Γ or, more generally, a rigid semigroup of partially defined endomorphisms.

Probably, the semidirect product $\Gamma \rtimes \Sigma$ is *inner rigid* except for a specific list of examples (possibly?) including certain (all) expanding endomorphisms of $\mathbb{Z}$ and of nilpotent groups of nilpotency degree 2 with cyclic center.

(c) The quasi-isometric completeness allows a canonical reconstruction of G in terms of a given co-compact subgroup $\Gamma \subset G = iso(X)$ which is, probably, functorial for *quasi-isometric embeddings* (injective homomorphisms?) $\Gamma_1 \to \Gamma_2$ in many cases.

Thus, for example, some compact locally symmetric Kähler manifolds Y can be reconstructed from their fundamental groups Γ in an essentially combinatorial Markovian fashion.

(d) What are non-cocompact $\Sigma \subset G$ for which the action on G/Γ is Markov hyperbolic? Is it true whenever $vol(G/\Sigma) < \infty$?

(e) What is the story for the real and, most interestingly, for the complex hyperbolic spaces X?

One still can reconstruct $iso(X)$ in terms of a $\Gamma \subset iso(X)$ as follows.

Let $\partial_\infty(X)$ be the ideal (hyperbolic) boundary of X. Assume for the simplicity sake, that the action of Γ on X is free and orientation preserving and let $\mathcal{C}_n$, $n = dim(X)$ denote the space of Γ-invariant Borel measures C on $(\partial_\infty(X))^{n+1}$ of finite mass, i.e. the (simplicial) L_1-norm $||C|| < \infty$, where $C \in \mathcal{C}$ represent the fundamental homology class $[X/\Gamma]$ of X/Γ, that is a generator of the (infinite cyclic!) group $H_n(\Gamma, \mathbb{Z})$.

Let $\mathcal{G}$ be the group of self-homeomorphisms h of $\partial_\infty X$ the action of which on $\mathcal{C}_n$ satisfies $||h(C)|| = ||C||$.

If $X = H^n_{\mathbb{R}}$ is real hyperbolic and $n \geq 3$, then the group $\mathcal{G}$ equals G since the support of C of minimal mass equals the set of *regular* ideal n-simplices in X by Milnor -Haagerup-Munkholm theorem.

Probably, $\mathcal{G} = G$ also in the complex hyperbolic case (and, properly stated, in *all* symmetric spaces with no flat and $H^2_{\mathbb{R}}$-factors).

But regardless of whether this is true or not for the complex hyperbolic spaces $X = H^m_{\mathbb{C}}$ of complex dimension $m \geq 2$, one can replace $[X/\Gamma] \in H_n(\Gamma; \mathbb{Z})$ by the *Kähler (Toledo) class* $\boldsymbol{C} \in H_2(\Gamma; \mathbb{Z})$. that is the Poincare dual of ω^{m-1} for the Kähler class $\omega \in H^2(X/\Gamma; \mathbb{Z}) = H^2(\Gamma; \mathbb{Z})$. Equivalently, C is the class in $H_2(X/\Gamma)$ which maximizes the ratio $\omega(C)/||C||_{l_1}$ for the simplicial l_1-norm on homology (in the sense of [27]).

In either case, we need a distinguished class $\omega \in H^2(X/\Gamma; \mathbb{Z})$, and then, it is easy to see that the group $\mathcal{G} = \mathcal{G}(\omega)$ of self-homeomorphisms of $(\partial_\infty(X))^3$ preserving the norms of 2-cycles equals $G = iso(H^m_{\mathbb{C}})$.

(If we choose a "wrong" class $\omega \in H_2(\Gamma)$, then, most likely, $\mathcal{G}(\omega)$ will be equal to Γ itself or a finite extension of Γ.)

(f) What are a measure-theoretic (instead of quasi-isometric) counterparts to rigidity, quasi-orbits, etc. that would be applicable to *non-cocompact* lattices $\Gamma \subset G$ in the context of Margulis-Zimmer super-rigidity theory that would give, in particular, a canonical reconstruction of G from Γ?

Notice that the l_1-homology makes sense with the Poisson-Furstenberg boundary in lieu of the hyperbolic boundary and it is compatible with *measurable* *equivalences* between groups (that are "ergodic bimoduli" where Nekrashevitch type constructions may be possible) but all this was not studied systematically.

(g) Besides the simplicial l_1-norm on homology, there is another norm, which is defined via the *assembly map* α from the group of *reduced rational bordisms* $\mathcal{B}_*(\Gamma)$ to the *rational Wall surgery groups* $\mathcal{L}_*(\Gamma)$.

Recall that each $L \in \mathcal{L}_i$ is represented by a free module M over the rational group ring of Γ with some extra structure (e.g. a non-singular quadratic form for $i = 4j$) on M. Let *rank*(L) denote the minimum of *rank*(M) over all M representing L and

$$||L|| =_{def} \lim_{n\to\infty} \frac{1}{n} rank(n \cdot L).$$

Non-vanishing of this norm on $\alpha(B)$, $B \in \mathcal{B}_*(\Gamma)$, strengthens *Novikov higher signatures conjecture* that claims just non-vanishing of $\alpha(B)$ for $B \neq 0$. It is known, that that this norm does not vanish on the fundamental classes of compact locally symmetric Hermitian spaces of non-compact type with non-zero Euler class, [34, 59] but its role in rigidity remains unclear.

3.2 Stable Factorization of Rigid Flows

Actions of such (softish) groups as $\mathbb{R}$ can not be stable in the above sense, since one can always reparametrize an action along the orbits. Accordingly, (super)stability is defined as preservation of the *partitions (foliations) into orbits* rather than of the actions themselves.

(This notion of stability is poorly adjusted to many "real life systems", e.g. to the amazing stability of the 24-h *circadian rhythm* under variation of temperature, which, at the same time, can be greatly perturbed by a physically insignificant factor, e.g. by a bad news said in a low voice. Finding an adequate functorial concept of "selective stability" is a challenge for mathematicians.)

Examples of such (super)stable actions of $\mathbb{R}$ are *suspensions* of (super)stable actions of $\mathbb{Z} \subset \mathbb{R}$ and geodesic flow on manifolds of negative curvature.

(The Σ^+-suspension of an action of Σ on X for $\Sigma^+ \supset \Sigma$ is the natural action of Σ^+ on $(X \times \Sigma^+)/\Sigma_{diag}$.)

The "combinatorial reconstruction" in this context, say for $\mathbb{R}$-actions, defines the space of *unparameterized* orbits of such an action.

For example, if X is a δ-hyperbolic space X, then the corresponding (space of ideal) unparameterized geodesics correspond to (the space of) pairs of disjoint points in the ideal boundary of X. But this does not automatically deliver the "ideal geodesic flow space" Ω with an actual action of $\mathbb{R}$ with these orbits.

What one does have, however, at least in the cases at hand, is a larger space, say Ξ such that Ω, if it exists, comes as a quotient space of Ξ.

For example, if X is a geodesic hyperbolic space, one may take the space of isometric maps $\mathbb{R} \to X$ for $\varXi$. The images of these maps are distance minimizing geodesics in X and the natural action of $\mathbb{R}$ corresponds to the *geodesic flow*.

The quotient map $\varXi \to \varOmega$ must bring two geodesics together if they have the same ends at the ideal boundary of X but this does not tell you which points on geodesics must be actually identified.

But this ambiguity is "homotopically trivial" and can be easily resolved by a (soft) partition of unity argument as follows.

Let $\varXi$ be a metric space with a free continous action of a locally compact and compactly generated group $\varSigma$ (that is either $\mathbb{R}^n$ or $\mathbb{Z}^n \subset \mathbb{R}^n$ in what follows) and let $\lambda \geq 0$ be constant such that the following three conditions are satisfied.

1. The orbits maps $\varSigma \to \varXi$ are λ- bi-Lipschitz with respect to, say, maximal left-invariant metric, on $\varSigma$ which equals a given metric on a compact subset generating $\varSigma$.
2. If two orbits are *parallel* i.e. the Hausdorff distance between them is finite, then this distance is $\leq \lambda$.
3. Every $2R$-ball in $\varXi$ can be covered by at most $N = N(R)$-balls of radius R.

Let $h : \varSigma \to H = \mathbb{R}^n$ be a continuous homomorphism with compact kernel.

Then there exists a metric space $\varOmega$ with a Lipschitz H-action on it and a continuous quasi-isometric map $P : \varXi \to \varOmega$, which sends $\varSigma$-orbits to H-orbits and such that two orbits in $\varXi$ go to the same orbit in $\varOmega$ if and only if they are parallel.

Moreover, if $\varGamma$ is a discrete isometry group of $\varXi$ which commutes with the action of $\varSigma$, then $\varOmega$ also admits a discrete isometric action of $\varGamma$ which commutes with H and such that the map P is $\varGamma$-equivariant.

Proof. (Compare with [28, 62]). An orbit preserving map $P : \varXi \to \varOmega$ which identify parallel orbits defines a continuous *closed H-valued 1-cocycle* ρ on the set $\varPi \subset \varOmega \times \varOmega$ of the pairs of points that lie on *mutually parallel $\varSigma$*-orbits in $\varXi$, namely,

$$\rho(\chi_1, \chi_2) = P(\chi_1) - P(\chi_2)$$

which make sense for $(\chi_1, \chi_2) \in \varPi$, since $P(\chi_1)$ and $P(\chi_2)$ lie in the same H-orbit.

Conversely every continuous *closed H-valued 1-cocycle* ρ on $\varPi \subset \varOmega \times \varOmega$, that is an anti-symmetric function in two variables, $\rho(\chi_1, \chi_2)$, $(\chi_1, \chi_2) \in \varPi$, such that $\rho(\chi_1, \chi_3) = \rho(\chi_1, \chi_2) + \rho(\chi_2, \chi_3)$, defines a space $\varOmega$ with an H action and a map $P : \varXi \to \varOmega$.

For example if $h : \varSigma \to H = \mathbb{R}^n$ is an isomorphism, then $\varOmega$ is defined as the quotient space of $\varXi$ by the equivalence relation $R \subset \varXi$ that equals the zero set of ρ. In what follows, we stick to this case, since h *is* an isomorphism in most example and since the general case needs only extra notation.

Let us look at cocycles that are represented by locally finite sums

$$\rho = \sum_i \rho_i, \text{ where } \rho_i(\chi_1, \chi_2) = \varphi_i(\chi_1) - \varphi_i(\chi_2)$$

for some Lipschitz maps $\varphi_i : \Xi \to H = \mathbb{R}^n$ with bounded (compact in the present examples) supports.

Such maps are easy to come by. Indeed every Σ-orbit, say $S \subset \Xi$, which is bi-Lipschitz to $\mathbb{R}^n$, admits a Lipschitz retraction $\Phi_S : \Xi \to S = H$ (i.e. $\Phi_S | S = id$) with the Lipschitz constant $\leq n \cdot \lambda$. This Φ_S can be cut off (i.e. made zero) to a $\varphi_\xi = \varphi_\xi(\chi)$, $\xi \in S$, $\chi \in \Xi$, by multiplying it with the function $f_\xi(\chi)$ which equals 1 on the (large) r-ball $B_\xi(R) \subset \Xi$ around ξ, which vanishes outside the concentric $2r$-ball and which equals $r^{-1}(2r - dist(\xi, \chi))$ in the annulus $B(2r) \setminus B(r)$.

What remains is to find a *Γ-invariant* collection of balls $B_i(r) = B_{\xi_i}(r)$ with a large r_i, such that the concentric balls $B_i(r/2)$ cover Ξ, while the intersection multiplicity of the balls $B_i(2r)$ is bounded by a constant $N < \infty$.

For example, if the action of Γ is cocompact, one may take the Γ orbit of a single (large) ball, and the general case needs a minor additional effort.

Remarks and Questions.

1. The main applications of Ω for hyperbolic groups Γ is showing that (certain) cohomology classes in Γ are bounded. Is there something similar for other *globally split hyperbolic* $\mathbb{R}$*-actions*, e.g. for the suspensions of hyperbolic $\mathbb{Z}$-actions?
   The fundamental groups Γ of suspensions of known (all?) hyperbolic $\mathbb{Z}$-actions are amenable and their bounded cohomologies vanish. But possibly there is a meaningful notion of cohomology with *partially* defined cocycles.
2. Can one combinatorially reconstruct the $\mathbb{R}^k$-action on Weyl chambers in locally symmetric spaces X of $\mathbb{R}$-rank k in term of $\pi_1(X)$?
3. Does the above factorization property extend to *non-Abelian groups H*?

The above argument, possibly, can be extended to simply connected *nilpotent* Lie groups H. On the other hand, the conclusion may be even stronger for "rigid". i.e. (some) semisimple H.

3.3 Shadows of Leaves

When is a "quasi-leaf" in a foliated manifold shadowed by a leaf?

We know this is so for the geodesic foliations of (finite and infinite dimensional) manifolds (and singular, i.e. $CAT(-\kappa)$, spaces) of negative curvature and, at least *locally*, for the orbit 1-foliations of general Anosov flows, such as suspensions of hyperbolic automorphisms (and expanding endomorphisms) of infranilmanifolds (where "locally" is, in fact, "semi-global").

Let us look at examples of k-foliations with $k > 1$.

Submanifolds Foliations. Given a smooth manifold X, let $\mathcal{F}_k(X)$ denote the space of pairs (x, S), where $x \in X$ and $S \ni x$ is a germ of an k-submainifold in X. Notice that $\mathcal{F}_k(X)$ is tautologically foliated, where the leaves correspond to *immersed* (with possible self-intersections) submanifolds in X.

We are mostly concerned with "small" $\mathcal{F}$*-saturated* submanifolds $G \subset \mathcal{F}_k(X)$ which are unions of leaves in $\mathcal{F}_k(X)$.

Examples. Let X is a Riemannian manifold. Then the the space marked geodesics in X, denoted $Geo_1 \subset \mathcal{F}_1(X)$, is an instance of our G. The space Geo_1 is foliated into geodesics in X – the orbits of the geodesic flow in (the unite tangent bundle of) X.

Similarly one defines the space $Geo_k \subset \mathcal{F}_k(X)$ the space of marked *totally geodesic* submanifolds in X.

Unlike Geo_1, the space $Geo_{k\geq 2}$ is empty for *generic* Riemannian manifolds X. This suggests another extension of Geo_1 to $k > 1$, namely $G = Plat_k \subset \mathcal{F}_k(X)$ – the space of marked *minimal* k-subvarieties in X that are solutions to the Plateau problem.

This G is infinite dimensional, but it may contain "nice" finite dimensional subfoliations [29]

If X is a complex manifold and $k = 2l$ then one defines $Hol_l \subset \mathcal{F}_k(X)$ – the space of all complex analytic submanifolds $Y \subset X$ with $dim_{\mathbb{C}}(Y) = l$ which, for *Kähler* manifolds X, makes a subfoliation in $Plat_k$.

Also notice that Hol_1 is somewhat similar to Geo_1, since holomorphic curves can be parametrized either by $\mathbb{C}$ or by $H^2_{\mathbb{R}}$; accordingly, these leaves comes from the orbits of $\mathbb{C}$- and/or $SL_2(\mathbb{R})$-actions on the spaces of holomorphic maps of $\mathbb{C}$ and/or of $H^2_{\mathbb{R}}$ to X. (See [37, 61] for a study of the dynamics of such group actions.)

If $k > 1$, then "interesting" finite dimensional saturated $G \subset \mathcal{F}_1(X)$ are rather scarce – the main source of examples of foliations in finite dimensions comes from Lie groups. Namely, we assume that X is transitively acted by a Lie group L and let $Y \subset X$ be an orbit under a connected subgroup in L. This Y, can be regarded as a leaf, say $S_Y \subset \mathcal{F}_k(X)$ for $k = dim(Y)$, and we take the L-orbit of S_Y in $\mathcal{F}_k(X)$ for G.

Why "quasi-leaves" but not "quasi-orbits"? Apparently, smooth actions of Lie groups larger than $\mathbb{R}$, especially of semisimple ones, are very rare [85]. Such an action may be stable just because miracles do not happens twice. On the other hand, one entertains certain freedom in perturbing individual leaves, where stability seems more meaningful.

But what is a "quasi-leaf"?

If a foliated manifold G is endowed with a Riemannian metric, then an immersed k-submanifold $S_0 \to G$ with a *complete* induced Riemannian metric is called an ε*-leaf* if S_0 meets the leaves S of our foliation at the angles $\leq \varepsilon$.

This definition is OK for small $\varepsilon > 0$ when we deal with the *local* shadowing/stability problem, but making sense of it for $\varepsilon \to \infty$, apparently, needs a specific definition case by case.

Next, we say that an $S_0 \to G$ is δ*-shadowed* by a leaf, if there is a δ-small C^0-perturbation of an original immersion $S_0 \to G$ which sends S_0 onto a true leaf $S \subset G$. We express this in writing by $DIST(S, S_0) \leq \delta$. (This is almost, but not quite, the same as the δ-bound on the *Hausdorff distance* $dist_{Hau}(S, S_0)$.) Plain "shadowing" refers, as earlier, to $DIST(S, S_0) < \infty$.

"*ε-Stable*" ("prestable" in the terminology of [32]) means that every ε-leaf is δ-shadowed by a unique leaf where $\delta = \delta(\varepsilon) \to 0$ for $\varepsilon \to 0$, where we may forfeit the uniqueness of the shadow and where the the corresponding definition of "super-stability" i.e. where $\varepsilon \to \infty$ needs a special consideration.

A particular class of examples where the shadowing/stability problem can be, probably, fully resolved, at least for small $\varepsilon > 0$, is that of the foliations associated to *totally geodesic* submanifolds Y in (possibly infinite dimensional) *Riemannian symmetric* spaces X. This can be formulated without direct reference to $G \subset \mathcal{F}_k(X)$ as follows.

Let $Y \subset X$ be a totally geodesic k-submanifold. Say that a k-submanifold $Y_\varepsilon \subset X$ is an *ε-quasi-translate of* Y if for each point $x \in Y_\varepsilon$, there exists an isometry g of X by which Y is moved ε-close to Y_ε, where there several (essentially equivalent) possibilities for this "close". For example, this may mean the existence of an ε-diffeomorphism φ of the unite ball $B_x(1) \subset X$ which maps the intersection $g(Y) \cap B_x(1)$ onto $Y_\varepsilon \cap B_x(1)$, where the ε-property of φ signifies that $dist_{C^i}(\varphi, id) \leq \varepsilon$, for $dist_{C^i}$ being a certain "standard" metric in the space of C^i-smooth maps $B_x(1) \to X$ for a given i.

Shadowing/Stability Problem for Symmetric Spaces. Given the above X, $Y \subset X$ and $\varepsilon > 0$. What are the cases, where Y is *ε-geostable*, i.e. every *ε-quasi-translate of* Y lies within finite Hausdorff distance from the true g-translate of Y for some isometry $g : X \to X$?

Examples. (a) Let X be a symmetric space (possibly of infinite dimension) with non-positive curvature, $K(X) \leq 0$, and $Y \subset X$ be the union of *all* geodesics parallel to a given one. Then

Y is ε-geostable for some $\varepsilon = \varepsilon(X) > 0$. In particular, the maximal flats in X, i.e. maximal subspaces isometric to a Euclidean space, are ε-geostable.

This follows by the standard "transversal (co)hyperbolicity" argument (suitably articulated in in [32], albeit with a few unnecessary extra assumptions on X), which equally applies to some non-symmetric spaces, e.g. to *the Euclidean buildings* and also to the Cartesian products $X = \mathbb{R}^m \times X_1 \times X_2 \times \ldots \times X_k$, where all X_i have $K(X_i \leq -\kappa) < 0$.

The idea is to regard the leaves S corresponding to the translates of Y in the corresponding foliated space G as fixed points of the group Σ of diffeomorphisms σ (or the semigroup of selfmappings) of G which send every leaf of our foliation, say $\mathcal{S}$, into itself.

"Transversal (co)hyperbolicity" signifies that there are "many" $\sigma \in \Sigma$ which "strongly contract" X in "many" directions transversal to the leaves. Such a σ defines a foliation $\mathcal{S}_+$, every leaf S_+ of which is consist of entire S-leaves and σ acting on S_+ brings these leaves closer together. It follows that it also brings closer together quasi-leaves modulo the non-contracting directions, which allows a local shadowing argument a la Anosov.

The simplest case of this is where $X = \mathbb{R} \times X_0$, where $K(X_0) \leq -\kappa < 0$ and dim(Y)=2.

In fact, let $Y \subset X$ be a "quasi-flat" surface, i.e. such that

Y is *quasi-vertical*: its angles with the vertical lines $\mathbb{R}\times x$, $x \in X_0$, are separated away from $\pi/2$;
The intersection of Y with each horizontal slice $r \times X_0 \subset X$, $r \in \mathbb{R}$, is *quasi-geodesic* in $r \times X_0 = X_0$.
Since the (Hausdorff) distance between every two intersections $Y \cap (r_1 \times X_0)$ and $Y \cap (r_2 \times X_0)$ is (obviously) finite, the projection of Y to X_0 lies within a finite distance from a geodesic, say $Y_0 \subset X_0$; hence, Y lies within a finite distance from $\mathbb{R} \times Y_0 \subset X$. □

(b) Let X be a Cartesian square, $X = Z \times Z$ and $Y = Z_{diag} \subset Z \times Z = X$, where Z is a simply connected symmetric space with *no Euclidean, no real hyperbolic and no complex hyperbolic factors.*
Then Y is geostable: there exists an $\varepsilon = \varepsilon(Z) > 0$), such that every ε-quasi-translate $Y_\varepsilon \subset X$ of Y is shadowed by a translate of Y.

Indeed, as we know, these Z are quasi-isometrically rigid, i.e. their the isometry groups are inner rigid, where the general case easily reduces to that where $K(Z) \leq 0$.

Then, if $K(Z) \leq 0$, all one needs of "ε-quasi" is that the angles at which Y_ε meets the fibers of the two projections $X \to Z$ are confined to an interval $[\alpha_1, \alpha_2]$ for $0 < \alpha_1 \leq \alpha_2 < \pi/2$, because such a Y_ε serves as the graph of bi-Lipschitz map $Z \to Z$; hence, it is close to an isometry $g_Z : Z \to Z$. Then the graph of g_Z in X serves as the required translate of Y which lies within finite Hausdorff distance from Y_ε.

Counterexamples. If X is a non-rigid symmetric space then, apparently, the shadowing/stability property fails to be true for most totally geodesic $Y \subset X$ with a notable exception for those from the above (a) and some similar (split) examples.

Questions.

(a) Are there other sources of the failure of geostability in symmetric spaces?
Namely, is every *non-geostable* $Y \subset X$ *contained in a non-rigid* $Y' \supset Y$ in X?
Is this, at least, true for graphs $Y \subset X = Y_0 \times Z$ of isometric embeddings $Y_0 \to Z$?
(Here, $Y_\varepsilon \subset X$ with arbitrarily *large* $\varepsilon < \infty$ essentially correspond to quasi-isometric embeddings $Y_0 \to Z$.)
Are totally geodesic quaternionic geodesic subspaces in the quaternionic hyperbolic space $H^{4n}_{\mathbb{H}}$ geostable?
Is the complex hyperbolic $H^{2n}_{\mathbb{C}} \subset H^{4n}_{\mathbb{H}}$ (of complex dimension n) geostable?

(b) Is there a version (or versions) of geostability in dimensions $k > 2$ similar to that for $k = 1$, which would be stable under *perturbations of the Riemannian metrics* in X which destroy all totally geodesic submanifolds of dimension > 1?
A tangible possibility is offered by the Plateau foliated space $Plat_k$, which, moreover, may be of use in the geostability problem for symmetric spaces X, since many quasi-geodesic (and sometimes even quasi-minimal) k-subvarieties are shadowed by k-volume minimizing subvarieties $Y_{min} \subset X$ (see [29]).
One can show in some cases that such a Y_{min} is *unique*, and if Y is geostable, then Y_{min} is, a posteriori, totally geodesic.

Is there an a priori criterion for a $Y_{min} \subset X$ to be totally geodesic?
Do (the norms of) the second fundamental forms of some minimal subvarieties Y_{min} in symmetric spaces X of non-compact type ever satisfy some maximum principle or enjoy a Bochner-Simons type formula?

(c) Suppose that an ε-quasi-translate $Y_\varepsilon \subset X$ of (a possibly non-geostable) $Y \subset X$ is *periodic*, i.e. invariant under an isometry group Γ_0 of X, such that the quotient Y_ε/Γ_0 is compact. Is Y_ε shadowed by an isometric translate of our (totally geodesic) $Y \subset X$?

This is not true for the real hyperbolic space $X = H^n_{\mathbb{R}}$, even if Γ_0 includes into a discrete isometry group Γ with compact quotient X/Γ; one can "bend" closed totally geodesic submanifold $Y/\Gamma \subset X/\Gamma$ along totally geodesic codimension one submanifolds $Z \subset Y$. But we shall see in the next Sect. 4.1 that a *theorem of Grauert* delivers a *holomorphic periodic stability* in certain cases, which makes the *periodic geo-stability* rather likely, e.g. for some $Y \subset H^m_{\mathbb{C}}$.

Sometimes "periodicity" can be relaxed to "quasiperiodicity" where the action of Γ_0 only "slightly" move Y_ε. E.g. one may allow invariant Y_ε with $Y - \varepsilon\Gamma_0$ having finite volume rather than being compact.

This can be studied in a foliated measure-theoretic set-up (see [19,30,85]) where one has a (suitably understood) measurable family of submanifolds, (i.e. a foliation with transversal measure mapped to X) rather than an individual Y_ε, and where the local and global closeness are understood "on the average", but this picture has not been fully clarified yet.

4 Kähler Stability and Kähler Universality

Let look at the stability/rigidity problem from the angle of the Cauchy-Riemann equations, where we try to obtain "holomorphic objects" e.g. subvarieties, maps or sections of bundles, from (generously understood) *approximately holomorphic* ones.

Ultimately, starting from a "suitable" group Γ, we want to identify/construct some "universal" (generalized) complex analytic space (e.g. an algebraic or a Kähler manifold) $B = B(\Gamma)$, or a holomorphic family of such B, such that all (many) other complex analytic spaces (e.g. Kähler manifolds) V with given homomorphisms $\pi_1(V) \to \Gamma$ would admit canonical holomorphic maps $V \to B$.

Besides Abel-Jacobi-Albanese construction there are two fundamental "super-stability" results in the complex geometry: criteria for the existence of certain complex subvarieties (Grauert) and of holomorphic maps (Siu).

4.1 $\mathbb{C}$-Convexity and the Existence of Complex Subvarieties

Recall that a complex manifold X with a boundary is called $\mathbb{C}$-*convex (at the boundary)* or having a *pseudoconvex boundary* if no (local) holomorphic curve (i.e. Riemann surface) $S \subset X$ can touch the boundary ∂X at a non-boundary point $s \in S$.

A complex manifold X without boundary is called $\mathbb{C}$*-convex at infinity* if it can be exhausted by compact $\mathbb{C}$-convex domains $X_i \subset X$.

X is called *strictly* $\mathbb{C}$*-convex at infinity* if there exists an exhaustion of it by X_i with C^2-smooth boundaries, such that these X_i are C^2*-stably* $\mathbb{C}$*-convex*, i.e. the convexity property persists under all sufficiently small C^2-perturbations of their boundaries

Let $k = dim_{hmt}(X)$ denote the *homotopy dimension* of X, that is the minimum of dimensions of locally contractible topological spaces which are homotopy equivalent to X.

Grauert Exceptional Cycle Theorem. *Let X be a complex manifold which is strictly $\mathbb{C}$-convex at infinity. If its homotopy dimension satisfies $k = dim_{hmt}(X) > m = dim_{\mathbb{C}}(X)$, then k is even and X contains a unique maximal compact complex subvariety of complex dimension $k/2$, say $Y_0 \subset X$, such that the homology inclusion homomorphism $H_k(Y_0) \to H_k(X)$ is an isomorphism.*

Idea of the Proof. The strict $\mathbb{C}$-convexity of the boundary of a relatively compact domain $X_0 \subset X$ implies that *the coherent sheaves* over X_0 have *finite dimensional* cohomology. This implies that there are "many" holomorphic functions in the interior X_0 which extend to a small neigbourhood $X_1 \supset X_0$ with a pole at a complex hypersurface $Z \subset X_1$ which is tangent to the boundary ∂X_0 at a single point.

These functions provide a proper holomorphic map from X_0 to $\mathbb{C}^N$ (with large N) which is an embedding away from a subset $Y_0 \subset X_0$ where this map is locally constant.

The homomorphisms is $H_k(Y_0) \to H_k(X)$ is an isomorphism by the Lefschetz theorem for (possibly) singular complex *Stein spaces*, i.e. properly embedded complex subvarieties in some $\mathbb{C}^N$.

Periodic Hol-Stability Corollary. Let X be a Hermitian (thus, Kählerian) symmetric space with non-positive sectional curvatures and let $Y \subset X$ be a totally geodesic subspace, such that

no $\mathbb{C}$-line tangent Y admits a parallel translation in X normal to Y,

where, observe, such a line at a point $y \in Y$, say $L_{\mathbb{C}} \subset T_y(Y) \subset T_y(X)$ admits a parallel translation in the direction of a unit normal vector $\nu \in N_y(Y) = T_y(X) \ominus T_y(Y) \subset T_y(X)$, if and only if the sectional curvatures of X vanish on the bivectors (ν, l) for all $l \in L_{\mathbb{C}}$. (These curvatures *do not vanish*, for example, if $rank_{\mathbb{R}}(Y) = rank_{\mathbb{R}}(X)$.)

If $dim(Y) > \frac{1}{2}dim_X$ then Y is periodically holomorphically stable: every periodic ε-quasi-translate Y_ε is δ-close to a complex analytic submanifold $Y' \subset X$. Moreover, such a complex analytic Y' close to Y is unique; hence, periodic.

Proof. The "no-parallel translate condition" is equivalent to the *strict* $\mathbb{C}$-convexity of all ρ-neighbourhoods of Y in X and if "ε-quasi" is understood in some C^2-norm, then this property passes to the ρ'-neighbourhoods of Y_ε, for all $\rho' \geq \rho'(\varepsilon) \to 0$ for $\varepsilon \to 0$. □

Remarks. (a) We did not assume Y being complex analytic itself but this follows by sending $\varepsilon \to 0$.

(b) The actual "C^2-quasi" condition can be expressed by a bound on the second fundamental form of Y_ε. This can be relaxed to C^1 (and, probably, to C^0) by smoothing Y_ε.

(c) Probably, "periodic" can be replaced by a suitable "quasi-periodic" with an extension of the L_2-techniques from [42] to the foliated framework but it is unclear if "periodicity" it can be fully removed.

Singular Generalization. The exceptional cycle theorem remains valid for singular complex spaces X, where a hypersurface $\partial \subset X$ is called (strictly) $\mathbb{C}$-convex if the intersection of ∂ with a small neighbourhood $U_x \subset X$ of each point $x \in \partial$ equals the pullback of a (strictly) $\mathbb{C}$-convex hypersurface in $\mathbb{C}^N$ under a holomorphic embedding from U_x into some complex *manifold* U'_x (where one may assume $dim_{\mathbb{C}}(U'_x) = 2dim_{\mathbb{C}}(X)$).

Basic Question. We want to eventually address the followin global rather than ε-local holomorphic stability problem.

Let V be a closed (compact with no boundary) complex analytic space (e.g. manifold) and $\Gamma_0 \subset \Gamma = \pi_1(V)$ be a subgroup.

Let $X_{\Gamma_0} \to V$ denote the Γ_0-*covering* of V, i.e. $\pi_1(X_{\Gamma_0})$ is isomorphically send onto Γ_0 by this covering map and let $k_0 = dim_{\mathbb{Q}hmt}(\Gamma_0)$ be

the minimal number such that Γ_0 *admits a discrete action on a contractible locally contractible metric space* X (which may have fixed points under finite subgroups in Γ) *such that* $dim_{hmt}(X/\Gamma_0) = k_0$, where "discrete" means that, for every bounded subset $B \subset X$ there are at most *finitely many* $\gamma \in \Gamma$, such that $\gamma(B)$ intersects B.

If Γ_0 has no torsion, then

$$dim_{\mathbb{Q}hmt}(\Gamma_0) = dim_{hmt}(\Gamma_0) =_{def} dim_{hmt}(K(\Gamma_0; 1))$$

for the Eilenberg-MacLane classifying space $K(\Gamma_0; 1)$ of Γ_0.

Also, there is a counterpart of this dimension for locally compact groups. For example, if G is a Lie group then $dim_{\mathbb{Q}hmt}(G)$ can be defined as the dimension of the quotient space of G by the maximal compact subgroup in it. More generally, one may look at *proper* actions of G on contractible locally contractible metric spaces X and take the minimal dimension of such an X for $dim_{\mathbb{Q}hmt}(G)$, where "proper" means that for every bounded subset $B \subset X$ the set of those $g \in G$, such that $\gamma(B)$ intersects B is *precompact*.

When does the homology group $H_{k_0}(\Gamma_0; \mathbb{Q})$ *admit a basis which comes (via the Eilenberg-MacLane classifying map* $X_{\Gamma_0} \to K(\Gamma_0, 1)$*) from compact complex subspaces of dimensions* $k_0/2$ *in* X_{Γ_0} ?

Subquestions. What properties of V and/or of Γ_0 could ensure that the homotopy dimension $k_0 = dim_{\mathbb{Q}hmt}(\Gamma_0)$ is *even* and what are conditions for *non-vanishing* of $H_{k_0}(\Gamma_0; \mathbb{Q})$?

Below is an instance of a partial answer for Kähler manifolds V which globalizes the above hol-stability in X, at least in the case of $K(X) < 0$.

($\star$) Let X be a complete simply connected complete Kähler manifold and let Γ_0 be an *undistorted* finitely generated isometry group of X, where "undistorted" means that some (hence every) orbit map $\Gamma_0 \to X$ for $\gamma \mapsto \gamma(x)$ is a *quasi-isometry* on its image for the word metric in Γ_0.

If X has pinched negative curvature, $-\infty < -\kappa_- \leq K(X) \leq -\kappa_+ < 0$, if Γ_0 and if

$$k_0 = dim_{\mathbb{Q}hmt}(\Gamma_0) = dim_{hmt}(X/\Gamma_0) > m = dim_{\mathbb{C}}(X),$$

then

- *k_0 is even,*
- *The homology group $H_{k_0}(\Gamma_0; \mathbb{R}) = H_{k_0}(X/\Gamma_0; \mathbb{R})$ does not vanish. Moreover, the $k_0/2$-power of the Kähler class of X/Γ does not vanish on $H_{k_0}(\Gamma_0)$,*
- *The homology $H_{k_0}(\Gamma_0; \mathbb{R}) = H_{k_0}(X/\Gamma_0; \mathbb{R})$ is generated by the fundamental classes of irreducible components of a compact complex analytic subspace in $V_{hol}^{k_0/2} \subset X/\Gamma_0$.*

Proof. . Since Γ_0 is undistorted and X has *pinched negative* curvature, every orbit $\Gamma(x_0) \subset X$ admits a Γ_0-invariant neighbourhood $U \subset X$ which lies within finite Hausdorff distance from this orbit and the boundary of which is smooth strictly convex by *Anderson's lemma* (see [1] and a sketch of the proof in [$\frown$] below).

Since X is Kähler, (*strict*) *convexity* $\Rightarrow$ (*strict*) $\mathbb{C}$-*convexity*, and Grauert theorem applies to the quotient space X/X_0 (which may be singular as we do not assume the freedom of the action action of Γ_0 on X). □

($\star\star$). This ($\star$) is most interesting where X admits a cocompact isometry group, e.g.

1. *X equals the universal covering $\tilde{V}$ of a compact manifold V without boundary.* In this case the lower bound on $K(X)$ is automatic, all one needs is
2. *The strict negativity of sectional curvatures, $K(V) \leq -\kappa < 0$.* What is especially pleasant for $X = \tilde{V}$ is that the "undistorted" condition becomes a purely algebraic one:
3. The orbit maps $\Gamma_0 \to X$ are undistorted if and only if *the imbedding $\Gamma_0 \subset \Gamma = \pi_1(V)$ is a quasi-isometry for some (hence all) word metrics in Γ and Γ_0.*

Thus, our basic question gets a satisfactory answer under the $(1+2+3)$-condition, i.e. for

undistorted subgroups $\Gamma_0 \subset \Gamma = \pi_1(V)$ with $k_0 = dim_{\mathbb{Q}hmt} > m = dim_{\mathbb{C}}(V)$, where V is compact Kähler manifold with strictly negative sectional curvatures.

Discussion. What are these ($\star$) and ($\star\star$) good for?

As for X, the primely example is the complex hyperbolic space $H^m_{\mathbb{C}}$, $m = dim_{\mathbb{C}}(X)$. But are there good candidates for Γ_0?

It seems hard to come up with examples where the assumptions of $(\star)$ are satisfied, but the existence of the complex subvariety $V^{k_0/2}_{hol} \subset X/\Gamma_0$ was not apparent beforehand.

A pessimistic conjecture (which may thrill champions of "rigidity") is that every top-dimensional irreducible component of $V^{k_0/2}_{hol}$ equals the image of a totally geodesic $H^{k_0/2}_{\mathbb{C}} \subset H^m_{\mathbb{C}}$ in the present case.

On the other hand, $(\star)$ implies that certain groups Γ_0, e.g. with *odd* $k_0 = dim_{\mathbb{Q}hmt}(\Gamma_0) > m$, admit *no undistorted actions* on $H^m_{\mathbb{C}}$.

Notice that the bound $k_0 > m$ is sharp:

cocompact groups Γ_0 of isometries on $H^m_{\mathbb{R}}$ naturally act on $H^m_{\mathbb{C}} \supset H^m_{\mathbb{R}}$, where this action is, clearly, undistorted. On the other hand, according to $(\star)$,

these Γ_0 admit no undistorted actions on $H^{m-1}_{\mathbb{C}}$ for odd k_0.

We shall see in the next section that this remains true for *even* $k_0 \geq 4$ as well, but it remains unclear if one can drop the "undistorted" condition in this case.

Questions and Conjectures.

(a) Possibly, on can relax $K(X) < 0$ to $K(X) \leq 0$ complemented by some extra condition, on Γ_0, e.g. by requiring that the action is

fully undistorted, that is some, (hence, every) orbit $O \subset X$ of Γ_0 admits an *eventually contracting* retraction $R : X \to O$, i.e. such that

$$dist(R(x_1), R(x_2)) \leq \varepsilon \cdot dist(x_1, x_2),$$

where ε depends on $d = dist(x_1, x_2)$ and $D = \min(dist(x_1, O), dist(x_2, O))$, such that $\varepsilon \to 0$ for $d, D \to \infty$.

This is akin to convexity but it has an advantage of being a quasi-isometry invariant. This is equivalent to "undistorted" if $K(X) \leq -\kappa < 0$, but may be strictly stronger, e.g. it is so for non-cocompact lattices acting on irreducible symmetric spaces of $\mathbb{R}$-ranks ≥ 2 according to (by now confirmed) Kazhdan's conjecture.

It is, in general, strictly stronger than "undistorted" but is equivalent to it for $K(X) \leq -\kappa < 0$.

The degenerate case of X/Γ_0 being quasi-isometric to the real line $\mathbb{R}$ allowed by this condition needs to be excluded, e.g. by insisting that X/Γ_0 is *δ-hyperbolic* with the ideal boundary of strictly positive topological dimension.

(b) It seems harder to replace "$K < 0$" by some *purely topological* conditions.

The strongest topological/algebraic substitute(s) for "$K < 0$" is the assumption that X is δ-hyperbolic with the ideal boundary homeomorphic to the sphere S^{2m-1}, $m = dim_{\mathbb{C}}(X)$.

This becomes a "true topology" if we also assume that X admits a cocompact isometry group, say a discrete group Γ freely acting on X.

Also, we need to assume that X is contractible or, at least that the fundamental homology class $[V]_{2m} \in H_{2m}(V)$ of $V = X/\Gamma$ does not vanish in the homology $H_{2m}(\Gamma : \mathbb{R})$.
Would all these allow the conclusions of ($\star$) and ($\star\star$) to remain valid?
Can, at least, one show that X is Stein?
(This seems likely even for **semi**hyperbolic X.)

(c) Can one do anything without "Kähler"?
For example, suppose that all assumptions from (b) are satisfied.
Does then the fundamental group $\Gamma = \pi_1(V)$ admit a *formal Kähler class* $\kappa \in H^2(\Gamma) = H^2(V)$, i.e. such that $\kappa^m = [V] \neq 0$?
If not, is this condition of any use for us?
Probably one handle "non- Kähler" for $m = 2$ in view of *astheno-Kählerian* approach to *Kodaira theory of complex surfaces* [50]

(d) Probably, ($\star$) and/or ($\star\star$) (as well as their conjectural generalizations) remains valid for *singular* Kähler spaces X, where, by definition, a singular Kähler metric is locally induced from an ordinary Kähler structure, (i.e. each point $x \in X$ admits a small neighbourhood $U_x \subset X$ and a holomorphic embedding of U_x to a complex manifold U'_x, such that our singular metric on U_x comes from a smooth Kähler metric on U'_x) and where the condition $K(V) < -\kappa$ must be understood as $CAT(-\kappa)$ in the singular case.

The bottleneck here is Anderson's lemma which, in fact, holds (this is easy) for all $CAT(-\kappa)$ spaces which have the following $[\frown]$-*property* that is obviously satisfied by Riemannian manifolds with *both side bounded curvatures*, $-\kappa \leq K(X) \leq \kappa$, and with the *injectivity radius bounded from below by some* $R > 0$.

$[\frown]$ There exists, for all $\varepsilon_1, \varepsilon_2 > 0$ and each point $x_0 \in X$, a function $\varphi_\frown : X \to \mathbb{R}_+$ with the support in the ε_1-ball $B_{x_0}(\varepsilon_1) \subset X$, such that

the second derivative of $\varphi_\frown$ on every geodesic in X is bounded in the absolute value by $\varepsilon_2 > 0$ and

$$\varphi_\frown(x_0) \geq \delta = \delta(X, \varepsilon_1, \varepsilon_2) > 0.$$

If a $CAT(-\kappa)$-space X satisfies $[\frown]$, then, by an easy argument,

the geodesic convex hull an ε-quasiconvex subset in X is contained in the δ-neighbourhood of U with $\delta \to 0$ for $\varepsilon \to 0$.

This, applied to the ρ-neigbourhood of U with a large ρ, implies Anderson's lemma:

the convex hull of a quasi-convex U lies within finite Hausdorff distance from U.

It is unknown if Anderson's lemma holds for *all* $CAT(-\kappa)$ spaces X, where the test questions are the following. Let $U = B_{x_1}(R) \cup B_{x_2}(R) \subset X$.

Is the convex hull of U contained in the ε-neigbourhood of U where $\varepsilon \to 0$ for $R \to \infty$?

Does, at least, ε admit a bound independent of $dist(x_1, x_2)$ for $R \to \infty$?

The above $[\frown]$ has an obvious $\mathbb{C}$-counterpart, where the bound on the full Hessian (second derivatives) of $\varphi_\frown$ is replaced by such a bound on the complex Hessian.

The relevant version of Anderson's lemma holds (by $[\frown]$ or by an additional argument) for some (all?) singular Kählerian $CAT(-\kappa)$-spaces X with cocompact isometry groups, e.g. for X with isolated singularities.

Singular Kähler Examples. Compact non-singular quotients of the $2m$-ball, $V = H^m_{\mathbb{C}}/\Gamma$ often have many complex *totally geodesic* submanifolds $V_0^{m_0} = H^{m_0}_{\mathbb{C}}/\Gamma_0 \subset V$.

By the Grauert blow-down theorem [25] the space $\underline{V} = V/V_0$ obtained from V by shrinking V_0 to a point has a complex analytic structure for which the obvious map $V \to \underline{V}$ is complex analytic outside V_0.

These $\underline{V}$ carry natural structures of *Artin algebraic spaces* moreover, some of these $\underline{V}$ admit singular Kähler metrics with $K < 0$ and often (always?) these $\underline{V}$ are *projective algebraic*.

On the other hand, the universal covers $\tilde{\underline{V}}$ of these are generically hypebolic by the (generalized) small cancellation theory. However, only relatively few among these are known to carry singular metrics of negative curvature, where the sufficient condition for this a lower bound on the size of the maximal *ρ-collar* $U_\rho \supset V_0$ of V_0 that is the maximal ρ-neighbourhood of V_0 in V which admits a homotopy retraction to V_0. Namely, if $\rho \geq \rho_0$ for some universal ρ_0, (something about $\pi/2$), then $\underline{V} = V/V_0$ carries a singular metric of negative curvature which, apparently (I did not carefully check this) can be chosen Kählerian.

More interestingly, this also applies to singular subvarieties $V_0 \subset V$ that are *immersed* (i.e. with non-trivial selfintersections) totally geodesic in $V = H^m_{\mathbb{C}}/\Gamma$, provided their self-intersection loci are "sufficiently sparse".

Namely, suppose that the self-intersection angles are bounded from zero by some $\alpha > 0$ and let V_0 admits a ρ-collar with $\rho \geq \rho_0(\alpha)$

Then such a collar can be approximated by a locally convex $U \supset V_0$ and by Grauert theorem V/V_0 is complex analytic. Also, such a V has a singular Riemannian (probably, Kählerian) metric of negative curvature. But it is unclear if there are other complex analytic subsets in these V with large approximately locally convex collars.

The picture is somewhat opposite for $V_0 \subset X$ with *many* self-intersections, where such V_0 tend to be "mobile" (ample) rather than exceptional. For example, let $V = H^m_{\mathbb{C}}/\Gamma$ and let V_0 be a immersed totally geodesic (reducible or irreducible) subvariety of complex dimension $m-1$ with $vol_{2m-2}(V_0) \geq const \cdot vol_{2m}(V)$ for a large $cost$. Then most (probably, all) of such V_0 are *connected* and

1. The homomorphism $\pi_1(V_0) \to \pi_1(V)$ is *onto*;
2. V_0 can be included into a *family* of divisors $V_q \subset V$ which are generically *non-singular.*

It is usually easy to see how (2)⇒(1) but the the opposite implication, probably, needs (?) extra conditions on V_0.

It is also not fully clear what happens if $dim_{\mathbb{C}}(V_0) \leq dim_{\mathbb{C}}(V) - 2$, where the cases $dim_{\mathbb{C}}(V_0) \geq dim_{\mathbb{C}}(V)/2$ and $dim_{\mathbb{C}}(V_0) < dim_{\mathbb{C}}(V)/2$ need separate treatments. (Possibly, every $V_0 \subset V$ with $dim_{\mathbb{C}}(V_0) < dim_{\mathbb{C}}(V)/2$ and very many self-intersections is contained in an immersed totally geodesic $V_1 \subset V$ with $dim(V_0) < dim(V_1) < dim(V)$.)

Also the arithmetics of the self-intersection loci of totally geodesic $V_0 \subset V = H_{\mathbb{C}}^m$ (e.g. their *definition fields* and/or arithmetic Galois groups) seems poorly understood.

Besides shrinking subvarieties in compact quotients $H_{\mathbb{C}}^m/\Gamma$ one obtains a attractive class of analytic spaces with "interesting" fundamental group by compactifying $V = H_{\mathbb{C}}^m/\Gamma$ for *non-cocompact* lattices Γ, where the fundamental group $\Gamma_\bullet$ of such a compactification $V_\bullet$ is hyperbolic if V has sufficiently large cusps. In fact, these $V_\bullet$ carry singular (Kählerian?) metrics with negative curvatures.

There are two somewhat opposite constructions of a different kind which deliver spaces of negative (or close to that) curvatures both, in (singular) Riemannian and the Kählerian categories. (Specific instances of this will come up in the next section).

1. *Ramified Coverings* $V_1 \to V$ tend to be more negatively curved than V.
   This is literally true if the branching locus $\Sigma \subset V$ (that is a possibly singular subvariety of codimension two) is totally geodesic; immersed totally geodesics Σ (with self-intersections) also serves well in many cases.
   The above $V_0 \subset H_{\mathbb{C}}^m/\Gamma$ as well as unions of translates of coordinate complex $(m-1)_{\mathbb{C}}$-subtori in $\mathbb{C}^m/\mathbb{Z}^{2m}$ provide examples of such Σ.
2. *Quotient Spaces of non-Free Group Actions.* If the action of Γ, say on $X = H_{\mathbb{C}}^m$, has fixed points, the quotient space $V = X/\Gamma$ may be (not necessarily) singular. However, if the fixed point locus is "sparse" this V may still carry a metric of negative (or close to that) curvature.
   Probably, there are lots of a smooth projective algebraic varieties (defined over number fields) which are biholomorphic to quotients $H_{\mathbb{C}}^m/\Gamma$. (Possibly, there are Kählerian counterexamples in view of [82].)
   More modestly, does every irreducible (smooth?) algebraic $\mathbb{C}$-variety V admit a dominating regular (only rational?) map from some $H_{\mathbb{C}}^m/\Gamma$ to some deformation V' of V?

We shall look closer on specific instances of (1) and (2) in the next section.

Non-holomorphic Problems. It seems that groups Γ_0, e.g. subgroups in a given discrete or a Lie group Γ, with large $dim_{\mathbb{Q}hmt}(\Gamma_0)$ are rather exceptional. (Not nearly as exceptional as arithmetic groups, but something in the same spirit.)

In fact, such subgroups $\Gamma_0 \subset G$ in semisimple (real and p-adic) Lie groups G seems to be associated with (possibly reducible) Γ_0-invariant "subvarieties" in the space G/K (in the Euclidean building for p-adic G) for the maximal compact subgroup $K \subset G$ (and in the Euclidean building in the p-adic case).

On the other hand if a group Γ has large $dim_{\mathbb{Q}hmt}$, then most (sometimes all) infinite quotient groups $\underline{\Gamma}$ of Γ have $dim_{\mathbb{Q}hmt}(\underline{\Gamma}) \geq dim_{\mathbb{Q}hmt}(\Gamma)$. Moreover, these Γ are unlikely to have isometric actions on low dimensional spaces, such as trees, for instance.

Besides large $dim_{\mathbb{Q}hmt}$, what makes a group Γ_0 "rigid" and its actions on X "special" is its connectivity at infinity.

The standard condition of this kind for hyperbolic groups Γ_0 is that its ideal boundary $\partial_\infty(\Gamma_0)$ is connected, locally connected and has no local *cut-points*: every connected subset $U \subset \partial_\infty(\Gamma_0)$ remains connected upon removing a finite subset $S \subset U$.

Alternatively one may bound from below the *cut dimension of* Γ_0, i.e. the minimal topological dimension of an $S \subset \partial_\infty(\Gamma_0)$ such that removal of S from U *does change* the connectedness of U.

There are several candidates for dim_{cut} for non-hyperbolic Γ_0 (e.g. in terms of asymptotic dimensions of subsets in Γ_0), which disrupt the connectedness of Γ_0 at infinity, but I am not certain what the working definition should be.

Questions. Does the "no cut point" condition, or, at least, a stronger bound $dim_{cut} \geq d_0$ for some $d_0 > 0$, imply that every embedding of Γ_0 into a hyperbolic group Γ is *undistorted*?

What characterizes symmetric spaces (and Euclidean buildings) X, such that every discrete isometric action of a Γ_0 on X with $dim_{\mathbb{Q}hmt} \geq d_0$ and $dim_{cut}(\Gamma_0) \geq d_1$ is undistorted? Fully undistorted?

The "undistorted" and "fully undistorted" conditions are accompanied by similar ones, such as absence of parabolic elements in $\Gamma_0 \subset G$ (e.g. satisfied by all Γ_0 in cocompact discrete $\Gamma \subset G$) and/or by *stability* of the action (see the next section and [9, 15, 39, 56]).

But the full scope of relations between such properties remains unclear, where such relations, probably, become more pronounced for "large" $dim_{\mathbb{Q}hmt}$ and/or dim_{cut}.

Call a subgroup Γ_0 in a Lie group (or a p-adic Lie group) a *quasi-lattice* if "an essential part" of Γ_0 is a *lattice* in a Lie subgroup $G' \subset G$, i.e. $G'/(G' \cap \Gamma_0)$ *has finite volume*, and where the "essential part" condition is expressed by

$$dim_{\mathbb{Q}hmt}(G' \cap \Gamma_0) = dim_{\mathbb{Q}hmt}(\Gamma_0).$$

What is the maximal dimension $dim_{\mathbb{Q}hmt}$ of discrete *non-quasi-lattices* $\Gamma_0 \subset G$?

For example, what is the maximal dimension $d_{max} = dim_{\mathbb{Q}hmt}$ of non-quasi-lattices Γ_0 in the isometry group of the hyperbolic quaternion space $H^{4m}_{\mathbb{H}}$?

Does every discrete non-quasi-lattice $\Gamma_0 \subset iso(H^{4m}_{\mathbb{H}})$ with no parabolic elements has $dim_{\mathbb{Q}hmt} \leq 2m$?

How many (undistorted) subgroups Γ from a given class a locally compact (e.g. discrete) group G may contain?

Let us formulate this precisely for hyperbolic groups G and a given set $\mathcal{G}$ of subgroups Γ as follows.

Take the closure $\mathcal{C} = \mathcal{C}(G, \mathcal{G})$ of the set of the ideal boundaries $\partial_\infty(\Gamma) \subset \partial_\infty(G)$, $\Gamma \in \mathcal{G}$, in the Hausdorff (distance) topology and ask ourselves:

What condition(s) on $\Gamma \in \mathcal{G}$ would imply a bound on the dimension of $\mathcal{C}$?

Notice in this regard that if G equals the isometry group of a symmetric space X of negative curvature and $\Gamma \subset G$ is an undistorted subgroup with $dim_{\mathbb{Q}hmt} < dim X$, then there are lots of undistorted subgroups in G isomorphic to the free product of as many copies of Γ as you wish by the (obvious) *Schottky combination theorem*. This makes $dim(\mathcal{C}) = \infty$ in this case.

In order to rule this out, one needs, besides a lower bound on $dim_{\mathbb{Q}hmt}$ for all $\Gamma \in \mathcal{G}$ also such a bound on dim_{cut}, or something like that.

For instance let $\mathcal{G}$ consist of all hyperbolic groups Γ with $dim_{\mathbb{Q}hmt}(\Gamma) = d$, for a given d, and such that the ideal boundaries of these Γ are homeomorphic to the sphere S^{d-1}.

Can one describe the symmetric spaces of a given dimension n (not very large compared to d) and/or hyperbolic groups G with $dim_{\mathbb{Q}hmt}(G) = n$ such that $dim(\mathcal{C}(G,\mathcal{G})) \geq D$ for a given (large) D?

Now a few words about the quotient problem. The only (known) systematic construction of quotients $\underline{\Gamma}$ of hyperbolic groups Γ depends on "collapsing" undistorted subgroups in Γ, (or suitable subgroups in something like a free product $\Gamma * \Gamma'$) where, the generalized small cancellation theory (see 1.5 in [39]) shows that the dimension $dim_{\mathbb{Q}hmt}$ may only *increase* in the process (e.g. if $dim_{\mathbb{Q}hmt}(\Gamma') > dim_{\mathbb{Q}hmt}(\Gamma)$ and Γ' *injects* into $\underline{\Gamma}$.)

What are condition on Γ which would rule out other kind of infinite quotient groups?

For example, does every infinite quotient group $\underline{\Gamma}$ of a Kazhdan's T hyperbolic Γ has $dim_{\mathbb{Q}hmt}(\underline{\Gamma}) \geq dim_{\mathbb{Q}hmt}(\Gamma)$?

In particular, let Γ be a cocompact lattice in $iso(H^{4m}_{\mathbb{H}})$ for $m \geq 2$ and $\underline{\Gamma}$ be an infinite quotient group of Γ.

Can the induced homology homomorphism $H_{4m}(\Gamma) \to H_{4m}(\underline{\Gamma})$ vanish?

Finally, does the above Γ admit an isometric action with *unbounded* orbits on a 2-dimensional simply connected space X with $K(X) \leq 0$ (i.e. $CAT(0)$-space which is not assumed locally compact)?

About dim $= \infty$. Is there an infinite dimensional version of Grauert's exceptional cycle theorem?

To formulate this one needs a notion of "middle dimension" in an infinite dimensional complex variety X.

One option, e.g. for the space X of smooth maps of the circle S^1 to a Kähler manifold V, is suggested by "quasi-splitting" of this X into two "halves" similarly how the space of functions $S^1 \to \mathbb{C}$ "splits" into two *Hardy spaces* of *holomorphic functions* on the Northern and to the Southern hemispheres into which an equatorial $S^1 \subset S^2$ divides the sphere S^2.

Then "$dim_{hmt}(X) > dim_{\mathbb{C}}(X)$" may signify that the gradient flows of C-convex functions on X from a suitable class, can not bring all of X to (a Fredholm perturbation) of such a "half".

Another possibility is suggested by the concept of "mean dimension" for infinite dimensional spaces acted upon by transformation groups [37, 61]

"Philosophical" Questions.

(a) Are there examples where something like the above ($\star\star$) makes sense but which are not ultimately depend on locally symmetric spaces?
Is there something of the kind in the Riemann moduli space of curves, or rather in its universal orbi-covering space X acted upon by the mapping class group Γ?

(b) The $\mathbb{C}$-*convexity* condition is vaguely similar to the *contraction* property in dynamics and Grauert's blow down theorem has a unique fixed point flavour to it. Accordingly, one may draw a similarity between $\mathbb{C}$-*concavity* (associated with *ampleness* or "mobility" of subvarieties) and *expanding* maps (or vice versa?).
What is the holomorphic counterpart of split hyperbolicity? Is there a holomorphic [$\mathbb{C}$-concave]×[$\mathbb{C}$-convex] version of Frank's super-stability theorem?
(*Non-super* stability seems easy: "[exceptional]×[ample]" is, apparently, a stable property of complex subvarieties.)

(c) Looking from a different angle, the existence/nonexistence of a subvariety (algebraic cycle) Y_0 in an algebraic variety V is an essentially *Diophantine* question which can be often resolved by reducing it to linear algebra where the rationality of a solution is automatic.
Grauert's theorem gives a convexity inequality criterion for the existence of certain Y_0 where the proof goes via *infinite* dimensions, similarly to the proof of the Shub-Franks superstability theorem which depends on a contraction/expansion inequality in an *infinite* dimensional space.
Is there a general picture where both arguments would be simultaneously visible?

(d) Is there an algebraic-geometric version of ($\star\star$), at least for locally symmetric V, where everything should be expressed in terms of *finite* coverings $\tilde{V} \to V$ (or of *finite dimensional* representations of Γ) and where "homotopy" must refer to the Grothendieck étale (and/or Nisnevich) topology?
(Probably, it is not hard to identify totally geodesic submanifolds Y_0 and their fundamental (sub)groups $\Gamma_0 \subset \Gamma$ in the purely algebraic/arithmetic language in the spirit of *Kazhdan's theorem on arithmetic varieties*.)

(e) Is there anything in common between (finitary!) Markov presentations of hyperbolic systems and the approximation of the first order theory of algebraic varieties over $\mathbb{C}$ by that over finite fields?

4.2 Existence and Non-existence of Holomorphic Maps

Define the *homotopy rank* of (the homotopy class of) a continuous map $f : X \to Y$, denoted $rank_{hmt}[f]$, as the minimal number r such that f is homotopic to a composed map $X \to P^r \to Y$, where P^r is an r-dimensional cellular space. For example, if Y itself is a cellular (e.g. triangulated) space, then $rank_{hmt}[f] \leq r$ if and only if f is homotopic to a map into the r-skeleton of Y. (At the end of the next Sect. 4.3, we shall refine the concept of $rank_{hmt}[f]$ for maps into non-compact, possibly, infinite dimensional, spaces Y.)

Call a homotopy class of maps $f : V \to W$ between complex analytic spaces ±-*holomorphic*. and write $[f] \in \pm\mathcal{HOL}$ if f is homotopic to a *holomorphic or to an anti-holomorphic* map.

Recall that the Galois group $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$ of $\mathbb{R}\backslash\mathbb{C}$ acts on the category of complex spaces V by conjugation, $V \leftrightarrow \overline{V}$, where the arrow "$\leftrightarrow$" establishes a homeomorphism between V and $\overline{V}$.

For example, if $V \subset \mathbb{C}P^N$ is a complex projective subvariety, then $\overline{V}$ equals the image of V under the conjugation involution $\mathbb{C}P^N \to \mathbb{C}P^N$ given by $z_i \mapsto \bar{z}_i$, $i = 0, 1, \dots, N$.

A map $V \to W$ is antiholomorphic, if the corresponding map $V \to \overline{W}$ (for $\overline{W}$ topologically identified with W) is holomorphic.

Complex Structures on Symmetric Spaces. The main targets of holomorphic maps in relevant examples are *Hermitian symmetric spaces* Y with non-positive sectional curvatures, $K(Y) \leq 0$, and their quotients $W = Y/\Gamma$ by groups of holomorphic isometries $\Gamma \subset iso_{hol}(Y)$. These Y are Kähler and the groups $iso_{hol}(Y)$ are transitive on Y.

The simplest such Y is the *complex hyperbolic space* $H^m_{\mathbb{C}}$ which is biholomorphic (but by no means isometric) to the unit ball in $\mathbb{C}^m$.

If a Hermitian symmetric Y is *irreducible as a Riemannian space*, i.e. does not (non-trivially) isometrically split as $Y = Y_1 \times Y_2$ then it carries *exactly two* complex analytic (Hermitian Kähler) structures invariant under $iso_{hol}(Y)$ – the original one and its conjugate, since the action of the isotropy subgroup $iso_{hol}(Y, y_0)$ on the tangent space $T_y(Y)$ is *irreducible*.

If $Y = Y_1 \times Y_2 \times \ldots \times Y_k$, where Y_i are irreducible Hermitian, then, obviously, Y admits 2^k invariant complex structures. Sometimes, maps into Y and the quotient spaces $W = Y/\Gamma$, which are holomorphic with respect to some of these structures, are called ±holomorphic or just holomorphic maps.

(The complex Euclidean space $\mathbb{C}^n$ is also Hermitian symmetric, but this space, as well as $Y = Y_0 \times \mathbb{C}^n$, must be addressed slightly differently.)

These definitions are justified by a theorem of Y. T. Siu who, in 1980, found a Hodge-Bochner type formula for *harmonic maps* from Kählerian to Riemannian manifolds which has led to a variety of results by Siu and his followers. (see [78, 80] and references therein).

Example: the ***S****ui-****S****ampson-****C****arlson-****T****oledo* $[H^m_{\mathbb{C}}/\Gamma]$*-Theorem.* Let W be covered by the complex hyperbolic space, i.e. $W = H^m_{\mathbb{C}}/\Gamma$ for a free discrete isometry group $\Gamma \subset iso_{hol}(H^m_{\mathbb{C}})$, let V be a compact Kähler manifold and $f_0 : V \to W$ be a continuous map. Denote by $\Gamma_0 \subset \Gamma$ the image of the fundamental group $\pi_1(V)$ in $\Gamma = \pi_1(W)$ under the induced homomorphism $\pi_1(V) \to \Gamma = \pi_1(W)$ (for some choice of the base points in V and W).

Let Γ_0 contains no nilpotent subgroup $\Gamma_0' \subset \Gamma_0$ of finite index (i.e. with card$(\Gamma_0/\Gamma_0') < \infty$) and with the nilpotency degree of Γ_0' at most 2 (e.g. $H^m_{\mathbb{C}}/\Gamma$ is compact and Γ_0 contains no cyclic subgroup of finite index).

If rank$_{hmt}(f_0) > 2$, then $[f_0] \in \pm\mathcal{HOL}$, i.e. f_0 is homotopic to a holomorphic or anti-holomorphic map. Moreover this ±-holomorphic map is unique.

Furthermore, if *rank*$_{hmt}[f_0] = 1, 2$, then

f_0 is homotopic to an "almost holomorphic" map $f : V \to W$, in the sense that f decomposes as $V \to S \to W$, where S is a Riemann surface and the map $V \to S$ is holomorphic.

Moreover, if $H^m_{\mathbb{C}}/\Gamma$ is compact, "no nilpotent subgroup of finite index" can be relaxed to "no cyclic subgroup of finite index".

(In other words, the covering $\tilde{V} \to V$ with the Galois group Γ_0 contains a one parameter family of compact complex subvarieties $V'_s \subset \tilde{V}$ with $dim_C(V'_s) = dim_{\mathbb{C}}(V) - 1$.)

Corollaries. (1: $[V]_{sing}$) This theorem, as stated, remains valid for *singular locally irreducible* Kählerian, e.g. algebraic, varieties V.

Indeed, by Hironaka's theorem, V admits a surjective holomorphic map $V' \to V$ with *connected* fibers, where V' is non-singular.

The composition of a continuous map $f_0 : V \to W$ with $V' \to V$ contains a holomorphic representative $f' : V' \to W$ in its homotopy class, where f' sends every fiber $S_v \subset V'$, $v \in V$, of the map $V' \to V$ to a single point in W, since every contractible holomorphic map of a *connected* S to W with $K(W) \leq 0$ is constant. This gives us a holomorphic map $f : V \to W$ which is, clearly, homotopic to f_0.

(2: $[W_0 \subset W]$) Let $W = H^m_{\mathbb{C}}/\Gamma$, for a discrete torsion free isometry group Γ of $H^m_{\mathbb{C}}$, contain a *unique* maximal compact *connected* complex analytic subspace $W_0 \subset W$ of positive dimension and let the inclusion $W_0 \subset W$ be a homotopy equivalence.

Let f_0 be a continous map from a compact Kähler manifold V to W_0 and let $\Gamma_0 \subset \Gamma = \pi_1(W_0)$ denote the image of $\pi_1(V)$ under f_0. If $rank_{hmt}[f_0] > 2$ and if Γ_0 contains no nilpotent subgroup of finite index (this seems redundant), then f_0 is homotopic to a holomorphic map $V \to W_0$.

Possibly, there are lots of such W and $W_0 \subset W$ but the only examples I see offhand are immersed totally geodesic $W_0 \subset W$.

If the selfintersection of such a W_0 in W is sufficiently sparse (as in "Singular Kähler Examples" of Sect. 4.1) then the inclusion homomorphism $\pi_1(W_0) \to \pi_1(W)$ is injective and, by passing to the $\pi_1(W_0)$-covering of W if necessary, we achieve the above "homotopy equivalence" property.

Also one can show in the "sparse case" that W admits a proper positive $\mathbb{C}$-convex function which vanishes on W_0 and is strictly $\mathbb{C}$-convex away from W_0, so that W_0 is *the only* compact analytic subspace in W.

These W_0, in the interesting cases where they do have self-intersections, are singular with locally reducible singularities; hence, every holomorphic map from a *non-singular* V to W_0 lifts to the normalization W'_0 of W_0 that, if connected, equals $H^{m_0}_{\mathbb{C}}/\Gamma'_0$ for $m_0 = dim_{\mathbb{C}}(W_0)$. (If W'_0 is disconnected, then its each connected component equals $H^{m_i}_{\mathbb{C}}/\Gamma'_i$.)

(3:$[V \times H^m_{\mathbb{C}}]_{cycle}$) Let $X \to V \times W$ be the "diagonal" covering map, i.e. $\pi_1(X)$ *isomorphically* projects onto $\pi_1(V)$ and *surjectively* onto $\Gamma_0 \subset \pi_1(W)$. Observe, that the projection $X \to V$ is a homotopy equivalence – a fibration with the fibers $H^m_{\mathbb{C}}$.

If the generator $[V]_X \in H_{2k}(X) = H_{2k}(V) = \mathbb{Z}$, $k = dim_{\mathbb{C}}(V)$, goes to a *non-zero* class in $H_{2k}(W)$ under the projection $X \to W$ and $k > 1$, then $[V]_X$ can be realized a *compact k-dimensional complex subvariety $V' \subset X$*, namely by the *graph of the holomorphic map $V \to W$ (lifted to X)* which is guaranteed by the SSCT theorem.

On the other hand,

if V is *non-singular*, then, *every* complex subvariety $V' \subset V \times W$ which *projects to* V *with degree 1*, equals the *graph of a holomorphic map* $V \to W$.

Indeed, the fibers of the projection $V' \to V$, that is a regular rational map, are *rationally connected*, while W contains *no rational curves*.

But singular V (e.g. those indicated in Sect. 4.1) may admit non-regular rational maps into W.

(4:[$H^m_{\mathbb{R}}$]) Let $W = H^n_{\mathbb{R}}/\Gamma$, let V be a compact (possibly singular) locally irreducible Kähler space, let $f_0 : V \to W$ be a continuous map and $\Gamma_0 \subset \Gamma$ be the image of $\pi_1(V)$ under f_0.

If Γ_0 *contains no Abelian subgroup of finite index, then* $rank_{hmt}(f_0) \leq 2$*; moreover, if* $rank_{hmt}(f_0) = 0, 1, 2$*, then* f_0 *is homotopic to a map* f *which factors through a holomorphic map* $V \to S$ *for a Riemann surface* S.

To see this observe that $\Gamma \subset iso(H^m_{\mathbb{R}}) \subset iso(H^m_{\mathbb{C}})$ for $H^m_{\mathbb{R}} \subset H^m_{\mathbb{C}}$ and that "no Abelian" implies "no nilpotent" in $iso(H^m_{\mathbb{R}})$.

Since the squared distance function from $H^m_{\mathbb{R}}/\Gamma \subset H^m_{\mathbb{C}}/\Gamma$ is *strictly* $\mathbb{C}$*-convex* on $H^m_{\mathbb{C}}/\Gamma$, every complex submanifold $V \subset H^m_{\mathbb{C}}/\Gamma$ of positive dimension must be contained in $H^m_{\mathbb{R}}/\Gamma$. But $H^m_{\mathbb{R}}/\Gamma$ is totally real in $H^m_{\mathbb{C}}/\Gamma$; hence, $H^m_{\mathbb{C}}/\Gamma$ receives no non-constant holomorphic map from a compact connected analytic space.

We conclude by a corollary that combines SSCT with the Grauert's exceptional cycle theorem. (See Sect. 4.1. It is unclear if the linear analysis, which underlies Grauert's argument, is truly necessary here.)

(5) Let X be a complete simply connected Kähler manifold with strictly negative sectional curvature, $K(X) \leq -\kappa < 0$ (e.g. $X = H^m_{\mathbb{C}}$) and Γ_X be a discrete undistorted torsion free isometry group of X. Assume that $dim_{hmt}(\Gamma_X) = dim_{hmt}(X/\Gamma_X) > m = dim_{\mathbb{C}}(X)$, i.e. $V = X/\Gamma_X$ is not contractible to the middle dimensional skeleton of some (hence, any) triangulation of V.

Let Γ_Y be a a discrete torsion free isometry group of $Y = H^N_{\mathbb{C}}$, such that the quotient space Y/Γ_Y contains no compact complex analytic subvariety of positive dimension, e.g. $\Gamma_Y \subset iso(H^N_{\mathbb{R}}) \subset iso(H^N_{\mathbb{C}})$.

Let $h : \Gamma_X \to \Gamma_Y$ *be a homomorphism such that the image* $\Gamma_0 = h(\Gamma_X) \subset \Gamma_Y$ *contains no nilpotent subgroup of finite index, (where "no Abelian" suffices if* $\Gamma_Y \subset iso(H^N_{\mathbb{R}})$*). Then* $dim_{hmt}(\Gamma_0) \leq 2$*, i.e. the quotient space* $H^N_{\mathbb{C}}/\Gamma_0$ *is contractible to the* 2*-skeleton of some triangulation of this space.*

In particular, no cocompact lattice in $iso(H^n_{\mathbb{R}})$*,* $n \geq 3$*, admits an undistorted action on* $H^{n-1}_{\mathbb{C}}$.

Proof. Apply the SSCT $[H^N_{\mathbb{C}}/\Gamma]$-theorem to the normalization of the subvariety $V \subset W$ delivered by Grauert's exceptional cycle theorem (see Sect. 4.1).

Disclaimer. The $[H^m_{\mathbb{C}}/\Gamma]$-theorem looks even prettier than the **A**bel-**J**acobi-**A**lbanese maps into tori – you do not have to bother with choosing a complex structure in the target. It may look as a possible tool comparable to AJA for constructing "new" holomorphic objects, e.g. for a realization of homology classes in a complex manifold W by *complex analytic* subvarieties – the images of *holomorphic* maps $V \to W$.

However, this possibility seems as remote as in the Grauert case from the previuos section, since producing *Kähler* manifolds satisfying given requirements on their homotopy types seems more difficult than finding complex subvarieties in a given W.

How, on earth, for instance, can you construct a compact Kähler manifold V with fundamental groups Γ admitting a discrete co-compact actions on $H^m_{\mathbb{C}}$, rather than by *first* constructing $H^m_{\mathbb{C}}/\Gamma$ and *then* taking a *subvariety* $V \subset H^m_{\mathbb{C}}/\Gamma$?

Apparently, the only realistic message one can extract from the $[H^m_{\mathbb{C}}/\Gamma]$-theorem in the available examples is "just" *holomorphic rigidity*:

"the only holomorphic representatives in a certain class of continuous objects are the obvious ones if at all."

In fact, "holomorphic realization" of continuous maps $V \to W$ imposes strong restriction on the topologies of both manifolds. For example, the image $[V]_* \in H_{2n}(W)$ of the fundamental class $[V] \in H_{2n}(V)$, $n = \dim_{\mathbb{C}}(V)$ must be (n,n) (as a current) in the Hodge decomposition in W. In particular, if a class $h \in H^{2n}(W)$ is representable by a *holomorphic* $2n$-form on W, then $h[V]_* = 0$. (Compact manifolds $W = H^m_{\mathbb{C}}/\Gamma$ usually carry lots of such n-forms for $m = 2n$.)

Yet, there is some reason for optimism as we shall see by looking at *infinite dimensional* examples.

4.3 Dirichlet Flow into Harmonicity for K ≤ 0

Recall that the *Dirichlet energy* of a C^1-smooth map between Riemannian manifolds, $f : V \to W$ is

$$E(f) = \int_V ||Df||^2 dv.$$

Equivalently, let $UT(V) = Geo_1(V)$ be the unit tangent bundle regarded as the space of marked geodesics $g : \mathbb{R} \to V$. Then

the energy of f equals the energy of the curves $f \circ g : \mathbb{R} \to V$ integrated against the Liouville measure in $UT(V) = Geo_1(V)$,

where, observe, so defined energy makes sense for an *arbitrary* metric space W and a geodesic metric space V with a distinguished (Liouville-like) measure on the space of geodesics in V invariant under the geodesic flow. (Loosely speaking, $E(f) = E(f \circ g)$ for a random geodesic g in V.)

If $dimV = 2$, i.e. V is a surface, then the energy of an f is a *conformal invariant* of the metric in V: it does not change if the Riemannian length (metric) is multiplied by a positive function $\varphi(v)$, since the squared norm of the differential $||Df||^2$ is divided by φ^2, while dv, being a 2-form on surfaces, is multiplied by φ^2 which keeps the integrant $||Df||^2 dv$ unchanged.

It is also clear that $E(f) \geq area(f)$, where the equality holds if and only the map f is conformal.

If W is *Kählerian* with the fundamental 2-form denoted ω_W, then every surface $f : V \to W$ satisfies

$$E(f) \geq area(f) \geq \Big| \int_V f^*(\omega_W) \Big|,$$

where $E(f) = \int_V f^*(\omega_W)$ if and only if the map f is ±-holomorphic, and where, if V is a *closed oriented* surface, the integral $\int_V f^*(\omega_W)$ is a homotopy invariant of f which, in fact, depends only the homology class $f_*[V] \in H_2(W;\mathbb{R})$.

Thus, holomorphic maps $V \to W$ for $dim_{\mathbb{R}}(V) = 2$ are energy minimizing in their homotopy classes.

An elementary computation shows that this remains true for all *Kähler* manifolds V of an arbitrary dimension.

If V and W are Kählerian and V is compact, then every holomorphic map $f : V \to W$ is energy minimizing in its homotopy class, where the minimal energy $E_{min}[f]$ depends only on the homomorphism $f_* : H_2(V) \to H_2(W)$, namely

$$E_{min}[f] = \Big| \int_V \omega_V^{n-1} \wedge f^*(\omega_W) \Big| \text{ for } n = dim_{\mathbb{C}}(V).$$

If $V = V^n \subset \mathbb{C}P^N$ is a projective algebraic variety, and $C \subset V$ is an algebraic curve which equals the intersection of V with a generic $\mathbb{C}P^{N-n+1} \subset \mathbb{C}P^N$, then $E_{min}[f]$ equals the energy of $f : C \to W$. In particular, this energy of $f|C$ does not depend on C for holomorphic maps $V \to W$.

This suggests the definition of the energy for arbitrary (non-holomorphic) f associated to a probability measure μ on the space $\mathcal{C}$ of these curves C, (that equals the Grassmannian $Gr_{N-n+1}(\mathbb{C}P^N)$), as "the energy of f on a random holomorphic curve in V", namely,

$$E_\mu(f) = \int_{\mathcal{C}} E(f|C)d\mu.$$

If f is holomorphic all these energies are equal $E_{min}[f]$; thus,

if a holomorphic f in the homotopy class $[f_0]$ of a continous map $f_0 V \to W$ exists, it necessarily equals the energy minimizing map in this class (where one needs the measure μ to have full $\mathcal{C}$ for its support).

The energy minimizing maps are especially appealing if W has *non-positive* sectional curvature, where the energy $f \mapsto E(f)$ is, obviously, *a geodesically (almost strictly) convex function* on the space of maps $f : V \to W$ in a given homotopy class, and as we shall explain below,

if V is compact, then every homotopy class $[f_0]$ of maps contains a smooth energy minimizing representative f_{min} under a (necessary) mild restriction on this class; moreover, this f_{min} is unique up to properly understood "translations by constants".

But even if one a priori knows that a homotopy class $[f_0]$ does contain a holomorphic map, which necessarily equals f_{min}, one can not (?) directly show that f_{min} is holomorphic, unless one imposes rather stringent assumptions on the local geometry of W. Yet, these assumptions are satisfied for a variety of meaningful examples.

The Euler-Lagrange equations for the Dirichlet energy on C^2-smooth maps f between Riemannian manifolds make a second order non-linear elliptic written as $\Delta f = 0$, where the solutions of this are called *harmonic* maps.

It is easy to see that a C^2-smooth map $f : V \to W$ is harmonic if and only if its second differential at each point $v \in V$, that is the quadratic map between the tangent spaces, $\varphi = D_f^2 : T_v(V) \to T_{f(v)}(W)$ satisfies $\Delta f(v) =_{def} \Delta\varphi(0) = 0$ for $0 \in \mathbb{R}^n = T_v(V)$, where the second differential is defined with the local geodesic coordinates in (V, v) and $(W, f(v))$, and where the Laplace operator $\Delta = \Delta_{\mathbb{R}^n}$ on C^2-maps from a Euclidean space $\mathbb{R}^n$ to a linear space T (that is the tangent space $T_{f(v)}(W)$ in the present case) is defined in the usual way.

Since $\Delta_{\mathbb{R}^n}$ does not depend on the metric in the target space T, but only on the affine structure in T, the equation $\Delta f(v) = 0$ does not fully uses the metric in W but rather the corresponding affine connection. On the other hand, the equation $\Delta f = 0$ is Euler-Lagrange for maps between Riemannian (and pseudo-Riemannian) manifolds, i.e. it represents the stationary points of Dirichlet's energy $E(f) = \int_V ||Df||^2 dv$ and global minima of $E(f)$ is the major (but not the only) source of harmonic maps in geometry.

Similarly to the stationary Euler-Lagrange equation $\Delta f = 0$, one sees that the vector field for the *downstream Dirichlet energy gradient flow* f_t on the space $\mathcal{F} \ni f$ of maps $f : V \to W$ is $f \mapsto \Delta f$ where the vectors $\Delta f(v) \in T_{f(v)}(W)$ are defined as above via the local geodesic coordinates at the points $v \in V$ and with the geodesic coordinates in W at $w = f(v)$.

If W is a Riemannian manifold with *non-positive curvature,* $K(W) \leq 0$ then, by a *theorem of Eells-Sampson*, this flow behaves very much the same as the usual heat flow on functions $V \to \mathbb{R}$.

In fact, $K(W) \leq 0$ makes this flow "more contracting" then for $W = \mathbb{R}^n$; in particular, it is strictly contracting away from a finite dimensional space of directions.

For example, if $K(W < 0)$, then the flow is strictly contracting on the subspace of maps f with *rank*$(Df) \geq 2$, which implies that the fixed point set of the flow – the set of harmonic maps in the homotopy class of $f_0 : V \to W$, consists of a *single* point (or it is empty if the flow slides to infinity in W which makes "strictness not fully strict" after all).

Furthermore the flow f_t, whenever it exists (i.e. if the image $f_t(V) \subset W$, does not "slide to infinity" in W and/or does not "hit the boundary" of W), satisfies the usual *parabolic estimates*, where the most important one is the bound on the norm of the differential of f at each point $v_0 \in V$ in terms of the full energy

$$||Df_{t+1}(v_0)||^2 \leq const_V \int_V ||Df_t(v)||^2 dv.$$

Notice, that the dimension of W does not enter these estimates, as it is especially clear for maps $V \to \mathbb{R}^N$; therefore, this flow is defined and satisfies all Eels-Sampson estimates for *infinite dimensional* Riemannian-Hilbertian manifolds W with $K(W) \leq 0$.

If V and W are compact manifolds without boundaries, then, according to Eells-Sampson,

the energy gradient flow is defined for all $t \geq 0$ *and* f_t *converges, for* $t \to \infty$ *to a harmonic map* $f_\infty : V \to W$.

If W is non-compact, then $f_t(V) \subset W$ may "slide to infinity" as it happens, for example, in $W = S^1 \times \mathbb{R}$ with the metric $\varphi^2(t)ds^2 + dt^2$ for the obvious maps $f_t = S^1 \to S^1 \times t \subset W$ if the function $\varphi(t) > 0$ is decreasing.

If $\lim_{t\to\infty} \varphi(t) = l_{inf} > 0$, then one still obtains a harmonic map f_∞ from the circle S^1 (that is a closed geodesic in the present case), however, not into W but into *the marked Hausdorff limit space* $W_\infty = lim_{t\to\infty}(W, f_t(s_0))$ that equals $S^1(l_{inf}) \times \mathbb{R}$ for the circle $S^1(l_{\mathrm{inf}})$ of length l_{inf}.

But if $l_{inf} = 0$ the family $f_t, t \to \infty$, "collapses" to a constant map into $\mathbb{R}$.

In general, for more complicated V and W, e.g. for $V = S^1\times S^1$ and $W = V\times R$ with the metric $\varphi_1^2(t)ds_1^2 + \varphi_2^2(t)ds_2^2 + dt^2$, one may have a partial collapse of the limit map.

If $dim(W) = \infty$ then $f_t(V)$ may slide to infinity "dimension-wise" rather than "distance-wise". For example, let $\varphi : \mathbb{R}^\infty \to \mathbb{R}_+$ be a convex function with bounded sublevels $\{\varphi(\overline{x}) \leq l\} \subset \mathbb{R}^\infty$ which *does not* achieve its minimum $l_{inf} \geq 0$ on our Hilbert space $\mathbb{R}^\infty$.

Accordingly, the harmonic flow on the circles in $W = S^1 \times \mathbb{R}^\infty$ with the metric $\varphi^2(\overline{x})ds^2 + d\overline{x}^2$ (for $d\overline{x}^2$ denoting the Hilbert metric) will not converge in W, even if it remains in a bounded regions. Yet, the situation here is no worse than that in the $S^1 \times \mathbb{R}$-example: the limit of maps $f_t : S^1 \to W$ goes to the limit space, that is to $S^1(l_{inf}) \times \mathbb{R}^\infty$, except that "limits of pointed metric spaces" must be understood as *ultralimits* as in [41, 55, 63] and in Sects. 6.A.slowromancapiii@, 6D$_3$, 7.A.slowromancapiv@ in [33] (See more on this "stability" at the end of this section).

If W is complete, finite or infinite dimensional, manifold of strictly negative curvature, $K(W) \leq -\kappa < 0$, or even, a possibly singular, complete $CAT(-\kappa)$-space for this matter, then a family of *uniformly Lipschitz maps* $f_t : V \to W$, e.g. finite energy f_t of an Eells-Sampson Dirichlet flow,

can not slide to infinity, except for the case where the action of the group Γ_0 – the image of $\pi_1(V)$ in $\Gamma = \pi_1(W)$ – on the universal covering $X = \tilde{W}$ is parabolic,

where, recall, "parabolic" means, that this action fixes a point, say b, in the ideal boundary $\partial_\infty(X)$ and preserves the *horospheres* centered at b, that are limits of spheres $S_x(R) \subset X$, where $R = dist(x, x_0)$ for a fixed point $x_0 \in X$ and where $x \to b$.

If W is a complete simply connected *finite dimensional* manifold with δ*-pinched* negative curvature, i.e. $-\infty < -\kappa \leq K(X) \leq -\delta\kappa < 0$, then every *discrete parabolic* group $\Gamma \subset iso(X)$ is *virtually nilpotent*, i.e. it contains a nilpotent subgroup of finite index, according to *Margulis' lemma.* Moreover the nilpotency degree of such a Γ is bounded by $\delta/2$.

For example, if $\delta < 4$, e.g. X has constant curvature, then Γ is *virtually Abelian* (which explains the presence of these conditions in the SSCT theorems $[H^m_{\mathbb{C}}]$ and $[H^m_{\mathbb{R}}]$ of the previous section).

Equivariant Maps. It is appropriate to regard maps $V \to W$ as equivariant maps between their respective universal coverings.

In fact, all of the above automatically translates and generalizes to equivariant maps $f : X \to Y$, where X and Y are acted by (not necessarily discrete) isometry groups Γ_X and Γ_Y and "equivariant" refers to a given homomorphism $h : \Gamma_X \to \Gamma_Y$.

One needs (?) the action of Γ_X to be *proper*, e.g. discrete in order to define the energy $E(f)$ as the integral of the (Γ_X-invariant!) energy density $||Df||^2dv$ over X/Γ_X and one needs no assumptions on the action of Γ_Y what-so-ever, except of being isometric.

Despite the apparent *technical* triviality of such a generalization, it significantly broadens the range of *applications* of the Eells-Sampson theorem.

Foliated Harmonicity. The equivariant setting for "periodic" metrics and maps admits the following "almost periodic" generalization.

Let X and Y foliated spaces with Riemannian leaves where the leaves in Y have $K \leq 0$. If the foliation in X comes with a *transversal measure*, on may speak of the energy of such maps and, under suitable (and not fully understood) stability conditions the leaf-wise Dirichlet gradient flow in the space of maps $X \to Y$ which send leaves to leaves converges to a leaf-wise harmonic map $f : X \to Y$ (see [30])

This is relevant for our present purpose if the leaves in both spaces are Kählerian, where we are after leaf-wise holomorphic maps (see next section).

Harmonic Maps with Infinite Energy. The radial projection f_{mdl} from $\mathbb{C} \setminus 0$ to the unit circle $S^1 \subset \mathbb{C}$, and/or the projection $S^1 \times \mathbb{R} \to S^1$ serve as models for general maps with infinite energy, where one exercises a sufficient control on the energy-density.

The existence of similar harmonic maps f from *quasi-Kählerian*, e.g. *quasi-projective*, varieties $V = \overline{V} \setminus \Sigma$ (that are complements to complex subvarieties Σ in compact Kählerian $\overline{V}$, e.g. projective algebraic varieties $\overline{V}$) to spaces with $K \leq 0$ is established in [48, 51, 58, 83].

These maps, in the directions transversal to V_0 near V_0, behave like f_{mdl} and satisfy the natural bounds on the energy density.

Since, eventually, one wants to prove that these f are *pluriharmonic* (see below) it is immaterial which Kähler metric on V is used for harmonicity. Yet, it is technically convenient to work with a *complete* Kähler metric on V, where such metrics are readily available. (See [48, 51], where all this is implemented for maps into finite dimensional spaces with the techniques which, probably, apply to $dim = \infty$ as well.)

About Stability. Let X and Y be acted by isometry groups Γ_X and Γ_Y, let $h : \Gamma_X \to \Gamma_Y$ be a homomorphism, where one can assume without loss of generality that Γ_Y equals the full isometry group $iso(Y)$ the most (but not only) relevant part of it is the image $h(\Gamma_X) \subset iso(Y)$.

Let us formulate several stability properties of h-equivariant maps $f : X \to Y$ which allow a harmonic limit f_∞ of the Eells-Sampson flow $f_t : X \to Y$, where, in the best case, f_∞ sends $X \to Y$, or, if this flow "slides to infinity" in Y, we want to guarantee a very similar harmonic map $X \to Y_\infty$, where the limit space Y_∞ should not be much different from Y.

We assume that Y is a complete simply connected (the latter can be dropped) manifold with $K(Y) \leq 0$, where we do allow $dim(Y) = \infty$. Besides it is convenient (and possible) to admit singular $CAT(0)$-spaces into the picture, keeping in mind that even if Y itself is non-singular the space Y_∞ may be singular.

Also, we can afford Y with *convex* boundary; moreover, if we expect the limit map f_∞ to be holomorphic, then just $\mathbb{C}$-*convexity* of the boundary will do.

As for X, we concentrate on the *co-compact* case, i.e. where the action of $\varGamma_X$ on X is *proper* (i.e. $\gamma \to \infty \Rightarrow \gamma(x) \to \infty$) and $X/\varGamma_X$ is *compact*. In fact, everything equally applies to the case where X is complete, the action of $\varGamma_X$ is proper and the starting map $f_0 : X \to Y$ has $\varGamma_X$*-finite* energy – the integral of the energy density over $X/\varGamma_X$ is finite.

The strongest stability condition, which, due to the point-wise bound on $||Df_t||$, i.e. a uniform Lipschitz bound on f_t, prevents "sliding to infinity" in Y and, thus, insures the limit map $f_\infty : X \to Y$, reads as follows (see [9, 15, 39, 56]).

$[Stab_{strong}]$ There are a geodesically convex subset $Y_0 \subset Y$ (which may have $dim(Y_0) < dim(Y)$) invariant under some isometry group $\varGamma_0$ of Y such that $h(\varGamma_X) \subset \varGamma_0 \subset \varGamma_Y$ and a finite (compact if $\varGamma_X$ is locally compact rather than just discrete) subset $\varDelta \subset \varGamma_X$, such that, for every $C > 0$, the subset $U(\varDelta, C) \subset Y_0$ defined by

$$U(\varDelta, C) \ni y_0 \Leftrightarrow dist_Y(h(\gamma)(y_0), y_0) \leq C \text{ for all } \gamma \in \varDelta$$

is $\varGamma_Y$*-precompact* in Y_0, i.e. it is covered by a $\varGamma_0$-orbit of a compact subset in Y_0.

For example, this condition is (obviously) satisfied by the actions of the above non-purely parabolic $h(\varGamma_X) \subset \varGamma_Y$ on spaces Y with $K(X) \leq -\kappa < 0$.

Notice that $Stab_{max}$ makes sense for *infinite dimensional* Y but this is not particularly interesting, since the normal projection $Y \to Y_0$, which is distance *decreasing*, brings the flow f_t to $Y_0 \subset Y$ and the full geometry of Y (most of which is situated away from Y_0) remains out of the picture.

One can do slightly better with *non-locally compact* convex $Y_0 \subset Y$, $i = 1, 2, \ldots,$ which isometrically split, $Y_0 = \times Z_i$, $i = 1, 2, \ldots$, with locally compact factors Z_i, but this is not very exciting either.

What serves better and covers a wider class of examples, is (as in [7, 8, 80]) a lower bound on the k*-volume* $vol_k[f_0]$ *of the equivariant homotopy class* of $f_0 : X \to Y$ which is similar to but more accurate than the homotopy rank $rank_{hmt}[f_0]$ defined earlier (at the beginning of Sect. 4.2).

Namely, if $n = dim(X/\varGamma_X)$ and $f : X \to Y$ is a smooth equivariant map, we define $vol_n(f)$ as the volume of the integral over $X/\varGamma_X$ of the absolute value of the Jacobian of f. (Typically, $\varGamma_X$ is discrete and $dim(X/\varGamma_X) = dim(X)$, but we do not, a priori, exclude $dim(\varGamma_X)$, where maps f must be regarded locally as maps from $X/\varGamma_X$ rather than from X in order to have correct Jacobians.)

The volume of the equivariant homotopy class of f, denoted $infvol_n[f]$, is defined as the infimum of all smooth maps in the homotopy class $[f]$.

Furthermore, this is especially relevant where $X/\varGamma_X$ is non-compact, we denote by $[f]_L$ the set of maps which can be joined with f by a homotopy of equivariant L*-Lipschitz* maps (if f is *not* L-Lipschitz itself this class is empty) and let $infvol_n[f]_L$ denote this infimum of the volumes of the maps in this class.

Define $infvol_k[f]_L$ for $k \leq dim(X/\varGamma_X)$ by restricting maps f to $\varGamma_X$-invariant piece-wise smooth $X' \subset X$ with $dim(X'/\varGamma_X) = k$ and setting

$$infvol_k[f]_L = \sup_{X'} \inf_{f \in [f]_L} vol_k(f|X')$$

Finally, let $rank_{hmt}[f]_{Lip}$ be the maximal k such that $infvol_k[f]_L > 0$ for all $L > 0$.

Notice that $rank_{hmt}[f]_{Lip} \geq rank_{hmt}[f]_{Lip}$, where the inequality is *strict*, for example, for the identity maps on complete *non-compact* manifolds with *finite* volumes.

The role of this Lipschitz homotopy rank, is to bound from below the rank of the (differential of the) limit map $f_\infty : X \to Y_\infty$, since, clearly,

$$rank(f_\infty) =_{def} \max_{x \in X} rank(D_x f_\infty) \geq rank[f_0]_{Lip}$$

for the Eells-Sampson equivariant Dirichlet flow $f_t : X \to Y$, which starts with a Lipschitz f_0 and converges to harmonic $f_\infty : X \to Y_\infty$.

Notice that

if Y is a (possibly infinite dimensional) symmetric space then the (ultra)limit space Y_∞ is isometric to Y,

but the homomorphism $h_\infty : \Gamma_X \to iso(Y)$ does nor have to be equal to the original h.

Also observe that

if Y has $-\kappa_1 \leq K(Y) \leq -\kappa_2$, then the limit space Y_∞ also has its sectional curvatures pinched between $-\kappa_1$ and $-\kappa_2$.

4.4 From Harmonic to Pluriharmonic for $K_\mathbb{C} \leq 0$

A map f from a complex V manifold to Riemannian one is called *pluriharmonic* if its restriction to *every holomorphic curve (Riemann surface)* in V is harmonic.

Equivalently, pluriharmonicity can be expressed by $Hess_\mathbb{C} f = 0$ for $Hess_\mathbb{C} f(v)$ being "one half" of the second differential $D^2_f : T_v(V) \to T_{f(v)}(W)$ made of the values of Laplacians of D^2_f at $0 \in T_v(V)$ on all holomorphic lines in $T_v(V)$.

Unlike harmonicity, pluruharmonicity is a very stringent condition on maps $f : V \to W$ at the points $v \in V$ where the ranks of the differentials $D_v f : T_v(V) \to T_{f(v)}(W)$ are > 2 – generic W do not receive such maps at all. In fact, maps $V_1 \times V_2 \to W$ which are harmonic on all coordinate "slices" $v_1 \times V_2$ and $V_1 \times v_2$ satisfy *two* determined PDE systems, which make such maps exceptionally rare.

On the other hand, there is the following

List of Standard Pluriharmonic Maps.

$\bullet_{\pm holo}$ ±-*Holomorphic* maps from complex manifolds to Kählerian ones are pluriharmonic.

$\bullet_{plu \circ hol}$ Composed maps $V \overset{holo}{\to} S \overset{harmo}{\to} W$, where if S is a Riemann surface are pluriharmonic.

In fact, composed maps $V \overset{holo}{\to} S \overset{pluri}{\to} W$ are pluriharmonic. for all complex manifolds V and S.

$\bullet_{subm}$ *Riemannian submersions* $f : V \to W$, where V is Kähler and the fibers $f^{-1}(w) \in V$ are complex analytic, are pluriharmonic. ("Riemannian submersion" means that the differential Df *isometrically* sends every the normal space to each fiber say $N_v \subset T_v(V)$, $v \in f^{-1}(w)$, *onto* $T_w(W)$.)

$\bullet_{geod\circ plu}$ *Composed* $geod \circ pluri$ *maps* $V \overset{pluri}{\to} W_0 \overset{geod}{\to} W$, are pluriharmonic, where "geod" refers to *locally isometric geodesic maps* which sends geodesics to geodesics.

$\bullet_{\times}$ *Cartesian products* of pluriharmonic maps $pluri_i : V_i \to W_i$ are pluriharmonic $\times_i pluri_i : \times_i V_i \to \times_i W_i$.

$\bullet_{dia}$ *Diagonal* Cartesian products of pluriharmonic maps $pluri_i : V \to W_i$, that are composed maps $diag : V \to \times_i (V_i = V) \to \times_i W_i$, are pluriharmonic $V \to \times_i W_i$.

[$*$] *Basic Class of Examples.* If W contains a totally geodesic split submanifold, $W \supset W_1 \times W_2$, where W_1 is Kählerian and where W_2 receives a harmonic maps from a Riemann surface, $harmo : S \to W_2$, then, for every holomorphic map $holo : V \to W_1 \times S$, the resulting composed map $V \overset{holo}{\to} W_1 \times S \overset{id\times harmo}{\to} W_1 \times W_2$ is pluriharmonic $V \to W$ with the image in $W_1 \times W_2 \subset W$.

The ***H**odge-**B**ochner-**S**ui-**S**ampsonformula* for *harmonic* maps f of a *Kählarian* V to a (possibly infinite dimensional) Riemannian W can be schematically written as

$$(\star)\quad ||Hess_{\mathbb{C}}||^2 dv = d\langle Df \cdot Hess_{\mathbb{C}} f\rangle_{k-1} + [K_{\mathbb{C}}(W)(Df(\tau_1), \dots, Df(\tau_4))]dv,$$

where $\langle \dots \cdot \dots \rangle_{k-1}$ is a certain bilinear form which takes values in $(k-1)$-forms on V, $k = dim_{\mathbb{R}}(V)$, and $K_{\mathbb{C}}(W)(t_1, t_2, t_3, t_4)$, $t_i \in T(W)$, is "a certain part" of the curvature tensor of W, where $[\dots]$ means that this $K_{\mathbb{C}}(W)$ applies to the images of the tangent vectors $\tau \in T(V)$ under the differential $Df : T(V) \to T(W)$ and then averages in a certain way at each point $v \in V$.

Since

$$\int_V d\langle Df \cdot Hess_{\mathbb{C}} f\rangle_{k-1} = \int_{\partial V} \langle Df \cdot Hess_{\mathbb{C}} f\rangle_{k-1}$$

by Stoke's formula, the inequality $K_{\mathbb{C}}(W) \leq 0$, i.e.

$$K_{\mathbb{C}}(W)(t_1, t_2, t_3, t_4) \leq 0 \text{ for all } t_1, t_2, t_3, t_4 \in T(W),$$

implies

★HBSS: *Every harmonic map f from a Kähler manifold V into W is pluriharmonic,*

provided V is compact without boundary, or it can be exhausted by compact domains with "small" boundaries so that the boundary term in the Stoke's formula goes to zero.

Miraculously, a seemingly impossible problem of the existence of a *pluriharmonic* map $f : V \to W$ is reduced to a realistic one of finding a harmonic map, where, if also $K(X) \leq 0$, one knows (Eells-Sampson) that every map $f_0 : V \to W$ is homotopic to a harmonic one under suitable stability assumptions, e.g. if W is compact.

(Explicit writing down $K_{\mathbb{C}}(W)$, which we have no intention of doing, see [80] for this, shows that $K_{\mathbb{C}}(W) \leq 0 \Rightarrow K(W) \leq 0$; thus we do not have to make any additional assumption on W if $K_{\mathbb{C}}(W) \leq 0$.)

But how could it happen that the Dirichlet flows that, in general, *do not commute* in the space of maps $V \to W$ for *different* Kählerian metric in V, have the *same* fixed point set?

Apparently, these flows "semi-commute" in a certain way, which is partly reflected in $[Tol]$*-convexity* (see Sect. 4.6) and which, if one could expressed this *fully and precisely*, would be useful in other contexts, e.g. for harmonic maps from Riemannian manifolds with special holonomy groups (which for locally symmetric spaces amounts to Margulis superrigidity, [19,49,64]) and for actions of groups with "connected families of virtually split subgroups" (similar to those in Sect. 2.3.) on $CAT(0)$ spaces of finite and infinite dimensions.

Also, it would be nice to find more sophisticated (semi)group actions defined by PDE in function spaces that would display (super)stability similar to that in Sects. 2 and 3.

Since pluriharmonicity is a very restrictive condition, one concludes that the map f must be a very special one, which, in turn, imposes strong constrains on the *homotopy class* of f. (see below).

Flat Targets. If W is a flat manifold, e.g. $W = \mathbb{R}^n/\mathbb{Z}^n$, and the $K_{\mathbb{C}}$-term in $(\star)$ vanishes, then the implication *harmonic* $\Rightarrow$ *pluriharmonic* for maps $f : V \to W$ follows from the *Hodge decomposition* on 1-forms, applied to the differential $\varphi = df$,

every harmonic $\mathbb{R}^n$*-valued (including* $n = \infty$ *[63])* 1*-form* φ *on a compact Kähler manifold* V, that is locally equals the differential of a harmonic map $f : V \to \mathbb{R}^n$, *decomposes into the sum* $\varphi = \varphi_+ + \varphi_-$*, where the forms* $\varphi_\pm$ *are* $\pm$*-holomorphic:*

locally, $\varphi_\pm = df_\pm$*, where the (local) map* $f_+ : V \to \mathbb{C}^n$ *is holomorphic while* $f_- : V \to \mathbb{C}^n$ *is antiholomorphic.*

In particular, the homotopy class $[f]_{Ab}$ of continuous Abel's maps f from V to the torus $W = A(V) = H_1(V;\mathbb{R})/H_1(V;\mathbb{R})$ admits a *pluriharmonic* representative $f_{plu} : V \to A(V)$ for every compact Kähler manifold and hence, for every variety with normal singularities, where this map, as we know, does not depend on any metric in W, but only on the flat affine structure in it.

Moreover, the Hodge decomposition provides the complex structure in this W for which f_{plu} is holomorphic, thus, furnishing the proof of the Abel-Jacobi-Albanese theorem.

This can be also seen by translating the classical

"pluriharmonic function equals the real part of a holomorphic function" to the Hodge theoretic language:

the complex valued 1*-form* $df_{plu} + \sqrt{-1} d_J f_{plu}$ *for* $d_J f_{plu}(\tau) = df_{plu}(\tau/\sqrt{-1})$, $\tau \in T(V)$*, is a closed holomorphic* 1*-form.*

Therefore, $f_{plu} : V \to A(V)$ is holomorphic for the complex structure on $A(V)$ corresponding to the linear anti-involution $J : \varphi \to \varphi_J$ on the linear space $H_1(V;\mathbb{R})$ which is realized by pluriharmonic 1-forms φ on V.

Since the harmonic map theory is invariant under isometries, one automatically obtains, as was explained at the end of the previous section, the following

Equivariant Hodge-Albanese Theorem. Let Γ_Y be an isometry group of the Euclidean space $Y = \mathbb{R}^n$, let X be a complete normal Kähler space with a proper cocompact isometric action of a group Γ_X, let $h : \Gamma_X \to \Gamma_Y$ be a homomorphism and $f_0 : X \to Y = \mathbb{R}^n$ be a continuous h-equivariant map. ("Proper" means $[\gamma \to \infty] \Rightarrow [\gamma(x) \to \infty]$) Then

there exists an isometric action of Γ_Y *on some* $\mathbb{C}^N$*, an* $\mathbb{R}$*-affine surjective equivariant map* $\mathbb{C}^N \to \mathbb{R}^n$*, and a holomorphic equivariant map* $X \to \mathbb{C}^N$ *such that the composed map* $X \to \mathbb{C}^N \to \mathbb{R}^n$ *is equivariantly homotopic to* f_0*.*

This theorem is limited by scarcity of "interesting" isometric group actions on $\mathbb{R}^n$ for finite n. But there are by far more such actions on $\mathbb{R}^\infty$ where the above applies under suitable *stability conditions*, see [13, 63] for instances of this.

Another kind of generalization of Hodge-Albanese concerns mappings of *quasi-projective* algebraic (and *quasi-Kähler*) varieties $\overline{V} \setminus V_0$ to commutative algebraic groups build of Abelian varieties and the multiplicative group $\mathbb{C}^\times$ [46,75]; probably, there is a full equivariant $\mathbb{R}^\infty$-valued version of this.

Discussion on $K_\mathbb{C} \neq 0$. The general HBSS formula ($\star$) is not that surprising in view of the corresponding Hodge formula where $K_\mathbb{C} = 0$, since

every invariant Euclidean formula involving at most second derivatives has its Riemannian counterpart with an extra curvature term.

Actually, $K_\mathbb{C}$ could be *defined* as such an "extra term" in the Hodge formula.

What is remarkable, however, that this $K_\mathbb{C}$ is *non-positive* in a variety of *significant* cases where one can use the Eels-Sampson theorem.

For example, one shows by a local/infinitesimal (sometimes quite involved, [7,8]) computation that the following spaces do have $K_\mathbb{C} \leq 0$.

List of Spaces with $K_\mathbb{C} \leq 0$.

$\bullet_2$ 2-*Dimensional* Y with $K(Y) \leq 0$, i.e. *surfaces* with non-positive curvatures.

$\bullet_{-1/4}$ Riemannian manifolds Y with $-1/4$-*pinched* curvature, i.e. where $-1 \leq K(Y) \leq -1/4$ have $K_\mathbb{C}(Y) \leq 0$. [44]

Furthermore, if the pinching is *strict*, i.e. $-1 < K(Y) < -1/4$, then also $K_\mathbb{C} < 0$, where this inequality, is, by definition (whatever it is), stable under C^2-perturbation of metrics.

Moreover, the above remains true for *the local* $1/4$-pinching condition. i.e. where $-\kappa(y) \leq K_y(Y) \leq -(1/4)\kappa(y)$ for some positive *function* $\kappa(y)$, $y \in Y$.

$\bullet_{sym}$ *Symmetric* spaces Y with $K(Y) \leq 0$ have $K_C(Y) \leq 0$, [7, 72, 78, 80].

Notice that the real hyperbolic spaces $H^n_\mathbb{R}$, $n = 2, 3, \dots, \infty$, of *constant* negative curvature have strictly negative $K_\mathbb{C}$, while the inequality $K_\mathbb{C}(Y) \leq$

0 is non-strict for all other symmetric spaces Y, not even for $Y = H^n_{\mathbb{C}}$ (which have $-1 \leq K(Y) \leq -1/4$) – an arbitrary small perturbation of the metric may bring $K_{\mathbb{C}} > 0$.

$\bullet_{WP}$ *Weil-Peterson* metric on the moduli space of curves. [74] (This metric is non-complete but it is *convex at infinity* which is enough for Eells-Sampson, where, however, the stability condition needs a special attention [11] and where the measurable dynamics provides an alternative to harmonic maps [52].)

$\bullet_{\times}$ Cartesian products of manifolds with $K_{\mathbb{C}} \leq 0$ also have $K_{\mathbb{C}} \leq 0$.

$\bullet_{sing}$ There are some singular spaces, e.g. Euclidean and hyperbolic buildings where $K_{\mathbb{C}} \leq 0$ and the HBSS formula ($\star$) applies, [10, 41].

$\bullet_{\Lambda}$ Let Y, topologically a ball, be a smooth Riemannian manifold and let $\Sigma \subset Y$ be a union of k mutually orthogonal totally geodesic submanifolds of real codimension 2.

Let $\widetilde{Y \setminus \Sigma}$ be the obvious $\mathbb{Z}^k$-covering and Y_Λ be the completion of $\widetilde{Y \setminus \Sigma}/\Lambda$ for some discrete isometry group Λ of $\widetilde{Y \setminus \Sigma}$ isomorphic to $\mathbb{Z}^k$.

The simplest example of this is $\Lambda = \lambda\mathbb{Z}^k$, where λ is an integer and Y_Λ equals a *ramified* covering of Y.

A pleasant instance of $Y_{\lambda\mathbb{Z}^k}$ being defined for *all* $\lambda > 0$, is where Y equals the Euclidean space as well a real or complex hyperbolic space.

If $K_{\mathbb{C}}(Y) \leq 0$, then the natural singular metric on such Y_λ for $\lambda \geq 1$, probably, also has $K_{\mathbb{C}} \leq 0$ which is easy to see for $Y = \mathbb{R}^n$, $H^n_{\mathbb{R}}$ and $H^n_{\mathbb{C}}$.

One can not have $K_{\mathbb{C}}(Y_{\lambda\mathbb{Z}^k}) \leq 0$ for $\lambda < 1$, e.g. where $\lambda = 1/m$ and $Y_{\lambda\mathbb{Z}^k} = Y/\mathbb{Z}^k_m$. But one can, in some cases (where Y harbor much negative curvature away from Σ) make $K_{\mathbb{C}} \leq 0$ by suitably smoothing the metric, where this is sometimes possible even in the Kähler category, e.g. for the *Mostow-Siu* examples (see [78] and references therein).

All being said, the range of possibilities and applications of spaces *locally isometric* to these Y_Λ, which are abundant, especially (but not only) for (finite and infinite dimensional) symmetric spaces Y, remains mainly unexplored.

$\bullet_{dim=\infty}$ Since the condition $K_{\mathbb{C}} \leq 0$ is expressible with quadruples of vectors it makes sense (like sectional curvature and unlike, say, Ricci curvature) for all infinite dimensional Riemannian-Hilbertian manifolds Y, and all of the above applies to these Y.

Our essential examples are infinite dimensional symmetric spaces Y which are completions of the unions of $Y_1 \subset Y_2 \subset \ldots$ for increasing chains of finite dimensional symmetric spaces Y_i and geodesic isometric embeddings $Y_i \subset Y_{i+1}$ (see [57, 69] and references therein).

The simplest infinite dimensional irreducible symmetric space is the real hypebolic $H^\infty_{\mathbb{R}}$, which, as we know, has $K_{\mathbb{C}} < 0$. It admits, for instance, a natural isometric action of every countable subgroup Γ in the *Cremona group* $Bir_2 = Bir(\mathbb{C}P^2)$ of birational automorphisms of the projective plane $\mathbb{C}P^2$. (The group Bir_2 naturally acts on the cohomology $\overline{H}^2$ of the projective limit of all rational surfaces over $\mathbb{C}P^2$ and regular rational maps between them, where this action preserves the intersection form of the type $(+ - - - - - - - - \ldots)$ on $\overline{H}^2$.)

This is used in [14] to show, among other things, that, for instance, "most" such Γ are *not* Kazhdan T and that they *can not* serve as fundamental groups of Kähler manifolds.

4.5 From Pluriharmonic to Holomorphic

Despite the fact that pluriharmonic maps are very special, it is not apparent what they actually are in specific examples.

There are several cases, however, where one knows that such maps either have small ranks or they are holomorphic [7, 8, 78, 80], where these properties can be expressed with the following *local* invariants of Y.

Define $rank_{plu}(Y)$ of a Riemannian manifold Y (possibly infinite dimensional and/or singular) as the maximal number r such that Y receives a pluri-harmonic map of rank r from a complex manifold.

Define $rank_{plu/hol}(Y)$ of a locally irreducible Hermitian manifold Y as the maximal number r such that Y receives a *non-*$\pm$*holomorphic* pluri-harmonic map of rank r from a complex manifold.

If Y (locally) reducible, i.e. if the universal covering $\tilde{Y}$ isometrically splits into a Cartesian product of $\tilde{Y} = \times \tilde{Y}_i$, where Y_i are Hermitian manifolds, then "±holomorphicity" of a map $f : X \to Y$ means ±holomorphicity of the local projections of f to each Y_i.

Siu Rank. These two ranks equal 2 for *generic* Y; they are most significant for *symmetric* spaces Y.

Their main role is to provide a lower bound on $rank_{Siu}(W) = rank_{hmt/hol}(W)$ of a complex manifold W, where $rank_{hmt/hol}$ is the minimal number k such that, for every compact Kähler manifold V, the space of continuous maps $f : V \to W$ with $rank_{hmt} > k$ contracts to the the subspace of ±holomorphic maps i.e. the inclusion $\pm\mathcal{HOL} \subset \mathcal{CONT}$ is a homotopy equivalence.

Here, as earlier, we need to make a provision for locally split W, i.e. where the universal covering $Y = \tilde{W}$ admits a complex analytic splitting $Y = \times_{i=1,...,j} Y_i$ such that the j foliations of W into the Y_i-slices, $i = 1, \dots, j$, are invariant under the Galois group of the covering $Y \to W$ (and so W comes with j mutually transversal holomorphic foliations which locally split W).

Here ±holomorphicity, is understood for such a W, as ±-holomorphicity of the j local projections of our maps on the Y_i-factors.

Remark about the h-Principle. Non-trivial lower bounds on $rank_{hmt/hol}$ are formally similar to the *Oka-Grauert h-principle*, but the underlying mechanisms of the proofs are opposite in nature: the h-principle is a manifestation of "softness" of holomorphic maps, while $rank_{hmt/hol}$ is about "rigidity".

Examples. ($\mathrm{A}_{1/4}$) If a Riemannian manifold Y has $-1 < K(X) < -1/4$, then $rank_{plu}(Y) = 2$; moreover every pluriharmonic map $X \to Y$ locally factors as $X \overset{holo}{\to} S \overset{plu}{\to} Y$ for $dim(S) = 2$. Moreover, this remains true under the corresponding *local* strict $1/4$ negative pinching assumption (as on the $K_{\mathbb{C}}$-list in the previous section.)

This implies, with the ★HBSS from Sect. 4.4 and the stability discussion in Sect. 4.3 the following

Kählerian $1/4$-*non-Pinching Theorem.* (see [7]) Let X be a normal Kähler space with a proper isometric cocompact action of a group $\Gamma_X \subset iso_{hol}(X)$, let Y be a complete, possibly infinite dimensional, manifold which is *strictly negatively* $1/4$ *pinched* i.e. $-1 < -\kappa_1 \leq K(Y) \leq -\kappa_2 < -1/4$, and let $h : \Gamma_X \to iso(Y)$ be a homomorphism. Then

every h-equivariant map $f_0 : X \to Y$ has $rank[f_0]_{Lip} \leq 2$.

(Probably, this remains true for $K_{\mathbb{C}} \leq -\kappa < 0$, e.g. for strictly *locally* negatively $1/4$-pinched manifolds, i.e. for $-(1+\varepsilon)\kappa(y) \leq K_y(Y) \leq -\kappa(y)/4$, where a technical difficulty arises if $\inf_{y\in Y} K_y(Y) = -\infty$ and the limit space Y_∞ is singular.)

($\mathrm{B}_{H^n_{\mathbb{C}}}$) The complex hyperbolic spaces $Y = H^n_{\mathbb{C}}$, $n = 1, 2, \dots, \infty$, satisfy

$$rank_{plu/hol}(H^n_{\mathbb{C}}) = 2;$$

moreover, every *non-±holomorphic* local pluriharmonic map into $H^n_{\mathbb{C}}$ factors as $X \overset{holo}{\to} S \overset{plu}{\to} Y$ for $dim_{\mathbb{R}}(S) = 2$. (see [7, 80])

We have indicated essential corollaries of this in Sect. 4.2; the present terminology allows the following reformulation.

Given a Kählerian X as in the above (A), an $h : \Gamma_X \to iso_{hol}(H^n_{\mathbb{C}})$ and an h-equivariant map $f_0 : X \to H^n_{\mathbb{C}}$.

If $rank[f_0]_{Lip} \geq 3$, then there exists a homomorphism $h_\infty : \Gamma_X \to iso_{hol}(H^n_{\mathbb{C}})$ and an h_∞-equivariant ±holomorphic map $f_\infty : X \to H^n_{\mathbb{C}}$, such that

$$rank_{\mathbb{R}}(f_\infty) \geq rank[f_0]_{Lip}.$$

(A×B) Let X/Γ_X do not fiber over Riemann surface in the sense that X admits no holomorphic equivariant map for any homomorphism h (of Γ_X to the isometry group of the target space) neither to the complex line $\mathbb{C}$ nor to the hyperbolic plane $H^2_{\mathbb{R}}$.

Let $f_0 : X \to Y = Y_1 \times Y_2$ be an equivariant map (for some homomorphism $h : \Gamma_X \to iso(Y)$), let Y_1 be strictly negatively $1/4$-pinched, e.g. $Y_1 = H^\infty_{\mathbb{R}}$, let $Y_2 = H^\infty_{\mathbb{C}}$ and let $rank[f_0]_{Lip} > 0$.

Then the corresponding limit map f_∞ ±-holomorphically send X to a single "holomorphic slice" $y_1 \times H^\infty_{\mathbb{C}}$.

Question. There are, apparently, lots of "interesting" isometry groups Γ acting on $H^\infty_{\mathbb{R}} \times H^\infty_{\mathbb{C}}$ but are there any candidates among them for our kind of groups Γ_X that also act on a Kählerian X, e.g. being the fundamental groups of a complex projective manifolds V?

(C) If Y is a Hermitian symmetric space with $K(X) \leq 0$ and with no flat (i.e. Euclidean) factor, then $rank_{plu/hol}(Y)$ equals the maximum of the real dimensions of totally geodesic subspaces $Y' = Y_0 \times H^2_{\mathbb{R}} \subset Y$, with a Hermitian Y_0

In fact, every non-±-holomorphic pluriharmonic map $X \to Y$ of maximal rank $= rank_{plu/hol}(Y)$ lands in some $Y_0 \times Y_1 \subset Y$ and equals $[X \overset{holo_0}{\to} Y_0] \times [X \overset{holo}{\to} S \overset{harm}{\to} Y_1]$ [7, 8, 80].

(C′) A similar evaluation of $rank_{plu/hol}(Y)$ in terms of split totally geodesic $Y' \subset Y$ is available (albeit the proofs are more involved, see [8]) for most (non-Hermitian) symmetric spaces Y.

Question. Let $f_0 : V \to W$ be a pluriharmonic map, where V is Kähler and W is Hermitian locally symmetric. Suppose that f_0 respects the Hodge filtrations on the cohomologies of the 2 manifolds. Does it help, this map to be holomorphic? Would it be useful to look at the corresponding the cohomologies with coefficient in flat finite and infinite dimensional unitary bundles with more cohomology available?

Discussion on Infinite Dimensional X and Y. Irreducible symmetric spaces Y with $rank_{\mathbb{R}} \geq 2$ admit more "interesting", e.g. discrete and proper, isometric group actions than $H^\infty_{\mathbb{R}} \times H^\infty_{\mathbb{C}}$, e.g. proper actions of Kazhdan's T-groups.

On the other hand the ranks $rank_{plu/hol}(Y)$ for these spaces Y go to infinity for $dim(Y) \to \infty$ (see [7, 8]) and it is unclear under which (global) conditions pluriharmonic maps from *finite dimensional* X into such Y are holomorphic.

There are several possibilities for *infinite dimensional* X, e.g. one can make such X by taking unions of increasing families $X_1 \subset X_2 \subset \dots$ (and inductive limits in general); however, it is unclear what are *meaningful* examples.

(A potentially more promising path to infinite dimensions via "symbolic" varieties is indicated in Sect. 4.10.)

4.6 [Tol]-Convexity and Deformation Completeness

The essential property of a W and/or of a continuous map $f_0 : V \to W$ we are concerned with is $[f_0] \in \pm\mathcal{HOL}$, that is, we look for a ±holomorphic (preferably unique) representative in the homotopy class of f_0.

This has nothing to do with any metric on W and with its curvature, and we naturally wish to have a condition expressible in global, purely complex analytic, even better topological, terms.

Are there global/asymptotic conditions on the universal covering Y of W adequately reflecting $K_{\mathbb{C}} \leq 0$ and $K_{\mathbb{C}} < 0$, similarly to how hyperbolicity captures $K < 0$?

Is there some "generalized CR-structure" on the ideal boundary $\partial_\infty(Y)$?

What does $K_{\mathbb{C}}$ tell you about *Pansu's conformal dimension* of $\partial_\infty(Y)$?

Suppose W is a topological manifold such that the universal covering Y is bi-Lipschitz equivalent to a complete Riemannian manifold Y' with $K_{\mathbb{C}}(Y') \leq -\delta < 0$.

Does W admit maps f_0 from compact Kähler manifolds V with $rank_{hmt}[f_0] \geq 3$?

Would it help to require $K(Y') \geq -const > -\infty$?

Can one relax "bi-Lipschitz" to "quasiisometric" in the case where Y is contractible?

A step toward this direction is suggested by the following

$\mathbb{C}$-*Convexity Lemma* [81]. Let W be a complete Riemannian manifold let C be a smooth connected projective algebraic curve (Riemann surface) and let $[f_0]$ be a homotopy class of maps $C \to W$. The minimum of the Dirichlet energies of maps $f \in [f_0]$ (which, as we know, depends only on the conformal structure on C) defines a function, call it E_{min}, on the moduli space $\mathcal{M}$ of conformal structures on C. If E_{min} is *assumed* by some (harmonic) map $f \in [f_0]$ with *rank* $= 2$ ("assumed" is probably, unnecessary in most cases).

The Toledo Lemma says that *if* $K_{\mathbb{C}}(W) \leq 0$*, then*

$[\mathrm{Tol}]_{conv}$ *the function* E_{min} *is* $\mathbb{C}$*-convex on* $\mathcal{M}$.

($\mathbb{C}$-Convex = pluri-sub-harmonic in the traditional terminology).

Furthermore, let C_b be a family of compact connected Riemann surfaces parametrized by a holomorphic curve $B \subset \mathcal{M}$ and $f_b : C_b \to W$ be *non-constant harmonic* maps with *non-equal images* in W,

$$f_{b_1}(C_{b_1}) \neq f_{b_2}(C_{b_2}) \subset W \text{ for } b_1 \neq b_2.$$

If $K_{\mathbb{C}} < 0$*, then*

the energy $E(f_b)$ *is strictly* $\mathbb{C}$*-convex in* b *at almost all* $b \in B$*,*

where "$\mathbb{C}$-convex" is the same as subharmonic and "almost all" means "from an *open dense* subset in B".

This Toledo's "almost strict" $\mathbb{C}$-convexity is, essentially, as good as $K_{\mathbb{C}} < 0$ for ruling out Kähler subgroups in $\pi_1(W)$, since, obviously,

every continous map f_0 *from a projective algebraic manifold* V *to such a* $[Tol]_{alm.strict}$ *space* W *has rank*$[f_0]_{hmt} \leq 2$.

Remarks. (a) If the condition $f_{b_1}(C_{b_1}) \neq f_{b_2}(C_{b_2})$ is not satisfied, then, this was pointed out to me by Toledo, the energy $E(f_b)$ may be *constant* in b, where the basic examples are families of ramified holomorphic covering maps between Riemann surfaces, $C_b \to W$, $dim_{\mathbb{C}}(W) = 1$. Probably, if $dim_{\mathbb{C}}(W) > 1$, then the only way the strict convexity fails is where the maps f_b factor through such ramified coverings over a minimal surface in W.

(b) The condition $[\mathrm{Tol}]_{conv}$ (unlike $K \leq 0$) makes sense for an *arbitrary* metric space W and this also applies to (properly reformulated) property $[\mathrm{Tol}]_{alm.strict}$.

Also both conditions have a global flavour and they are (slightly) more robust then the corresponding $K_{\mathbb{C}}$-curvature inequalities.

However, they limited to harmonic maps of *closed* surfaces into W, and it would be nice to have something similar for open surfaces – non-compact and/or with boundaries. This would, in particular, allow a local version of [Tol]-convexity and might imply stability of this property, and consequently of $K_{\mathbb{C}} \leq 0$, under weak (Hausdorff?) limits of metric spaces.

From now on, $[\mathrm{Tol}]_{conv}$*-manifolds* W are those satisfying this $\mathbb{C}$-convexity property.

"Holomorphic" Corollary. Let V be a projective algebraic manifold, let $F : V \to W$ be a *continuous* map and let $C_q \subset V$ be an algebraic family of algebraic curves in V, where q runs over a smooth connected projective algebraic curve Q, such that C_q is connected non-singular for generic $q \in Q$.

If W is [Tol]-convex, then the minimal energy $E_{min}(q)$ of the (homotopy class of the) map F restricted to non-singular curves is constant in q.

Proof. If C_q is a singular curve in our family and $f_q : C_q \to W$ is a continuous map which is smooth away from the singular points of C_q, then the energy of this map is defined by integrating $||Df_q||^2$ over the non-singular locus of C_q. Thus, the function $E_{min}(q)$ is defined for all $q \in Q$.

Let us show that the function $E_{min}(q)$ is continuous on Q for all (not necessarily [Tol]-convex) Riemannian manifolds W.

This is obvious at those q where the curves C_q are non-singular. It is also clear that E_{min} is semicontinuous at all q: the energy may only jump up for $q \to q_0$.

To see that, in fact, there is no "jump", i.e. $E_{min}(q_0) \leq \lim E_{min}(q)$, assume (the general case trivially reduces to this) that the curve $C_0 = C_{q_0}$ has a double point singularity, say at $c_0 \in C_0$. Let $\tilde{C}_0 \to C_0$ be the non-singular parametrization (by normalization) of C_0 and observe that $E_{min}(\tilde{C}_0) = E_{min}(C_0)$. (The extremal harmonic map $\tilde{C}_0 \to W$ does not, typically, factors through the parametrizing map $\tilde{C}_0 \to C_0$, which makes the extremal map $C_0 \to C_0$ discontinuous).

The curves C_q, $q \to q_0$, are obtained near c_0 by attaching "arbitrarily narro" 1-handles H_ε to $\tilde{C}_0 = C_0$ by gluing the two branches of C_0 along the two infinitesimally small circles in C_0 around c_0.

These handles can be implemented by maps $H_q \to W$ with the energies $E(H_q) \to 0$, that makes

$$(\circ) \qquad \limsup_{q \to q_0} E_{min}(C_q) \leq E_{min}(\tilde{C}_0) = E_{min}(C_0).$$

All one needs to construct such $H_q \to W$ is a family φ_δ, $\delta > 0$, of smooth functions $\varphi_\delta : D \to \mathbb{R}$, where $D \subset \mathbb{C}$ is the unit disc, where $E(\varphi_\delta) \leq \delta$ for all $\delta > 0$ and where

- The functions φ_δ vanish on the boundary of D.
- The functions φ_δ equal 1 at the center $0 \in D$.

(Such φ_δ are made out of $\delta \log |z|$ in an obvious way.)

Thus, by $(\circ)$, the function $E_{min}(q)$ is continuous at all, including singular, curves C_q in our family, and if W is [Tol]-convex $E_{min}(q)$ is constant on Q, since *every continuous* function on D which is smooth $\mathbb{C}$-convex on Q minus a finite subset is constant on Q. (I am not certain, but this is, undoubtedly, known, if every bounded $\mathbb{C}$-convex function on $D \setminus 0$ is continuous at 0.)

Curve Deformation Completeness. A complex manifold W is called cdc, if the following holds for all above (V, C_q, f_0):

if the continuous map $f_{q_0} = f_0|C_{q_0} : C_{q_0} \to W$ is homotopic to a holomorphic one for some $q_0 \in Q$, then all $f_q : C_q \to W$ are homotopic to holomorphic maps.

Observe that the cdc property, unlike [Tol]-convexity and/or $K \leq 0$, does not depend on any metric, but only on the complex structure in W.

On the other hand, the above Corollary implies that

Kählerian [Tol]$_{conv}$-manifolds are cdc – curve deformation complete.

Indeed, if a smooth map $f_q : C_q \to W$ is homotopic to a holomorphic $h : C_q \to W$, then $E(f_q) \geq E(h)$ by *Wirtinger's inequality*, where, a priori, the equality holds if and only if f_q is $\pm$-holomorphic. Since the $\pm$-involution amounts to the change of the orientation in C_q, a $(+)$-holomorphic map of positive rank into a *Kähler* manifold W can not be homotopic to a $(-)$-holomorphic one; hence all f_q are holomorphic. ("Kähler" is essential: Calabi-Eckmann-Hopf manifolds $H_{mn} = [(\mathbb{C}^m \setminus 0) \times (\mathbb{C}^n \setminus 0)]/\{e^z \times e^{\sqrt{-1}z}\}_{z \in \mathbb{C}}$ contain *contractible* holomorphic curves, $H_{11} \subset H_{mn}$.)

Let us indicate some simple properties of *curve deformation complete compact Kähler* manifolds W.

Holomorphic Extensions from Curves. Start with the case where a homotopy class $[e_0]$ of maps of a compact complex curve (Riemann surface) C to W admits *at most one holomorphic* representative $C \to W$ for every complex structure on C. This is so, for example, for all *non-contractible* maps $e_0 : C \to W$ if W admits a Kähler metric with strictly negative sectional curvature.

Remarks. (a) If the space E_0 of *continuous* maps $e : C \to W$ homotopic to e_0 has $H^2(E_0; \mathbb{R}) = 0$, then the space of *holomorphic* maps $C \to W$ homotopic to e_0 is *finite*.

Indeed, the space $H_0 \subset E_0$ of holomorphic maps $h : C \to W$ is a complex space, such that the c-evaluation map $\varepsilon_c : E_0 \to W$ for $\varepsilon_c : e \mapsto e(c) \in W$ is *holomorphic* on $H_0 \subset E_0$ for ecah $c \in C$. Since W is Kähler as well as compact the space H_0 is compact.

If $dim_{\mathbb{C}}(H_0) = d > 0$, then the pullback $\varepsilon_c^*(\Omega_W)$ of the Kähler class Ω_W of W to E_0 by ε does not vanish on H_0 for generic $c \in C$, since

$$\varepsilon_c^*(\Omega_W)^d[H_0] > 0;$$

for the fundamental class $[H_0]$ of (a d-dimensional irreducible component of) H_0; hence, $H^2(E_0) \neq 0$ for $d > 0$.

(b) If W is *aspherical*, i.e. the universal covering of W is contractible, then E_0 is homotopy equivalent to the Eilenberg-MacLane classifying space $K(Z_0; 1)$ of the centralizer Z_0 of the $[e_0]$-image of the fundamental group of C in $\pi_1(W)$. For instance,

(c) If W admits a Riemanninan metric with strictly negative curvature, then $H^2(E_0; \mathbb{R}) = 0$ for all non-contractible maps $e_0 : C \to W$.

(d) If W admits a Kähler metric with non-positive sectional curvature, then the space of holomorphic maps $h : C \to W$ in every homotopy class $[e_0]$ is connected; such an h is unique, if and only if the the e_0-image of $\pi_1(C)$ in $\pi_1(W)$, has trivial centralizer Z_0.

Probably, there are many algebraic manifolds W where the space of holomorphic maps $C \to W$ in a given homotopy class is disconnected, even for aspherical W. One could arrange this, for example, with an algebraic curve B in the moduli space $\mathcal{M}_g$ of curves C of a genus g, where B an irreducible component of the lift $\tilde{B}$ of B to the universal orbicovering $\tilde{\mathcal{M}}_g$ has a double point.

Assume the above algebraic manifold V is a non-singular compact complex surface, i.e. $dim_{\mathbb{C}}(V) = 2$, let the family C_q be *non-constant* so that the union of all curves $C_q \subset V$ equals V and let us show that all *compact Kähler cdc* manifolds W *without rational curves* satisfy the following *extension property*.

$\star$ *Every continuous map $f_0 : V \to W$ which is holomorphic on some non-singular curve $C_{q_0} \subset V$ is homotopic to a unique holomorphic map $V \to W$, provided the homotopy class of $e_0 = f_0|C_{q_0}$ contains at most one holomorphic representative $C_q \to W$ for all q in a small neighbourhood $U_0 \subset Q$ of q_0.*

Proof. Let $Z \subset V \times W$ be the union of (possibly singular) holomorphic curves $\tilde{C} \subset V \times W$ such that

- The projection $P_V : V \times W \to V$ biholomorphically sends each $\tilde{C}$ onto a curve $C_q \subset V$ for some $q \in Q$ and the projection $P_W : V \times W \to W$ is holomorphic on each $\tilde{C}$;

•• All curves $\tilde{C} \subset V \times W$ have "degrees" equal that of the graph $\Gamma(C_{q_0}) \subset V \times W$ of the map $f_0|C_{q_0} : C_{q_0} \to W$, i.e. the value of the Kähler class $\Omega = \Omega_{V\times W} = \Omega_V \oplus \Omega_W$ on each $\tilde{C}_q$, that is $\Omega[\tilde{C}_q]$, equals $\Omega[\Gamma(C_{q_0})]$.

Since • is comprised of sentences in the "first order holomorphic language" and •• is "linguistically Kähler", this Z is a *compact complex analytic* subset in $V \times W$. (One usually applies such "linguistic" argument to algebraic manifolds but it remains valid in the compact Kähler case as well.)

Let $Z_0 \subset Z$ be the irreducible component of Z which contains the graph of the (holomorphic!) map $f_0|C_{q_0} : C_{q_0} \to W$.

Since the projection $P_V : Z_0 \to V$ is one-to-one over U_0 it is *generically* one-to-one; hence, Z_0 serves as the graph of a rational map $V \to W$ that, in fact, is holomorphic since W contains no rational curves.

Remarks. The above argument does not work if "at most *one* holomorphic representative $C_q \to W$" ($q \in U_0$) is relaxed to "at most *finitely many* holomorphic representatives $C_q \to W$" or to "at most one holomorphic representative $C_{q_0} \to W$" (for a single q_0), but it seems hard to find an actual cdc manifold where the above extension property fails to be true under the so relaxed conditions.

It is also unclear whether "without rational curves" is relevant for cdc manifolds W.

Let us adjust the above to the case where there may be several mutually homotopic holomorphic maps of a curve C into W, keeping in mind split submanifolds $W' \times W''$ in a manifold W of non-positive curvature. (The proper condition was missing from the first version of this paper as was pointed out to me by Domingo Toledo).

Let W, assumed as earlier compact Kähler cdc, have contractible universal covering. Assume moreover, that W has the following

Split Deformation Property. Given a homotopy class $[e_0]$ of maps of a closed surface C_0 into W there exists a unique (possibly empty) compact split irreducible analytic space $W' \times W''$ and a holomorphic map $I : W' \times W'' \to W$ with the following property.

Let C_q be a complex curve (Riemann surface) homeomorphic to C_0 and $h_q : C_q \to W$ be a holomorphic map such that the composition of h_q with a homeomorphism $C_0 \leftrightarrow C_q$ belongs to the homotopy class $[e_0]$.

Then there exist unique *holomorphic* map $h'_q : C_q \to W'$ and a *constant* map $g''_q : C_q \to W''$, such that the map h_q factorizes as $C \overset{h'_q \times g''_q}{\to} W' \times W'' \overset{I}{\to} W$, i.e. $h_q = I \circ (h'_q \times g''_q)$.

Basic Example. The split deformation property is enjoyed by compact Kähler manifolds W with *non-positive* sectional curvatures.

Indeed, let $\Gamma_0 \subset \pi_1(W)$ be a subgroup let $\Gamma''_0 \subset \pi_1(W)$ be its centralizer and $\Gamma'_0 \supset \Gamma_0$ be the centralizer of Γ''_0.

Since $K(W) \leq 0$ – "Kähler" is irrelevant here – there exists, by the Gromoll-Wolf-Lawson-Yau splitting theorem, a split Riemannian manifold $W'_0 \times W''_0$ and a map $I_0 : W'_0 \times W''_0 \to W$ which is geodesic isometric on all coordinate slices $w'_0 \times W''_0$ and $W'_0 \times w''_0$ and such that the I_0-images of the fundamental groups of W'_0 and W''_0 in $\pi_1(W)$ are conjugate to Γ'_0 and Γ''_0 respectively.

Now, recall that W is Kähler and assume that W'_0 contains an irreducible complex analytic subset A_0 such that the image of the fundamental group of A_0 in $\Gamma'_0 = \pi_1(W'_0)$ contains Γ_0.

Then W''_0 identifies with the space of holomorphic maps $A \to W$ which send $\pi_1(A)$ onto Γ_0 while W'_0 appears as the space of holomorphic maps $W''_0 \to W$ homotopic to $I_0|w'_0 \times W''_0$. Thus, the manifolds W'_0 and W''_0 acquire complex analytic structures for which the map I is complex analytic.

Questions. Given a compact Kähler (e.g. algebraic) manifold W, call (the conjugacy class of) a subgroup $\Delta \subset \Gamma = \pi_1(W)$ *holomorphic*, if it equals the image of the fundamnetal group of a compact connected Kähler (possibly singular) space A under a *holomorphic* map $A \to W$.

What is the structure $\mathcal{H} = \mathcal{H}(\Gamma) = \mathcal{H}(\Gamma, W)$ of (the set of) holomorphic subgroups Δ in Γ?

For instance, for which W is the centralizer of a holomorphic subgroup is holomorphic (as for the above W with $K(W) \leq 0$)?

When can one reconstruct the complex structure in W in terms of $\mathcal{H}(\Gamma)$?

Which manifolds W have many and which few holomorphic subgroup?

For example, what are aspherical W (e.g. with $K(W) \leq 0$), where every holomorphic Δ either has finite index in Γ or equals $\{id\} \subset \Gamma$?

What is the corresponding structure for the *algebraic* fundamental group that is the profinite completion of Γ?

If W is an algebraic manifold, what is a similar structure on the full geometric Galois group Γ_{geo} of W that encodes all ramified coverings of W?

Does $\Gamma_{geo}(W)$ allow a reconstruction of W?

Split Deformation Property $\Rightarrow$ *Extension Property.* Return to the above non-singular compact complex surface V and a non-constant family of curves $C_q \subset V$, parametrized by an irreducible algebraic curve $Q \ni q$.

$\star\star$ *Let W be a compact Kähler cdc (curve deformation complete) manifold which also enjoy SDP and let $f_0 : V \to W$ be a continuous map which is holomorphic on some curve $C_{q_0} \subset V$. Then f_0 is homotopic to a holomorphic map $V \to W$ in the following two cases:*

(1) *Every continuous map $Q \to W$ is contractible, e.g. Q is a rational curve and W is aspherical;*
(2) *The inclusion homomorphism $\pi_1(C_{q_0}) \to \pi_1(V)$ is onto.*

Proof. Because of SDP everything reduces to maps into $W' \times W''$ and the proof of the above $\star$ applies to maps $V \to W'$.

Questions

(a) Let W be $[\mathrm{Tol}]_{conv}$ and let $f_q : C_q \to W$ be a family of harmonic (rather than holomorphic) maps for generic points $q \in Q$ for which the curves $C_q \subset V$ are non-singular.
When does such a family *continuously* extend to all of V with *no assumption* of having a holomorphic member among them?
If so, this would imply that every $f_0 : V \to W$ is homotopic to a pluriharmonic map, as in the case $K_{\mathbb{C}} \leq 0$.
(b) Can the curve deformation completeness property (or some slight modification of it) be expressed *algebraically* for algebraic manifolds W, i.e. is it invariant under the action of the Galois group?
Notice that the main (but not only) source of (known) $\mathcal{C}$-deformation complete manifolds are arithmetic varieties, the class of which *is* Galois invariant by a theorem of Kazhdan.
So we ask whether Kazhdan's theorem extends to the class of curve deformation complete manifolds.
Also this his question may be asked about $rank_{hmt/hol}(W)$: is it an algebraic invariant for algebraic manifolds W?
(c) Another question motivated by Kazhdan's theorem is as follows. Let W be an arithmetic variety, (possibly one has to assume it admits positive solution to the *congruence problem*) let V be a smooth projective manifold (defined over a number field?) and let h be a homomorphism of the profinite completion $\overline{\pi}_1(V)$ of the fundamental group of V to $\overline{\pi}_1(W)$.
Does there exist, under a suitable lower bound on some "topological rank" of h, a regular map f from V to a Galois transform of W (kind of ±-holomorphic, but now discontinuous, map), such that the homomorphism $\overline{\pi}_1(V) \to \overline{\pi}_1(W)$ induced by f equals h up to a Galois automorphism of $\overline{\pi}_1(W)$.

(d) There is a purely algebraic counterpart to $rank_{hmt/hol}(W)$ which is defined as follows.

Let $p : U \to B$ be a surjective holomorphic map between non-singular algebraic varieties and $V_b = p^{-1}(b) \subset U$ be (necessarily non-singular) fiber $V_b = p^{-1}(b) \subset U$ over a non-critical point $b \in B$.

Define $rank^{dfm}_{hmt/hol}(W)$ as the minimal number k, such that every holomorphic map $V_b \to W$ of $\mathbb{R}$-rank $> k$ extends to a holomorphic map $U \to W$ for *all* U, B, p and V_b as above (where we do not assume beforehand the existence of any *continuous* map $V \to W$ that extends our holomorphic $V_b \to W$).

It is easy to show that

$$rank^{dfm}_{hmt/hol}(W) \leq rank_{hmt/hol}(W).$$

Are there algebraic manifolds W with $rank^{dfm}_{hmt/hol}(W) < rank_{hmt/hol}(W)$?

(e) Let Y be a symmetric, say Hermitian, space with $K(Y) \leq 0$ and $\Gamma_C = \pi_1(C)$ be a surface group. Let $\mathcal{M}$ be the space of conformal structures on C and $\mathcal{R}$ the space of conjugacy classes of homomorphisms $\Gamma \to iso_{hol}(Y)$.

The minimal energy $E_{min} = E_{min}(\mu, \rho)$ of Γ_C-equivariant maps $\tilde{C} \to Y$, for $\tilde{C} = H^2_{\mathbb{R}}$, is a real analytic function on $\mathcal{M} \times \mathcal{R}$.

What is the algebraic/analytic nature of this function?

What are "natural" PDE satisfied by it?

What is the set of the critical points of $E_{min}(\mu, \rho)$?

Does the set of these functions for variable Y and/or Γ_S carry some meaningful structure?

Can one *effectively* describe pluriharmonic/holomorphic maps from Kähler manifolds into Y/Γ in terms of $E_{min}(\mu, \rho)$?

Example. Let V be a non-singular projective variety, $V \subset \mathbb{C}P^M$, and let $C \subset V$ be a "generic mobile" curve, e.g. the intersection of V with an M'-plane $\mathbb{C}P^{M'} \subset \mathbb{C}P^M$, $M' = M - dim(V) - 1$, in general position.

If W is a cdc (curve deformation complete) manifold (e.g. W is $[\mathrm{Tol}]_{conv}$), which also has SDP (split deformation property), and if W contains no rational curve (e.g. $K(W) \leq 0$), then

every continuous map $f_0 : V \to W$ which restricts to a holomorphic map on C is homotopic to a holomorphic map $V \to W$.

Furthermore let cdc be relaxed to $rank^{dfm}_{hmt/hol}(W) \geq 2k$ and let $V_0 \subset V$ be the intersection of $V \subset \mathbb{C}P^M$ with a generic M_0-plane $\mathbb{C}P^{M_0} \subset \mathbb{C}P^M$ for $M_0 > M - dim(V) + k$.

Then every holomorphic map $V_0 \to W$ of $\mathbb{C}$-rank $> k$ extends to a holomorphic map $V \to W$.

Questions.

1. We assumed from time to time that certain manifolds (varieties) were algebraic. Was it truly needed or would "Kähler" suffice?

2. Let W be a closed Riemannian $[\mathrm{Tol}]_{conv}$ manifold and suppose it receives a continuous map f_0 from an algebraic (Kählerian?) manifold V, such that $rank_{hmt}[f_0] = dim(W)$.
   Is (the Riemannian metric on) W "essentially" Kählerian?
   Can one non-trivially deform the Riemannian metric on a locally symmetric Kählerian W with $K_{\mathbb{C}} \leq 0$ keeping it $[\mathrm{Tol}]_{conv}$?
   For example, does the n-torus admit a non-flat $[\mathrm{Tol}]_{conv}$-metric?
3. Which "slowly growing" harmonic maps $X \to Y$ are pluriharmonic?
   For instance, are *Lipschitz* harmonic functions $Y \to \mathbb{R}$ on Abelian coverings Y of compact Kähler manifolds pluriharmonic?
4. Do non-trivial lower bounds on $rank_{hmt/hol}$ or $rank^{dfm}_{hmt/hol}$, e.g. the strongest $rank_{hmt/hol}(W) = 2$ or the weakest $rank^{dfm}_{hmt/hol}(W) < dim_{\mathbb{R}}(W)$, say for projective algebraic manifolds W, imply that the universal covering $\tilde{W}$ of W is *contractible*?
5. Conversely, let W be a compact Kähler manifold W with *contractible* $\tilde{W}$. Does it satisfy $rank_{hmt/hol}(W) < dim_{\mathbb{R}}(W)$ or at least $rank^{dfm}_{hmt/hol}(W) < dim_{\mathbb{R}}(W)$?
   Here one has to keep in mind certain exceptions/modifications. e.g. for $\pi_1(W) = \Gamma_0 \ltimes \mathbb{Z}^k$ as in Sect. 4.9.
   On the other hand, there are, apparently, no examples of Kähler manifolds W with *contractible* $\tilde{W}$ and *word hyperbolic* fundamental group $\pi_1(W)$ where one would be able to show that $rank_{hmt/hol}(W) > 2$.
6. Can you tell in terms of Γ when a compact Kähler space W with $\pi_1(W) = \Gamma$ and $rank_{hmt/hol} < dim_{\mathbb{R}}(W)$ is non-singular?
   When does a complete Kähler manifold W with $rank_{hmt/hol} < dim_{\mathbb{R}}(W)$ contain a compact (or just finite dimensional for $dim(W) = \infty$) "core" , i.e. an analytic subspace $W_0 \subset W$ (similar the above $\mathcal{B}$ in the moduli space of curves) such that the inclusion $W_0 \subset W$ is a homotopy equivalence, or, at least, such that the inclusion homomorphism $\pi_1(W_0) \to \pi_1(W)$ is an isomorphism?

For instance, let Γ admit a proper discrete (free?) action on a given (some?) infinite dimensional symmetric space Y.

Granted such an action when does Y/Γ admit a finite dimensional (compact?) complex analytic "core" $W_0 \subset W$?

There are few examples of "interesting" isometric group actions on infinite dimensional symmetric spaces Y with $K(Y) \leq 0$. The most studied are *Haagerup's a-T-menable* groups that *properly* act on $Y = \mathbb{C}^\infty$; these, according to A. Valette, include all *amenable*, e.g. solvable groups.

These actions are used to derive the following constrains on the fundamental groups $\Gamma = \pi_1(W)$ of compact Kähler manifold W.

if Γ is $1/6$-small cancelation group, then it contains a surface group of finite index [13],

and

if Γ is solvable then it contains a nilpotent subgroup of finite index [12] (where it remains unknown which nilpotent groups can serve as $\Gamma = \pi_1(W)$).

Also, there are interesting isometric actions of $PSL_2(\mathbb{R})$ and of the Cremona group $Bir(\mathbb{C}P^2)$ on $H^\infty_{\mathbb{R}}$ [14].

4.7 Algebra-Geometric Abel-Jacobi-Albanese Construction

The following classical construction of the Jacobian $W = A(V)$ of a projective algebraic variety V is vaguely similar to the combinatorial reconstruction of Shub-Franks group actions (see Sect. 2.4) where orbits of an action come as equivalence classes of quasi-orbits modulo $DIST < \infty$ equivalence relation.

Let $Z_0(V)$ be the group of 0-*cycles in* V that are formal integer combinations $\sum_{v \in V} m_v$, for functions $m : V \to \mathbb{Z}$ with *finite* supports. and Observe two tautological maps ± : $V \to Z_0(V)$ for $v \to \pm v = \pm 1 \cdot v$.

Every holomorphic map $\alpha : V \to A$ to a commutative algebraic group A extends, via the summation in A, to $\alpha^+ : Z_0(V) \to A$.

Assume that the image $\alpha(V) \subset A$ generates A as a group. Then α^+ is *onto* and in order to reconstruct A from V it remains to identify the equivalence relation on $Z_0(V)$ which reduces $Z_0(V)$ to A.

This can be equivalently seen in terms of the symmetric powers $V^{\frac{N}{sym}} = V^N/\Pi(V)$ of V (for the permutation group $\Pi(N)$ acting on the Cartesian power V^N) as follows.

Extend the above ±maps $V \to Z_0(V)$ by summation in $Z_0(V)$ to the corresponding maps $\pm_N : V^{\frac{N}{sym}} \to Z_0(V)$ and compose these maps with α^+.

Thus, for every pair of integers (N_+, N_-), we obtain a map $\alpha^{N_\pm} : V^{\frac{N_+}{sym}} \times V^{\frac{N_-}{sym}} \to A$, that are

$$[(v_1, \dots, v_{N_+}), (v'_1, \dots, v'_{N_-})] \to v_1 + \dots + v_{N_+} - v'_1 - \dots - v'_{N_-}.$$

which are onto for large $N_\pm$ and try to identify the fibers of these maps.

A rational curve in $Z_{v_0}(V) = Z_0(V)/\mathbb{Z}v_0$, is the image of a *non-constant holomorphic* map φ from the projective line $P^1 = \mathbb{C}P^1$, where "holomorphic" means that $\varphi : P^1 \to Z_0(V)$ descends from a pair of (ordinary) holomorphic maps $\varphi_\pm : P^1 \to V^{\frac{N_\pm}{sym}}$ for some (large) integers $N_\pm$ via the summation map on 0-cycles, $\pm_{N_\pm} : V^{\frac{N_+}{sym}} \times V^{\frac{N_-}{sym}} \to Z_0(V)$.

Two zero cycles z_1 and z_2 in $Z_0(V)$ are called *rationally equivalent* if they can be joined by chain of *rational curves* between them and the corresponding quotient space is denoted by $Z_0(V)/\sim_{rat}$

Metrically speaking, let $DIST(z_1, z_2) = DIST_{rat}(z_1, z_2)$ be the length of the shortest chain of rational curves between z_1 and z_2 (e.g. $DIST(x_1, z_2) \leq 1$ if and only if z_1 and z_2 lie on a rational curve in $Z_0(V)$) and rewrite $Z_0(V)/\sim_{rat} = Z_0(V)/[DIST_{rat} < \infty]$.

Since A contains no rational curve, every map $Z_0(V) \to A$ factors via a map $Z_0(V)/\sim_{rat} \to A$.

If V is a *curve*, i.e. $dim_{\mathbb{C}}(V) = 1$, then, classically, $Z_0(V)/rat$ equals the Jacobian $A(V) = H_1(V;\mathbb{R})/H_1(V;\mathbb{Z})$.

In fact, the map $\alpha^{N_+} : V^{\frac{N_+}{sym}} \to A(V)$ associated to an Abel-Jacobi map $\alpha : V \to A(V)$ (which is defined up-to translations in $A(V)$) is *onto* for $N_+ \geq \frac{1}{2} rank(H_1(V))$ and if $N_+ >> rank(H_1(V))$, in fact, $N_+ > rank(H_1(V))$ suffices, then the fibers of α^{N_+} are complex projective spaces.

(This agrees with algebraic topology: the homotopy types of the symmetric powers $V^{\frac{N}{sym}}$, for any V, converge as $N \to \infty$ to the product of the *Eilenberg-MacLane spaces*, $K(H_1(V), 1) \times K(H_2(V), 2) \times \ldots \times K(H_{2m}(V), 2m)$ for $m = dim_{\mathbb{C}}(V)$ by the *Dold-Thom theorem*.)

However if $dim_{\mathbb{C}}(V) \geq 2$, then the quotient space $Z_0(V)/\sim_{rat}$ may be much greater than the Albanese variety $A(V)$, in fact $Z_0(V)/\sim_{rat}$ is *infinite dimensional* in most cases (see [66]). I wonder if the rate of growth of the dimension of $V^{\frac{N}{sym}}/\sim_{rat}$ for $N \to \infty$ has ever been looked at. Also, the geometry of $(Z_0(V), DIST_{rat})$ may be interesting.

Possibly, one can remedy this by limiting $Z_0(V)$ to a certain subspace $Q_0(V) \subset Z_0(V)$ (similar to quasi-orbits in the Shub-Francs case) and/or by strengthening the $\sim_{rat}$ relation. For example, instead of chains of rational curves, one may try chains of *surfaces* $S \subset Z_0(V)$ *such that* $H_1(S;\mathbb{R}) = 0$, but this does not look pretty.

Traditionally, if $V \subset \mathbb{C}P^M$ is projective algebraic, one makes $A(V)$ from $A(C)$ for generic curves $C \subset V$, that are intersections of V with M'-planes $\mathbb{C}P^{M'} \subset \mathbb{C}P^M$, $M' = M - dim(V) - 1$, in general position, since

the Albanese variety $A(V)$ *equals the maximal common quotient space (Abelian variety) of the Jacobians* $A(C)$ *of generic curves* $C \subset V$.

(The suitably augmented category of the Abelian varieties that are Jacobians of *all* non-singular curves in V seems a nice comprehensive invariant of V; it, probably, has been studied by algebraic geometers, but I am not an expert.)

To see this, let $f_C : C \to A$ be a holomorphic map from a generic $C \subset V$ to a flat Kählerian torus A. Such a map extends to a continuous map $f : V \to A$ if and only if the homology homomorphism $(f_C)_* : H_1(C) \to H_1(A)$ factors as $H_1(C) \overset{emb_*}{\to} H_1(V) \overset{h}{\to} H_1(A)$ for some h, where emb_* is the inclusion homomorphism for $C \subset V$.

Granted such an f, we restrict it to all curves $C_g = V \cap \mathbb{C}P_g^{M'}$ in V, for $V \subset \mathbb{C}P^M \supset \mathbb{C}P_g^{M'}$ which are obtained by varying M'-planes in $\mathbb{C}P^M \supset V$ (parametrized by the corresponding Grassmannian $G \ni g$) passing through a given point $v_0 \in C \subset V \subset \mathbb{C}P^M$, where we want this v_0 to go to $0 \in A(V)$.

Deform the resulting maps $C_g \to A$ to *harmonic* ones, say $f_g : C_g \to A$, all sending $v_0 \to 0 \in A$. These maps, as we know, must be all holomorphic because of $[\text{Tol}]_{conv}$ (see the end of previous section) and, consequently, holomorphically depending on q.

It follows, that if two such curves, say C_{g_1} and C_{g_2} intersect at a point $v \in V$, then $f_{g_1}(v) = f_{g_2}(v)$, since C_{g_1} and C_{g_2} can be joined by a *rational* curve, say P, in the space of all C passing through v_0 and v and, since *every holomorphic* map, such as $p \mapsto f_p(v)$, from a rational P to A is constant, these f_g define a holomorphic map $V \to A$.

However, it is not apparent (at least to the author) how to see in a simple geometric way (without using Hodge theory or Picard varieties) that the maximal common factor (or, rather, the maximal quotient space) of the Jacobians $A(C)$ has the $\mathbb{C}$-dimension equal one half of the rank of $H_1(V)$.

4.8 Deformation Completeness and Moduli Spaces

Let us look from a similar perspective at other curve deformation complete, e.g. $[\mathrm{Tol}]_{conv}$, spaces W, such as $W = H^m_{\mathbb{C}}/\Gamma$ for a free discrete undistorted (e.g. cocompact) group $\Gamma \subset iso_{hol}(H^m_{\mathbb{C}})$.

Let C be a non-singular projective algebraic curve (Riemann surface) of positive (preferably large) genus and $\alpha_C : C \to W$ be a non-constant holomorphic map.

Such a map may be non-unique in its homotopy class, as in the above case of flat Kählerian tori and then we need to normalize this map as we did it with SDP in Sect. 4.6. This, however, is a minor issue and we assume α_C *is unique in its homotopy class* as for $W = H^m_{\mathbb{C}}/\Gamma$.

Denote by $\mathcal{M} = \mathcal{M}(C)$ the *Riemann moduli space* of deformation of this curve C and let $\mathcal{C} \to \mathcal{M}$ be the "universal curve" over $\mathcal{M}$ that is the space of C with marking points $c \in C$.

Let $b_C \in \mathcal{M}$ be the point corresponding to C and denote by $\mathcal{B} = \mathcal{B}(C, W)$ the union of all algebraic curves $B \subset \mathcal{M}$, such that $B \ni b_C$ and such that the map $\alpha_C : C \to W$ extends to a *continous* map α_S of the complex surface $S = \mathcal{C}|B$ over B, that is the restriction of the universal curve to B, to W.

Recall that these *continuous* maps $\alpha_S : S \to W$ are homotopic to *holomorphic* ones $\alpha_S^{hol} : S \to W$ in the curve deformation complete case and we additionally assume these are unique.

Denote by $\mathcal{D} = \mathcal{D}(C, W) \subset \mathcal{C}$ the universal curve restricted to $\mathcal{B} \subset \mathcal{M}$ and denote by $\mathcal{A} : \mathcal{D} \to W$ the map compiled by α_S^{hol} for all above $B \subset \mathcal{M}$ and S over B.

(Since $\mathcal{M}$ is an *orbifold*, rather than a manifold, where the orbi-singular points in $\mathcal{M}$ correspond to curves with non-trivial symmetries, one should, to be faithful to the truth, pass to the universal orbi-coverings $\tilde{\mathcal{M}}$ take the corresponding $\tilde{\mathcal{C}} \to \tilde{\mathcal{M}}$ and to formulate everything in terms of equivariant maps from $\tilde{\mathcal{C}}$ to the universal covering of W. And if one wants to stay in the algebraic category, one may use finite coverings of sufficiently high level.)

Notice that for $[\mathrm{Tol}]_{conv}$ manifolds W the subspace $\mathcal{B}$ equals the minimum set of the ($\mathbb{C}$-convex) minimal energy function $b \mapsto E_{min}(C_b)$, $b \in \mathcal{M}$, for maps from the fibers C_b of the universal curve into W.

Also observe that if W is finite dimensional, then $\mathcal{B} \subset \mathcal{M}$ is a $\mathbb{C}$-*algebraic* subvariety in the moduli space of curves and that $\mathcal{A} : \mathcal{D} \to W$ is a holomorphic map.

Let $W \subset \mathbb{C}P^M$ be projective algebraic, apply the above to a generic curve $C = W \cap \mathbb{C}P^{M'}$, $M' = M - m + 1, m = dim_{\mathbb{C}}(W)$, and observe that the closure of $\mathcal{B}$ in a suitably compactified $\mathcal{M}$ equals, in this case, the Grassmann manifold $G_k(\mathbb{C}P^M)$ of k-planes $\mathbb{C}P^k \subset \mathbb{C}P^M$ and that the fibers of the map $\mathcal{A} : \mathcal{B} \to W$ are *rational* varieties.

The essential specific feature of our W, besides the (assumed) injectivity of the tautological classifying map $G_k(\mathbb{C}P^M) \to \mathcal{M}$ for $g \mapsto C_g = \mathbb{C}P^k_g \cap W$ (defined on the Zariski open subset in $G_k(\mathbb{C}P^M)$ corresponding to non-singular curves) is that the image $\mathcal{B}$ of this map has a "semitopological" description, being the *maximal* algebraic subset in $\mathcal{M}$, such that $\mathcal{D} \subset \mathcal{C}$ over it admits a *continuous* map to W extending this from some curve $C \in \mathcal{D}$.

Questions. Can one *intrinsically,* in terms of the moduli spaces $\mathcal{M}$ of curves of *all* genera, relate these $\mathcal{B}$ and the maps $\mathcal{A} : \mathcal{D} \to W$ associated to different projective embeddings of W?

Is the *canonical embedding* $W \to \mathbb{C}P^{M_h - 1}$ associated to the space $\mathcal{H}_m$ (of dimension M_h) of holomorphic m-forms on W is of a particular interest in this picture?

Can all this (which is just a reformulation of Sui-Sampson-Carlson-Toledo theorem) be actually used for *(re)construction* of spaces like $H^m_{\mathbb{C}}/\Gamma$?

"Internalization" of Deformation of Surfaces. One can do a little of such "reconstruction" with moduli spaces of surfaces rather than of curves (where, of course, these spaces are not so readily available for sightseeing.)

Namely, let $\mathcal{S} \to \mathcal{M} = \mathcal{M}(S)$ be a "universal surface" where $\mathcal{S}$ and $\mathcal{M}$ are projective algebraic varieties and $\mathcal{P} : \mathcal{S} \to \mathcal{M} = \mathcal{M}(S)$ a surjective holomorphic *parametrization* map, where $\mathcal{S}$ is non-singular over a generic point in $\mathcal{M}$, where generic fiber $S \subset \mathcal{S}$ is a smooth surface and such that *all* (isomorphism classes of) smooth deformations of S are among the fibers of $\mathcal{P} : \mathcal{S} \to \mathcal{M}$.

Let W have $rank^{dfm}_{hmt/hol}(W) = 2$, e.g. $W = H^m_{\mathbb{C}}/\Gamma$. Then, by the very definition of $rank^{dfm}_{hmt/hol}$,

every holomorphic map $S \to W$ extends to a rational map $\mathcal{S} \to W$; moreover if $\mathcal{S}$ is non-singular and W contains no rational curve (which may follow from [Tol]–$conv$) then this rational map is holomorphic.

Notice that universality of $\mathcal{S}$ is non-essential for this statement but is relevant for identification/realization of spaces like $H^m_{\mathbb{C}}/\Gamma$.

For example, (this is significant starting from $dim(W) = 3$) assume $W \subset \mathbb{C}P^M$ is projective and take the intersections of W with all $(k + 1)$-planes $\mathbb{C}P^{M'+1}_g \subset \mathbb{C}P^M$, $g \in G_{M'+1}(\mathbb{C}P^M)$, $M' = M - dim(W) + 1$.

Then these intersection *surfaces* make the *full moduli* spaces $\mathcal{M}$ of these surfaces (i.e. every "abstract deformation" of $S = W \cap \mathbb{P}^{k+1}$ is "internal" – it comes from moving the k-plane $\mathbb{C}P^{k+1} \subset \mathbb{C}P^M$) and W comes as a quotient of the universal surface $\mathcal{S}$ over this $\mathcal{M}$ where the quotient map $\mathcal{S} \to W$ has rational fibers.

As we mentioned earlier, this "internalization" is a purely algebraic property which makes sense for varieties over an arbitrary field, but it is unclear, for example, in what form it holds (if at all) for mod p reductions of $W = H^m_{\mathbb{C}}/\Gamma$.

Questions. Let the inclusion homomorphism $\pi_1(S) \to \pi_1(W)$, for a non-singular surface $S \subset W = H^m_{\mathbb{C}}/\Gamma$, is an isomorphism.

Is such an S "mobile" in the sense that its deformations cover all of W?

Can *mobile generic* surfaces in other locally symmetric Hermitian spaces W without flat factors have "external" deformations?

(The probable answer is "No", which may (?) lead to examples where $rank^{dfm}_{hmt/hol} < rank_{hmt/hol}$)

A related group of questions is motivated by the concept of *Kähler hyperbolicity* (see [6, 31].)

Let Γ be a finitely presented group with even dimension $dim_{\mathbb{Q}hmt}(\Gamma) = 2m$ and such that $H^{2m}(\Gamma;\mathbb{Q}) = \mathbb{Q}$. Let $\kappa \in H^2(\Gamma;\mathbb{R})$ be a *hyperbolic* "Kähler" class, i.e. $\kappa^m \neq 0$ and where "hyperbolic" means that it is representable as the differential (coboundary) of a *bounded* (non-invariant) 1-cochain. For example, if Γ is *word hyperbolic*, then all 2-cocycles are hyperbolic.

Basic Questions. When does there exist a Kähler space Y with a discrete isometric action of Γ, such that the Kähler class of Y corresponds (in an obvious way) to κ?

If such spaces Y do exist, can one evaluate $rank_{hmt/hol}(Y/\Gamma)$ and identify those where this rank is minimal?

One knows (see [31]) that if such a Y is finite dimensional, topologically contractible and non-singular and if the quotient space $W = Y/\Gamma$ is compact, (probably these conditions can be significantly relaxed), then Y supports a huge Hilbert space $\Omega^m_{L_2} = \mathbb{C}^\infty$ of square integrable holomorphic m-forms, $m = dim_{\mathbb{C}}(Y)$, where the Γ-equivariant Bergman evaluation map $B : Y \to \mathbb{C}P^\infty$ (for this $\mathbb{C}P^\infty$ being the space of hyperplanes in $\mathbb{C}^\infty = \Omega^m_{L_2}$) is, generically, one-to-one.

When does the Bergman metric on Y induced by B (from the Fubini-Studi metric on $\mathbb{C}P^\infty$) has $K_{\mathbb{C}} \leq 0$?

Is there an *effective* criterion (criteria) expressible *entirely in terms of the group* Γ for the Γ*-invariant* $2m$-dimensional homology class in this $\mathbb{C}P^\infty$, corresponding to the B_*-image of the Γ*-invariant* fundamental class $[Y]_\Gamma$ of Y, (that is, essentially, the same as $[W] \in H^{2m}(W)$) to be representable by an m-dimensional Γ-invariant complex analytic subvariety $Y' \subset \mathbb{C}P^\infty$ with compact quotient $W' = Y/\Gamma$?

The first difficulty to overcome is to identify the Hodge component $\Omega^m_{L_2}$ of holomorphic L_2-forms in the full L_2-cohomology $H^m_{L_2}(Y;\mathbb{C})$ in terms of Γ *alone*.

Then you, probably, need to pinpoint the correct Γ-invariant Hermitian-Hilbert metric on the linear space $\mathbb{C}^\infty = \Omega^m_{L_2}$ in order to have a useful metric on our $\mathbb{C}P^\infty$.

Granted this, you can formulate the necessary and sufficient condition in term of the *minimal* $2m$*-volume* of the homology class corresponding $[Y]_\Gamma$. but it remains unclear how to compute this volume in specific examples.

Alternatively, one can consider *all* (suitably homologically normalized) Hermitian-Hilbert metrics HH on $H^m_{L_2}(\Gamma;\mathbb{C})$ and *minimize* this minimal $2m$-volume in the corresponding projective space (acted upon by Γ) over all HH.

But it remains highly problematic how to compute these minimal volumes in specific examples.

A more general and promising class of questions is as follows:

Which complex invariants of a Kähler manifold W "essentially" depend on the fundamental group $\Gamma = \pi_1(W)$?

For example, is there a Kählerian counterpart to *Novikov's higher signature conjecture*?

Namely, let $W \to K(\Gamma, 1)$ be the Eilenberg-MacLane classifying map and $h_{2i} \in H^{2i}(W;\mathbb{Q})$ be a cohomology class coming from $H^{2i}(K(\Gamma,1);\mathbb{Q})$ via this map. Let $c_{n-i} \in H^{2n-2i}(W;\mathbb{Q})$, $n = dim_{\mathbb{C}}(W)$, be a linear combinations of product of certain Chern classes of W.

Does the value $(h_{2i} \smile c_{n-i})[W]$, for the fundamental class $[W] \in H_{2n}(W) = \mathbb{Z}$, actually depend on the complex structure in W or only on the homotopy type of W?

If not, what should one add to make it true?

For example, do the Chern numbers of a W with *contractible* universal covering depend *only* on $\pi_1(W)$?

4.9 On Kählerian and Hyperbolic Moduli Spaces

The Abel-Jacobi-Albanese construction needs *a choice* of a complex structure in the target torus covered by $\mathbb{C}^n$. This can be compensated by considering the moduli space $\mathcal{B}_n$ of all such structures, where, however, some caution is needed, since some complex tori, e.g. $\mathbb{C}^n/\mathbb{Z}^{2n}$ for $n \geq 2$, admit *infinite* groups of complex automorphisms.

Accordingly, the moduli space of the complex $2n$-tori A, that is an orbispace which is locally at a point $b_0 = b_{A_0}$ corresponding to A_0 equals the quotient of a complex analytic space of deformations of the complex structure in A_0 by the automorphism group of A, has a pretty bad singularity at this b_0.

To remedy this, one fixes a *polarization* i.e. a translation invariant non-singular 2-form ω_0 on a torus and considers the moduli space of the isomorphism classes of the invariant complex structures where this ω_0 serves as the imaginary part of an invariant Hermitian metric.

The resulting moduli space $\mathcal{B}_n$ of Kählerian tori is a non-compact locally symmetric Hermitian orbi-space of finite volume, that is a quotient of a Hermitian symmetric space Y by a discrete isometry group $\Gamma = \Gamma_n = \pi_1^{orbi}(\mathcal{B}_n)$, also denoted Γ_Y.

These tori A themselves, parametrized by $\mathcal{B}_n$, make the *universal family*, say $\mathcal{A}_n \to \mathcal{B}_n$ where the fibers represent the isomorphism classes of all these A.

The Abel-Jacobi-Veronese theorem says in this language that

every continuous map map from a compact Kähler manifold V to a fiber $A_{b_0} \subset \mathcal{A}_n$, $b_0 \in \mathcal{B}$, which induces an isomorphism $H_1(V)/torsion \to H_1(A_{b_0})$, is homotopic to a holomorphic map $V \to \mathcal{A}$ with the image in a single fiber A_b, $b = b(V) \in \mathcal{B}_n$.

The universal orbi-covering space $\tilde{\mathcal{A}}_n$ of $\mathcal{A}_n$ has a natural structure of a holomorphic vector bundle over Y, where this bundle carries a Γ invariant flat $\mathbb{R}$-linear (but not $\mathbb{C}$-linear) connection.

On the other hand, since Y is *topologically contractible* as well as *Stein*, this bundle is holomorphically (non-canonically) isomorphic to the trivial bundle $Y \times \mathbb{C}^n$.

The Galois group $\Gamma_{\mathcal{A}}$ of the covering map $\tilde{\mathcal{A}}_n \to \mathcal{A}_n$ is the semidirect product $\Gamma \ltimes \mathbb{Z}^{2n}$, for the *monodromy action* of Γ on $\mathbb{Z}^{2n}$ and where the action of $\Gamma_{\mathcal{A}}$ on $\tilde{\mathcal{A}}_n = Y \times \mathbb{C}^n$ is $\mathbb{C}$-*affine* on the fibers $y \times \mathbb{C}^n \subset \tilde{\mathcal{A}}_n$

This monodromy action Γ on $\mathbb{Z}^{2n}$ is of the kind we met in Sect. 2.3 (where we discussed/conjectured the super-stability of such actions) and the Siu theorem for equivariant map $X \to Y$ can be reformulated in "dynamics" terms as well.

Namely, let X be a complex analytic space with a discrete action of $\Gamma_X \subset iso_{hol}(X)$ and $f : X \to Y$ be an h-equivariant continuous map for a homomorphism $h : \Gamma_X \to \Gamma = \Gamma_Y$. Regard the trivial bundle $X \times \mathbb{C}^n \to X$ as that induced from $\tilde{\mathcal{A}}_n \to Y$ and, thus, lift the action of Γ_X on X to a continous fiber-wise $\mathbb{C}$-linear action of Γ_X on $X \times \mathbb{C}^n \to X$.

If the map f is holomorphic, then so is this lifted action, and, whenever Siu theorem applies, the continuous action of Γ_X on $X \times \mathbb{C}^n \to X$ is equivariantly homotopic to a holomorphic action.

(The moduli space $\mathcal{B}_n$ and its finite orbi-covers contain lots of compact Hermitian totally geodesic subspaces, e.g. some W locally isometric the complex hyperbolic spaces $H^m_{\mathbb{C}}$. The Siu theorem applies, for example, if the image of h is contained in the fundamental (sub)group of such a W and $rank_{hmt}(f) > 2$.)

Besides the action of Γ_X, the map f induces an action of $\mathbb{Z}^{2n}$ on $X \times \mathbb{C}^n$ by parallel translations in each fiber $x \times \mathbb{C}^n$, where the two actions together define an action of the semidirect product $\Gamma'_X = \Gamma_X \ltimes_h \mathbb{Z}^{2n}$ where Γ_X acts on $\mathbb{Z}^{2n}$ as $h(\Gamma_X) \subset \Gamma_Y$.

If the map f is holomorphic, so is the action of Γ'_X on $X \times \mathbb{C}^n$.

Conversely,

if such an action is holomorphic to start with, it defines, by the universality of $\mathcal{A}_n$, an equivariant holomorphic map $f : X \to Y$.

This reformulation does not seem to help in proving Siu-type properties, but it raises the following

Questions

(a) Are there other instances of holomorphic Γ_X-spaces X, such that a continuous lift of a discrete Γ_X-action from X to a fiber-wise linear and/or fiber-wise affine action on a holomorphic vector bundle over X can be *predictably* deformed to a holomorphic lift?

(b) Let $W' \to W$ be a holomorphic fibration where the fibers are Kählerian tori. Let $f_0 : V \to W'$ be a continuous map, such that its projection to W is (known to be) homotopic to a holomorphic map $V \to W$. (We may additionally insists that this fibration admits a holomorphic section.)
Under what geometric conditions on W' (in particular, concerning the classifying map from W to the moduli space $\mathcal{B}$ of Kählerian tori) and homotopy conditions on f_0 is the map f_0 itself homotopic to a holomorphic map?
(We look for an answer that would embrace the Albanese theorem corresponding to W being a single point along with the Siu theorem where one has to require, in particular, that the homotopy rank $rank_{hmt}(f_0)$ equals that of the projection of f_0 to W.)

(c) The latter question has a counterpart in hyperbolic dynamics that we formulate below in the simplest case.
Let W be a manifold (or orbifold) of negative curvature and let $W' \to UT(W)$ be an n-torus fibration over the unite tangent bundle of W.
Let the geodesic flow on $UT(W)$ lift to an $\mathbb{R}$-action on W' which maps $\mathbb{T}^n$-fibers to fibers and such that the return map $\mathbb{T}^n_u \to \mathbb{T}^n_u$, $u \in G \subset UT(W)$, over every periodic orbit (closed geodesic) G in $UT(W)$, is a linear hyperbolic automorphism of $\mathbb{T}^n_u$.
(An inspiring example of such a W' is the lift of the universal elliptic curve $\mathcal{A}_1$ over the modular curve $W = \mathcal{B}_1$ to $UT(W)$, where the $\mathbb{R}$-action is Anosov hyperbolic and where the return maps $\mathbb{T}^2 \to \mathbb{T}^2$ simultaneously and coherently represent *all* hyperbolic automorphisms of $\mathbb{T}^2$ by monodromy transformations.)
What is the (super)stability range of this $\mathbb{R}$-action on W'?
Namely, let V' be a topological, say compact, space with an $\mathbb{R}$-action and $f_0 : V' \to W'$ be a continuous map. (If W is an *orbi*fold, one has to formulate this in terms of equivariant maps between universal orbicoverings of our orbispaces.)
Under what circumstances is f_0 homotopic to a continuos map $V' \to W'$ which sends $\mathbb{R}$-orbits in V' to $\mathbb{R}$-orbits in W'?

4.10 Symbolic and Other Infinite Dimensional Spaces

Interesting (semi)group actions on compact complex spaces X appear only sporadically, where some of these, e.g. holomorphic actions of $\mathbb{Z}_+$ on the 2-sphere (rational curve), have been extensively studied under the heading of *complex (holomorphic) dynamics*.

On the other hands there are lots of infinite dimensional spaces holomorphically acted upon by pretty large groups. Below is a particular construction of such spaces.

$\underline{\Gamma}$-Power Categories. Let $\mathcal{K}$ be a "geometric" category, e.g. the category of Kähler or complex algebraic spaces, or a category of dynamical systems and let $\underline{\Gamma}$ be a countable group. Define the category $\underline{\Gamma}^{\mathcal{K}}$ by Markovian recipe as follows.

The objects $X \in \underline{\Gamma}^{\mathcal{K}}$ are projective limits of finite Cartesian powers $K^{\underline{\Delta}}$ for $K \in \mathcal{K}$ and finite subsets $\underline{\Delta} \subset \underline{\Gamma}$. These X are naturally acted upon by $\underline{\Gamma}$ and the admissible *finitery morphisms* in our $\underline{\Gamma}$-category are $\underline{\Gamma}$*-equivariant* projective limits of morphisms in $\mathcal{K}$, where such a morphism $F : X = K_1^{\underline{\Gamma}} \to Y = K_2^{\underline{\Gamma}}$ is defined by a single morphism in $\mathcal{K}$, say by $f : K_1^{\underline{\Delta}} \to K_2$ where $\underline{\Delta} \subset \underline{\Gamma}$ is a *finite* (sub)set.

Namely, if we think of $x \in X$ and $y \in Y$ as K_1- and K_2-valued functions $x(\underline{\gamma})$ and $y(\underline{\gamma})$ on $\underline{\Gamma}$ then the value $y(\underline{\gamma}) = F(x)(\underline{\gamma}) \in K_2$ is evaluated as follows:

Translate $\underline{\Delta} \subset \underline{\Gamma}$ to $\underline{\gamma}\underline{\Delta} \subset \underline{\Gamma}$ by γ, restrict $x(\gamma)$ to $\gamma\Delta$ and apply f to this restriction $x|\underline{\gamma\Delta} \in K_1^{\gamma\underline{\Delta}} = K_1^{\underline{\Delta}}$.

In particular, every morphism $f : K_1 \to K_2$ in $\mathcal{K}$ tautologically defines a morphism in $\mathcal{K}^{\underline{\Gamma}}$, denoted $f^{\underline{\Gamma}} : K_1^{\underline{\Gamma}} \to K_2^{\underline{\Gamma}}$, but $\mathcal{K}^{\underline{\Gamma}}$ has many other finitery morphisms in it.

(Apparently, the right setting for $\mathcal{K}^{\underline{\Gamma}}$ is where $\mathcal{K}$ is a *multi-category* and where, in particular, $\underline{\Gamma}$-equivariant transformations of spaces $\mathcal{K}^{\underline{\Gamma}}$ represent certain "branches" of operads in $\mathcal{K}$.)

Moreover, there may exist extra (non-finitery) $\underline{\Gamma}$-equivariant morphisms in $\mathcal{K}^{\underline{\Gamma}}$ for certain categories $\mathcal{K}$. For instance, if $\mathcal{K}$ is the category of topological spaces, then continuos $\underline{\Gamma}$-equivariant maps do not need be, in general, finitery. Also one can naturally define continuos holomorphic maps $K_1^{\underline{\Gamma}} \to K_2^{\underline{\Gamma}}$ for complex analytic K_1 and K_2.

Let us enrich $\mathcal{K}^{\underline{\Gamma}}$ by adding new objects to it defined by "equivariant systems of equations" in $X = K^{\underline{\Gamma}}$, e.g. by $F_1(x) = F_2(x)$ for two morphisms $F_1, F_2 : X \to Y$ for some $Y \in \mathcal{K}'^{\underline{\Gamma}}$. Then introduce quotients of (old and new) objects X by equivalence relations $R \subset X \times X$, that are subobjects in our category.

Most of finite dimensional *questions* in algebraic geometry and complex analysis, as well as in dynamics, e.g. concerning $rank_{hmt/hol}$, superstability, etc., automatically extend to the corresponding $\underline{\Gamma}$-categories, where they acquire a dynamics component coming from the $\underline{\Gamma}$-actions. (See last section in [40] and references therein.)

Furthermore, some finite dimensional *results* pass to $\underline{\Gamma}$-categories by just taking projective limits, and, in rare cases, finite dimensional *proofs* also extend to $\mathcal{K}^{\underline{\Gamma}}$.

But the predominant number of *obviously formulated* problems in $\underline{\Gamma}$-categories still wait their solutions [38].

Sample Questions.

(a) Let $\mathcal{K}$ be the category of complex projective algebraic manifolds, let $X \subset K_1^{\underline{\Gamma}}$ be a subobject and $Y = K_2^{\underline{\Gamma}}$, where $K_2 = H^n_{\mathbb{C}}/\Gamma$.
If the group $\underline{\Gamma}$ is amenable then one can, probably, define $meanrank_{hmt}(F)$ (similarly to *meandim* in [37],) for continuous $\underline{\Gamma}$-equivariant maps $F : X \to Y$, and show that every continuous $\underline{\Gamma}$-equivariant maps $F : X \to Y$ with $meanrank_{hmt}(F) > 2$ is $\underline{\Gamma}$-equivariantly homotopic to a holomorphic map.

(b) An alternative to the $\mathcal{K}^{\underline{\Gamma}}$ setting is that of compact infinite dimensional *concentrated mm-spaces* (see [38]) foliated by Hermitian-Hilbertian manifolds. Can one develop the full fledged non-linear Hodge theory for such spaces and apply this for proving Siu-type theorems in the "maximally extended/completed" Kähler $\underline{\Gamma}$-categories?

(c) Let $\mathcal{K}$ be the category of algebraic varieties defined over $\mathbb{Z}$. Then the $\mathbb{F}_q$-points of an object X in the (extended) category $\mathcal{K}^{\underline{\Gamma}}$ make an ordinary Markov hyperbolic $\underline{\Gamma}$-dynamical system, say $X(\mathbb{F}_q)$.

What are the patterns in the behaviour, say, of the *topological entropies* of $X(\mathbb{F}_q)$ for $q = p^i$ and $i \to \infty$, or, more interestingly, for $p \to \infty$?

(One may expect a satisfactorily general/concise answer for $\underline{\Gamma} = \mathbb{Z}$, but one, probably, needs strong assumptions on X for more general amenable and *sofic* groups $\underline{\Gamma}$.)

(d) The Markovian dynamical $\underline{\Gamma}$-systems $X(\mathbb{F}_{p^i})$ converge in the model theoretic sense to $X(\mathbb{C})$. Is it has anything in common with Markovian presentations in Sects. 2.5 and 3.1?

Is there anything special from the dynamics point of view about the $\underline{\Gamma}$-systems $X(\mathbb{F}_q)$?

Notice that the $\underline{\Gamma}$-spaces $X(\mathbb{F}_{p^i})$ converge in the model theoretic sense to $X(\mathbb{C})$ for $p, i \to \infty$, [36], but it is unclear if there is any link between this and Markovian presentations in Sect. 2.5.

The above $\underline{\Gamma}$-spaces also make sense for continuous, e.g. Lie, groups $\underline{\Gamma}$, where the relevant equations defining "subobjects" are partial differential ones.

A particular instance of such a space is that of holomorphic maps from $\underline{\Gamma} = \mathbb{C}$ to an algebraic manifold, or more generally, an *almost complex* manifold K (see [37, 61]).

The dynamics of the action of $\mathbb{C}$ as well as of the group of $\mathbb{C}$-affine transformations of $\mathbb{C}$, on spaces of holomorphic maps $\mathbb{C} \to K$, e.g. the structure of invariant measures on such spaces can be seen as a part of the *Nevanlinna value distribution theory* (See [24, 61] for the first steps along these lines.)

Concluding Remarks. There are other parallels between complex geometry and dynamics. For example, conjectural lower bounds on the topological entropy and on the full spectrum of *intermediate entropies*, (defined in 0.8.F in[35]) of continuous actions on an X in terms of the corresponding asymptotic invariants of the induced actions on $\pi_1(X)$ (a correct/ definition of such invariants is not so apparent) are vaguely similar to the *Hodge conjecture.*

But the main question remains open:

Is there something more to all these "parallels" than just the universality of the categorical/functorial language?.

Acknowledgements I thank Domingo Toledo who acquainted me with his unpublished $\mathbb{C}$-*convexity* result and also suggested several corrections and improvement in the present manuscript.

I am also thankful to John Franks and Jarek Kwapisz who instructed me on the (relatively) recent development in the theory of hyperbolic, in particular pseudo-Anosov, systems associated with **B**.

References

1. M. Anderson, The Dirichlet problem at infinity for manifolds of negative curvature. J. Differ. Geom. **18**, 701–721 (1983)
2. D. Anosov, *Geodesic Flows on Closed Riemannian Manifolds with Negative Curvature*. Proceedings of the Steklov institute of mathematics, vol. 90 (American Mathematical Society, Providence, 1967)
3. G. Band, Identifying points of pseudo-Anosov homeomorphisms. Fund. Math. **180**(2), 185–198 (2003)
4. M. Barge, J. Kwapisz, Hyperbolic pseudo-Anosov maps a. e. embed into a toral automorpism. Ergod. Theory Dyn. Syst. **26**4, 961–972 (2006). http://www.math.montana.edu/~jarek/Papers/hirsh.pdf
5. R. Bowen, Markov partitions for axiom A diffeomorphisms. Am. J. Math. **91**, 725–747 (1970)
6. M. Brunnbauer, D. Kotschick, On hyperbolic cohomology classes (2008). arxiv.org/pdf/0808.1482
7. J. Carlson, D. Toledo, Harmonic mappings of Kähler manifolds to locally symmetric spaces. Inst. Hautes Études Sci. Publ. Math. **69**, 173–201 (1989)
8. J. Carlson, D. Toledo, Rigidity of harmonic maps of maximum rank. J. Geom. Anal. **3**2, 99–140 (1993)
9. K. Corlette, Harmonic maps, rigidity, and Hodge theory, in *Proceedings of the International Congress of Mathematicians (Zürich, 1994)*, vol. 1 (Birkhäuser, Basel, 1995), pp. 465–471
10. G. Daskalopoulos, C. Mese, A. Vdovina, Superrigidity of hyperbolic buildings. Geom. Funct. Anal. 1–15 (2011). Springer
11. G. Daskalopoulos, L. Katzarkov, R. Wentworth, Harmonic maps to Teichmuller space. Math. Res. Lett. **7**1, 133–146 (2000)
12. T. Delzant, L'invariant de Bieri Neumann Strebel des groupes fondamentaux des variétés kähleriennes. Math. Ann. **348**, 119–125 (2010)
13. T. Delzant, M. Gromov, *Cuts in Kähler Groups*. Progress in mathematics, vol. 248 (Birkhäuser, Basel/Switzerland, 2005), pp. 31–55
14. T. Delzant. P. Py, Kähler groups, real hyperbolic spaces and the Cremona group (2010). arXiv:1012.1585v1
15. S. Donaldson, Twisted harmonic maps and the self-duality equations, Proc. London Math. Soc. **55:1**(3), 127–131 (1987)
16. A. Eskin, D. Fisher, K. Whyte, Quasi-isometries and rigidity of solvable groups. Pure Appl. Math. Q. **3**(4), 927–947 (2007) (Special Issue: In honor of Gregory Margulis)
17. A. Fathi, Homotopical stability of pseudo-Anosov diffeomorphisms. Erg. Theory Dyn. Syst. **10**, 287–294 (1990)
18. D. Fisher, Local rigidity: past, present, future, in *Dynamics, Ergodic Theory and Geometry*. Mathematical sciences research institute publications (Cambridge University Press, Cambridge/New York, 2007), p. 45–98. A p.s file available on http://mypage.iu.edu/$\sim$fisherdm/paperwork.html
19. D. Fisher, T. Hitchman, Cocycle superrigidity and harmonic maps with infinite dimensional targets (2005). arXiv:math/0511666
20. D. Fisher, K. Whyte, Continuous quotients for lattice actions on compact manifolds. Joint with Kevin Whyte. Geome. Ded. **87**, 181–189 2001. A p.s file available on http://mypage.iu.edu/$\sim$fisherdm/paperwork.html

21. J. Franks, *Anosov Diffeomorphisms*. Proceedings of symposia in pure mathematics, vol. 14 (American Mathematical Society, Providence, 1970), pp. 61–93
22. J. Franks, M. Handel, Complete semi-conjugacies for psuedo-Anosov homeomorphisms. (2007). http://www.math.wisc.edu/$sim$oh/FranksHandel.pdf
23. D. Fried, Finitely presented dynamical systems. Erg. Theory Dyn. Syst. **7**, 489–507 (1987)
24. A. Gournay. A Runge approximation theorem for pseudo-holomorphic maps (2010). arXiv: 1006.2005v1
25. H. Grauert, Über modifikationen und exzeptionelle analytische Mengen. Math. Ann. **146**(4), 331–368 (1962)
26. M. Gromov, Hyperbolic manifolds, groups and actions, in *Riemann Surfaces and Related Topics: Proceedings of the 1978 Stony Brook Conference*. Annals of mathematics studies, vol. 97 (Princeton University Press, Princeton, 1981) pp. 183–213
27. M. Gromov, *Volume and Bounded Cohomology*. Institut des hautes études scientifiques (Paris, France).; Publications mathématiques, vol. 56 (I.H.E.S, Le Bois-Marie, Bures s/Yvette, 1982/1983), pp. 5–99
28. M. Gromov, Hyperbolic groups, in *Essays in Group Theory*, Mathematical sciences research institute publications, vol. 8 (Springer, New York, 1987) pp. 75–263
29. M. Gromov, Foliated plateau problem, part I. GAFA **1** 14–79 (1991)
30. M. Gromov, Foliated plateau problem, part II. GAFA **1** 253–320 (1991)
31. M. Gromov, Kähler hyperbolicity and L_2-Hodge theory. J. Differ. Geom. **33**(1), 263–292 (1991)
32. M. Gromov. Stability and pinching, in *Seminari di Geometria*, ed. by M. Ferri (Universitàdegli studi di Bologna, Dipartimento di matematica, Bologna, 1992), pp. 55–99
33. M. Gromov, *Asymptotic Invariants of Infinite Groups*. Geometric group theory (Sussex, 1991), vol. 2, London mathematical society lecture note series, vol. 182 (Cambridge University Press, Cambridge, 1993), pp. 1–295
34. M. Gromov. Positive curvature, macroscopic dimension, spectral gaps and higher signatures, in *Functional Analysis on the Eve of the 21st Century*, Volume II, ed. by S. Gindikin et al., (In honor of the eightieth birthday of I. M. Gelfand). Proceedings conference, Rutgers University, New Brunswick, 24–27 October 1993. Progress in mathematics, vol. 132 (Birkhäuser, Basel, 1996), pp. 1–213
35. M. Gromov, Carnot-CarathŽodory spaces seen from within, in *Sub-Riemannian Geometry*. Progress in mathematics, vol. 144, (Birkhäuser, Basel, 1996), pp. 79–323
36. M. Gromov, Endomorphisms of symbolic algebraic varieties. J. Eur. Math. Soc. **1**(2), 109–197 (1999)
37. M. Gromov, Topological invariants of dynamical systems and spaces of holomorphic maps: I. Math. Phys. Anal. Geom. **2** 323–415 (1999)
38. M. Gromov, Spaces and questions. Geom. Funct. Anal. **Special Volume**(Part I), 118–161 (2000). GAFA 2000
39. M. Gromov, Random walk in random groups. GAFA, Geom. Funct. Anal. **13**, 73–146 (2003)
40. M. Gromov, Manifolds: where do we come from ? What are we? Where are we going? (2010)
41. M. Gromov, R. Schoen, Harmonic maps into singular spaces and p-adic superrigidity for lattices in groups of rank one. Publ. Math. IHES **76**, 165–246 (1992)
42. M. Gromov, G. Henkin, M. Shubin, Holomorphic L2 functions on coverings of pseudoconvex manifolds. Geom. Funct. Anal. **8**(3), 552–585 (1998)
43. J. Hadamard, Les surfaces à courbures opposèes et leurs lignes gèodèsique. J. Math. Pures Appl. **4**, 27–73 (1898)
44. L. Hernandez, Kähler manifolds and 1/4-pinching. Duke Math. J. **62**(3), 601–611 (1991)
45. M. Hirsch, On invariant subsets of hyperbolic sets, in *Essays on Topology and Related Topics* (Springer, New York, 1970), pp. 126–135
46. S. Iitaka, Logarithmic forms of algebraic varieties, J. Fac. Sci. **23**, 525–544 (1976)
47. Y. Ishii, J. Smillie, Homotopy shadowing. Am. J. Math. **132**(4), 987–1029 (2010)
48. J.Jost, Y. Yang, Kähler manifolds and fundamental groups of negatively δ-pinched manifolds. arXiv:math/0312143v1

49. J. Jost, S.-T. Yau, Harmonic maps and superrigidity. Proc. Symp. Pure Math. **54**, 245–280 (1993)
50. J. Jost, S.-T. Yau, A nonlinear elliptic system for maps from Hermitian to Riemannian manifolds and rigidity theorems in Hermitian geometry. Acta Math. **170**(2), 221–254 (1993). Errata in 173 (1994), 307
51. J. Jost, K. Zuo, Harmonic maps of infinite energy and rigidity results for representations of fundamental groups of quasiprojective varieties. Math. Res. Lett. **1**, 631–638 (1994)
52. V. Kaimanovich, H. Masur, The Poisson boundary of the mapping class group. Invent. Math. **125**(2), 221–264 (1996)
53. A. Katok, J. Lewis, R. Zimmer, Cocycle superrigidity and rigidity for lattice actions on tori. Topology **35**, 27–38 (1996)
54. B. Kleiner, B. Leeb, Rigidity of quasi-isometries for symmetric spaces and Euclidean buildings. Inst. Hautes tudes Sci. Publ. Math. **86**, 115–197 (1997)
55. N. Korevaar, R. Schoen, Sobolev spaces and harmonic maps for metric space targets, Commun. Anal. Geom. **1**(3–4), 561–659 (1993)
56. F. Labourie, ÒExistence d'applications harmoniques tordues a valeurs dans les variétes á courbure negativeÒ. Proc. Am. Math. Soc. **111**(3), 877–882 (1991)
57. G. Larotonda, Geodesic Convexity Symmetric Spaces and Hilbert-Schmidt Operators Gabriel Larotonda, These, 2004, Instituto de Ciencias Universidad Nacional de General Sarmiento JM Gutirrez 1150 (1613) Los Polvorines Buenos Aires, Argentina
58. J. Lohkamp, An existence theorem for harmonic maps. Manuscr. Math. **67**(1), 21–23 (1990)
59. G. Lusztig, Cohomology of classifying spaces and Hermitian representations. Represent. Theory Electron. J. Am. Math. Soc. **1**, 31–36 (1996)
60. G. Margulis, N. Qian, Rigidity of weakly hyperbolic actions of higher real rank semisimple Lie groups and their lattices. Ergodic Theory Dyn. Syst. **21**, 121–164 (2001)
61. S. Matsuo, M. Tsukamoto, Instanton approximation, periodic ASD connections, and mean dimension. arXiv:0909.1141
62. I. Mineyev, N. Monod, Y. Shalom, Ideal bicombings for hyperbolic groups and applications. arXiv.org ¿ math ¿ arXiv:math/0304278
63. N. Mok, Harmonic forms with values in locally constant Hilbert bundles, in *Proceedings of the Conference in Honor of Jean-Pierre Kahane (Orsay, 1993)*. J. Fourier Anal. Appl. (Special Issue), 433–453
64. N. Mok, Y. Siu, S. Yeung, Geometric superrigidity. Invent. Math. **113**, 57–83 (1993)
65. H. M. Morse, A one-to-one representation of geodesics on a surface of negative curvature. Am. J. Math. **43**(1), 33–51 (1921). JSTOR
66. D. Mumford, Rational equivalence of 0-cycles on surfaces. J . Math. Kyoto Univ. **9**, 195–204 (1969)
67. V. Nekrashevych, Symbolic dynamics and self-similar groups (Preprint).
68. P. Pansu, Mtriques de Carnot-Carathodory et quasiisomtries des espaces symtriques de rang un. J. Ann. Math. **129**(1), 1–60 (1989)
69. B. Popescu, Infinite dimensional symmetric spaces Inaugural-Dissertation zur Erlangung des Doktorgrades der Matematisch-Naturwissenschaftlichen Fakultat der Universitat Augsburg vorgelegt von Augsburg 2005
70. Th. M. Rassias, Isometries and approximate isometries. Int. J. Math. Math. Sci. **25**, 73–91 (2001)
71. E. Rips, Subgroups of small cancellation groups. Bull. Lond. Math. Soc. **14**(1), 45–47 (1982)
72. J. Sampson, Applications of harmonic maps to Kähler geometry, in *Complex Differential Geometry and Nonlinear Differential Equations, Brunswick, 1984*. Contemporary Mathematics, vol. 49 (American Mathematical Society, Providence, 1986), pp. 125–134
73. B. Schmidt, Weakly Hyperbolic Actions of Kazhdan Groups on Tori. arXiv:math/0511314
74. G. Schumacher, Harmonic maps of the moduli space of compact Riemann surfaces. Math. Ann. **275**(3), 455–466 (1986)
75. J.-P. Serre, *Groupes algebriques et corps dc classes*. Actualités scientifiques et industielles (Hermann, Paris, 1959)

76. M. Shub, Endomorphisms of compact differentiable manifolds. Am. J. Math. **XCI**, 175–199 (1969)
77. Y. Sinai, Markov partitions and C-diffeomorphisms. Funct. Anal. Appl. **2**(1), 64–89 (1968). Funkts. Anal. Prilozh. **2**(1), 64–89 (1968)
78. Y.T. Siu, The complex-analyticity of harmonic maps and the strong rigidity of compact Kähler manifolds. Ann. Math. (2) **112**(1), 73–111 (1980)
79. S. Smale, Finding a horseshoe on the beaches of Rio. Math. Intell. **20**(1), 39–44 (1998)
80. D. Toledo, Rigidity theorems in Kähler geometry and fundamental groups of varieties, in *Several Complex Variables*. MSRI Publications, (Berkeley, CA, 1995–1996), vol. 37 (Cambridge University Press, Cambridge, UK/New York, 1999), pp. 509–533
81. D. Toledo, Energy is Plurisubharmonic. http://www.math.utah.edu/$\sim$toledo/energy.pdf
82. C. Voisin, On the homotopy types of Kähler manifolds and the birational Kodaira problem. J. Differ. Geom. **72**, 43–71 (2006)
83. M. Wolf, Infinite energy harmonic maps and degeneration of hyperbolic surfaces in moduli spaces. J. Differ. Geom. **33**, 487–539 (1991)
84. K. Wortman, A finitely presented solvable group with small quasi-isometry group. Mich. Math. J. **55**, 3–24 (2007)
85. R. Zimmer, Actions of semisimple groups and discrete subgroups, in *Proceedings of the International Congress of Mathematicians, Berkeley* 1247–1258 (1986)

A Smooth Multivariate Interpolation Algorithm

John Guckenheimer

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract This brief paper introduces an algorithm for smooth interpolation of a multivariate function. The input data for the algorithm consists of a set of function values and derivatives up to order r on a finite set of points E. The algorithm utilizes the Voronoi diagram of E to construct a partition of unity that isolates the points of E. Averaging locally defined functions with the partition of unity yields the interpolating function. There are no special cases; the algorithm treats all input data uniformly. The paper describes a test problem, the computation of a surface that is defined as the level set of a function of three variables. This test demonstrates that the algorithm produces approximations whose order corresponds to the degree of the derivatives in the input data.

1 Introduction

This paper is dedicated to Steve Smale, my PhD advisor. Steve has been a continuing inspiration to me throughout my career: his influence is apparent in this work. Numerous times, Steve investigated carefully chosen problems in areas that were new to him at the time. His approach has been to think about a subject from first principles and reshape it himself – typically with a perspective that reflects his mathematical roots in differential topology. Emulating his boldness, I tackle here a subject new to me – multivariate interpolation – from first principles. My viewpoint appears simple from the perspective of multivariable calculus, but different from

J. Guckenheimer (✉)
Department of Mathematics, Cornell University, Ithaca, NY, USA
e-mail: jmg16@cornell.edu

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_9, © Springer-Verlag Berlin Heidelberg 2012

those I have found in the numerical analysis literature on the subject. Hopefully, these ideas will lead to further developments of broad applicability.

The problem of multivariate interpolation is closely related to the Whitney extension theorem. A *Whitney field* of degree r on a finite set $E \subset \Re^n$ is a map $\tau_r : E \rightarrow P_r$ where P_r is the space of polynomials $P : \Re^n \rightarrow \Re$ of degree r. We seek a C^∞ function $f : \Re^n \rightarrow \Re$ whose Taylor expansion of degree r at each $x \in E$ is given by $\tau_r(x)$. In the context of manifolds, $\tau_r(x)$ is an r-jet of f at each $x \in E$ and the r-jet of f restricted to E is τ_r. This interpolation problem is a special (trivial) case of the Whitney extension theorem with many solutions, and further criteria will typically be imposed to select among these. For example, Fefferman et al. [2] have investigated the algorithmic construction of f with minimal or near minimal C^{r+1} norms. This paper approaches the problem experimentally, looking for simple constructions that are easily implemented and tested on examples. Criteria of simplicity are not readily captured in formal terms, so specification of what makes for a "good" extension would be premature in these early stages of this research. The criterion used here to assess the quality of interpolants is to begin with a Whitney field that is the restriction of a C^∞ function g to E, and then test how well the constructed interpolant $f = f_E$ approximates g as E and r vary. This begs the question of whether the function g would be chosen as a "good" interpolant, but our choices of g reflect the types of functions we seek to approximate. Fixing g, we would like to find a family of interpolants with the property that $f_E - g$ tends to 0 in the C^{r+1} norm as the mesh diameter of E tends to 0.

The interpolation problem arises in different areas. The immediate motivation here is accurate computation of level surfaces of a map $F : \Re^m \rightarrow \Re^k$. Individual points of such a surface are readily computed with root finding algorithms. Jets can also be determined as solutions of a hierarchy of equations obtained by differentiating F. Our goal is to smoothly blend patches defined by these jets in order to obtain high order accuracy in approximations of a level surface. We accomplish the blending by averaging the functions with carefully selected partitions of unity. Numerical tests indicate that this strategy does yield substantial improvements compared to prevailing methods for representing level surfaces, especially when compact representations are desired. Our tests take input data from Multifario [3], a multiparameter continuation toolbox for computing implicitly defined manifolds. Multifario produces discontinuous, polyhedral approximations to surfaces under investigation. Our methods seek smooth surfaces that give better approximations to the manifold based upon the same mesh.

Interpolation problems arise also in computer graphics. Specifically, many motion capture systems are based upon video tracking markers on a moving object. The surface of the object is reconstructed from the marker measurements. There are two differences from the interpolation problem formulated above: (1) the data is in the form a point cloud in $\Re^3$ that may lie on a folded surface that is not the graph of a function, and (2) many familiar objects are piecewise smooth and have ridges and corners that are an important part of their geometry. Nonetheless, reconstruction of local surface patches is essentially the problem described above. So it is instructive to look to computer graphics for rendering methods that have been developed in that setting. One class of prevailing methods produces *subdivision surfaces* [5]. Iterative piecewise linear refinements of a polygonal mesh converge to a C^1 (in a few cases C^2) surface. Each step of the iteration is based upon using a local stencil to produce new mesh points. The iteration is continued until the mesh spacing is comparable to the resolution of the rendering. Depending upon the scheme, the iterates may only approximate the input data or they may interpolate it exactly. One of the awkward considerations in the subdivision schemes is that the refinement rules depend upon the local connectivity of the mesh: the number of polygons adjacent to each vertex matters. Nonetheless, the methods are fast and readily implemented on graphics processing units. The limited smoothness of these methods results from self similarity: smooth manifolds become more and more linear on shorter length scales. In contrast, our methods rely upon forming smooth averages of patches defined by explicit formulas.

2 Algorithm

This section describes an algorithm for computing a C^∞ function $f : \Re^n \to \Re$ that interpolates a Whitney field τ_r on the finite set E. The Whitney field τ_r consists of $r - jets$ at each of the points of $E = \{x_j\}, 1 \leq j \leq N$. This is used to construct a C^∞ function f that interpolates the data τ_r. Given query points $y_k \in \Re^n$, the algorithm returns output $f(y_k)$.

The construction of f is based upon partitions of unity. Recall that a partition of unity of a C^∞ manifold M subordinate to the open covering U_j of M consists of C^∞ functions $\varphi_j : M \to \Re$ with the following properties:

- $\varphi_j \geq 0$ with support contained in $\bar{U}_j$.
- For each $x \in M$, only a finite number of $\varphi_j(x)$ are nonzero.
- For each $x \in M$, $\sum_j \varphi_j(x) = 1$

If $f_j : M \to \Re$ is a collection of C^∞ functions with the same index set as the partition of unity, $\sum_j \varphi_j(x) f_j(x)$ is the *average* of the f_j with respect to the partition of unity. Note that it suffices that the domain of f_j contains U_j and that the average is a C^∞ function.

We construct a partition of unity indexed by the set E so that x_k is in the support of φ_j if and only if $k = j$. (For brevity, we write $\varphi_{x_j} = \varphi_j$.) If the functions f_j represent the jets $\tau_r(x_j)$ and are defined in domains that contain the supports of the φ_j, then

$$f(x) = \sum_{j=1}^{N} \varphi_j(x) f_j(x) \tag{1}$$

is a smooth interpolating function for the Whitney field. Here N is the number of points in E. One specific choice for f_j is the polynomial P_j of degree r whose jet at $_j$ is $\tau_r(x_j)$. At the point x_j, the functions $\varphi_k,\ k \neq j$ vanish together with all their derivatives. Thus the only term that contributes to the r-jet of f is $\varphi_j P_j$. Since $\varphi_j = 1 + o(|x - x_j|^r)$, the r-jet of $\varphi_j P_j$ is the r-jet of P_j at x_j.

The key aspect of the construction of f is building a cover that isolates points of E and a partition of unity subordinate to that cover. The first step in this process invokes a standard algorithm [1] to compute the *Voronoi diagram* of E. This tessellation has cells V_j defined to be the points that are closer to $x_j \in E$ than to other points of E. Observe that if we scale V_j by a factor of 2, then the enlarged cell W_j contains no points of E other than x_j in its interior. Precisely, $W_j = \{x|\frac{1}{2}(x + x_j) \in V_j\}$. Because $\frac{1}{2}(x_k + x_j)$, $j \neq k$ is not in the interior of V_j, x_k is not in the interior of W_j. The W_j are a cover of R_n because V_j is in the interior of W_j and every point x is in a V_j.

The members of the partition of unity φ_j will be functions whose support are precisely the W_j. Let σ be a face of W_j and L_σ be a linear function that vanishes on σ and is positive on the interior of W_j. Then the function $\prod L_\sigma$, the product being over all the faces of W_j, is positive on the interior of W_j and vanishes on its boundary. If g_j are arbitrary smooth functions that are positive on the boundary of W_j, then

$$\psi_j(x) = \begin{cases} \exp\left(-\dfrac{g_j(x)}{\prod L_\sigma(x)}\right) & x \in W_j \\ 0 & x \notin W_j \end{cases}$$

is a C^∞ function. In particular, ψ_j and all its derivatives vanish as x approaches the boundary of W_j. We set

$$\varphi_j(x) = \frac{\psi_j(x)}{\sum \psi_j(x)}$$

to obtain a partition of unity. The functions g_j can be used to help "shape" the bump functions φ_j and modify their gradients near the boundary of W_j. This may help in reducing the derivatives of the blended function f due to large derivatives of the φ_j.

Evaluation of f at $y \in \Re^n$ requires a list of the W_j for which $y \in W_j$. With this list, the values of $\psi_j(y)$ are computed, and finally

$$f(y) = \frac{\sum_j \psi_j(y) f_j(y)}{\sum_j \psi_j(y)}.$$

3 Example

Henderson [3] developed *Mulitfario*, a collection of algorithms for multiparameter continuation of k-dimensional level sets of a map $R^{n+k} \rightarrow R^n$. He used the computation of a 2-dimensional torus embedded in R^3 as a test case. This torus T is the zero set of the function

$$\left(x^2 + y^2 + z^2\right)^2 + \frac{39}{50}z^2 - \frac{11}{50}x^2 - \frac{11}{50}y^2 + \frac{1521}{10000} \tag{2}$$

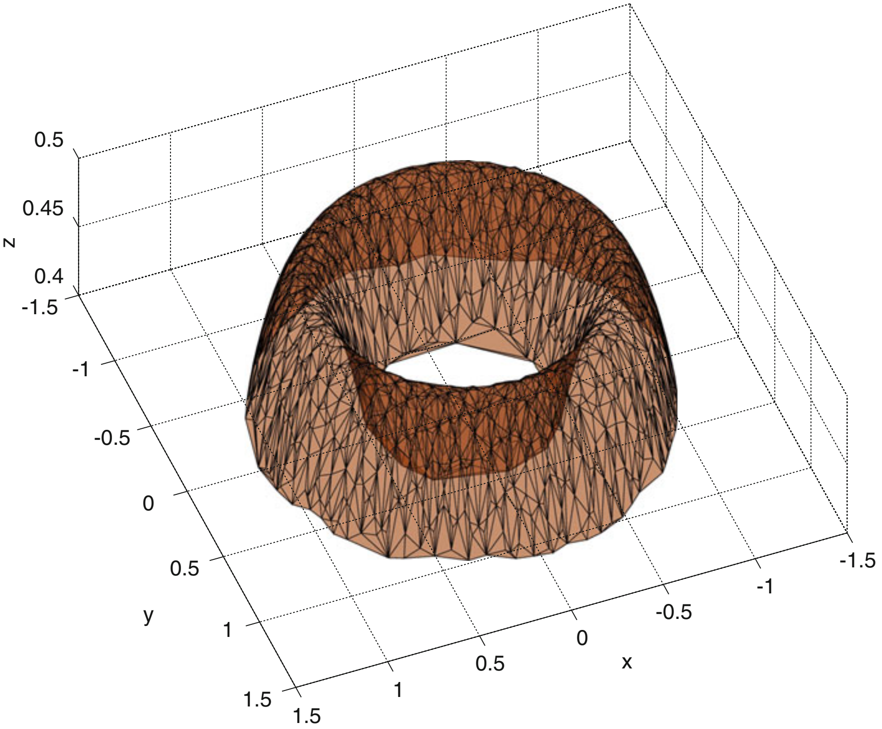


Fig. 1 The torus $T_{2/5}$

It is the surface of revolution whose intersection with the (x, z) plane is the circle of radius $1/2$ centered at $(4/5, 0, 0)$. See Fig. 1. Here we use computation of a portion of this torus to test the performance of the algorithm described above. We take output from a Multifario computation of this torus as our starting point. In Multifario, manifolds are represented by discrete sets of points p_j together with approximations to their tangent spaces at these points. The manifolds are approximated by polytopes whose boundaries are equidistant between a pair of p_j. This gives a discontinuous geometric object: there are gaps at the boundaries of the polytopes. Our algorithm produces smooth interpolations of a portion of the torus represented as a graph of a function $F_z(x, y)$. Denoting by T_c the part of the torus with $F_z > c$, we apply the algorithm described above to $T_{2/5}$. To avoid issues related to boundaries, we select the mesh points produced by Multifario on the larger surface $T_{1/5}$ and project these onto the (x, y) plane, obtaining the set of points E. We next evaluate the jets of degree 4 of F_z to obtain a Whitney field τ on E for $T_{1/5}$ via explicit formulas for F_z and its derivatives. The program Maple was used to generate formulas for these derivatives, and they were evaluated with Matlab.

We computed interpolations of τ with the algorithms described above, implemented in Matlab. Algorithms from the Computational Geometry Algorithms

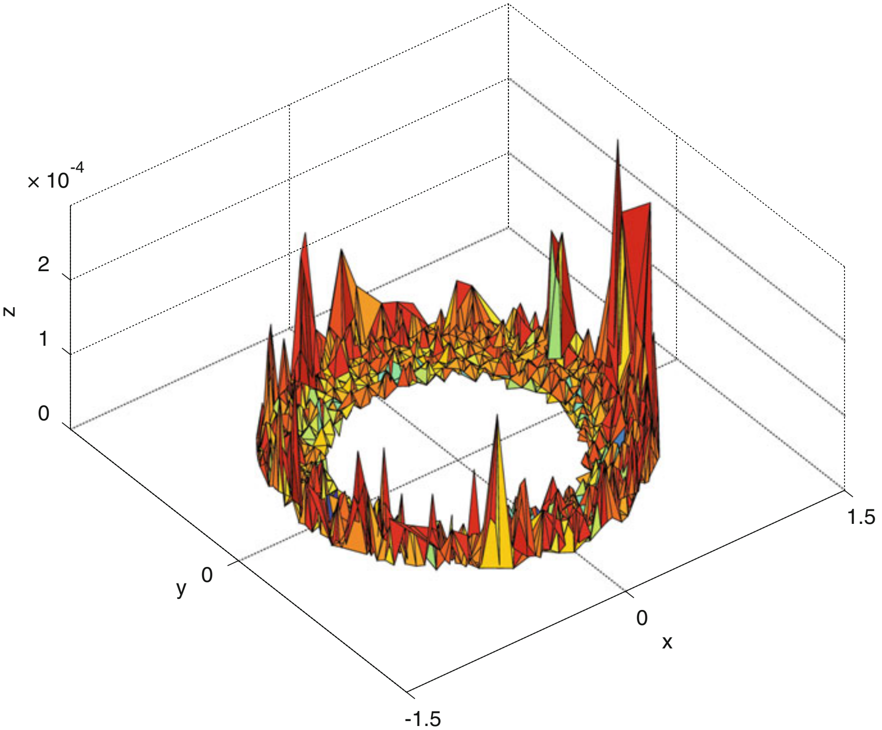


Fig. 2 A three dimensional plot of the residuals between the zero set of the function (2) and the interpolation of three jets at the (added) points of the barycentric subdivision of the mesh produced by Multifario

Library (CGAL) that have been incorporated into Matlab [4] were used to compute the Voronoi diagram of E. The cells of the Voronoi diagram were then expanded by a factor of 2 from the interior mesh point to obtain supports for a partition of unity of a region containing the annulus $A_{2/5}$ defined by $0.5 < r < 1.1$ in the (x, y) plane. The values of interpolations of the Whitney fields τ of degrees 1–4 at the vertices of the barycentric subdivision of E were compared with explicit evaluation of the function F_z at these points. We denote the differences of these values by R_j, the subscript labeling the degree of the interpolation.

Figure 2 plots a piecewise linear interpolation of the computed residuals of R_3 on the annulus $A_{2/5}$. There are 695 points in the mesh E output from Multifario; residuals of interpolations at 1,819 mesh points of the barycentric subdivision are plotted in the figure. The figure illustrates vividly two observations: (1) the largest residuals are close to the boundary of the annulus where the function F_z has a larger gradient and mesh triangles with a larger aspect ration, and (2) there are a few outliers at which the residuals have large relative magnitude. The largest magnitude of the residuals is approximately 3.35e−4. To obtain another perspective on the

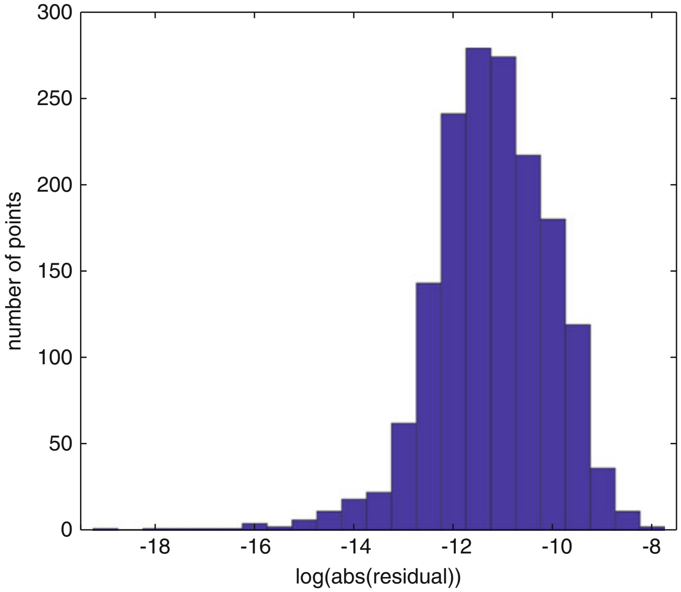


Fig. 3 A histogram of logarithmically transformed residuals between the zero set of the function (2) and interpolations of three jets at the (added) points of the barycentric subdivision of the mesh produced by Multifario

Table 1 Accuracy of the interpolation on $A_{2/5}$ by averaged jets of degrees 1–4 is summarized for two different meshes with the values of the maximum magnitude of the interpolation error and the mean of the logarithms of these magnitudes

Degree	Mesh points	Int points	Max residual	Mean log resid
1	695	1,632	0.0066	−6.13
2	695	1,632	8.04e−4	−10.0
3	695	1,632	3.35e−4	−11.2
4	695	1,632	1.44e−4	−14.0
1	4,075	9,757	0.0016	−8.09
2	4,075	9,757	8.54e−5	−13.4
3	4,075	9,757	2.16e−5	−15.2
4	4,075	9,757	3.87e−6	−19.4

accuracy of the interpolation, Fig. 3 plots a histogram of the logarithm of the residual magnitudes. The mean of these logarithms is approximately -11.2, representing a residual of magnitude approximately 1.32e−5. The largest residual is due primarily to a single cubic Taylor polynomial based at a point p of E near (1.07, –0.026) that is evaluated at a point q close to $(1.12, 0.035)$. Note that p lies near the boundary of $A_{2/5}$ and q lies outside it.

Table 1 summarizes the effects on accuracy of interpolation due to mesh refinement and increasing jet order. Both refinement and increasing jet order consistently improve accuracy as measured by maximum residuals and the mean of the logarithmically transformed residuals. These data can be compared with estimates of the residuals from polyhedral approximations in the tangent spaces of points of E like those used in of Multifario. We compared the values of these polyhedral approximations at the edge points added to the barycentric subdivision of E. These are midpoints of segments joining the mesh points in adjacent cells. The maximum size of the gaps at these points is approximately 0.0044 and the mean of the logarithms of the gap sizes is approximately -8.16. We also computed the residuals between the correct value of z and the average of the values from the two polyhedra. The magnitude of the residuals was smaller than 0.0090 and the mean of the logarithms of the residuals was approximately -6.58. The residual magnitudes are similar to those obtained from the partition of unity averages when using jets of degree 1.

These results are expected since the partition of unity is also averaging linear approximations to the function at the mesh points E. Note that averaging itself is insufficient to increase the accuracy of linear approximations. This is readily seen from the following example: consider a convex function $f : \Re \to \Re$ and linear approximations to f at two points $x_1 < x_2$. There will be a $x_1 < y < x_2$ so that both linear approximations give the same value at y. In this case, all weighted averages are the same and the residual with the value of f is $O(|y - x_j|^2)$. This indicates that the accuracy of approximation of our interpolation method rests with the accuracy of the individual patches: averaging with a partition of unity smooths the interpolating function while doing little to improve the C^0 accuracy of approximation.

4 Discussion

There is a vast gap in approximation theory for univariate and multivariate functions. Expansion of univariate analytic functions in suitable bases such as Fourier series or Chebyshev polynomials yields highly accurate approximations that can be used to interpolate function values. Analogous methods for multivariate functions do not seem to be prevalent. We demonstrate here how standard algorithms of computational geometry can be used as the foundation for smoothly interpolating functions with partitions of unity. The partitions of unity average or blend surface elements while preserving their values and the values of their derivatives at a set of mesh points. The methods are independent of the geometry of the mesh and simple to implement. We end with two additional remarks.

One of the slowest parts of the algorithm implemented for this paper is testing which domains W_j of the partition of unity contain a point y at which the approximation is to be evaluated. In most circumstances, we expect that the number of such domains will be small and that a local calculation will suffice to find the W_j. However, the following example illustrates the case that all domains intersect at a single point. Consider the set $E = \{\exp(2\pi i \frac{j}{N}, 0 \le j < N\}$ in $\Re^2$ regarded as the complex line. The Voronoi diagram of E consists of N sectors with apex at the origin. Each set W_j of the corresponding cover contains the origin in its interior.

This will remain true if the points of E are perturbed. Thus, there are no bounds on the number of sets W_j that can overlap at a single point. On the other hand, if the Voronoi cells within a region each are contained in a ball of radius R, then we need only examine sets W_j whose centers are within distance $2R$ of y to find all the W_j that contain y. We expect that this observation can be used to significantly decrease the run time of the algorithm.

Theoretically, we would like to be able to estimate the derivatives of partition of unity averages and to choose partitions of unity that optimize these averages with respect to a given criterion. The derivatives of $f = \sum_j \varphi_j f_j$ involve derivatives of both the f_j and the φ_j. Since $\sum_j \varphi_j$ is identically one, for any partial derivative ∂, $\sum_j \partial\varphi_j = 0$. Therefore ∂f is $\sum_j \varphi_j \partial f_j$ plus additional terms of the form $\sum_j \partial_1\varphi_j \partial_2(f_j - f)$. If the f_j are close to one another in a C^r norm, then the additional terms will be small. This basic estimate lends a glimmer of hope that partitions of unity might serve as a vehicle for implementing higher order methods for multivariate problems. For example, high order methods are commonplace in solving initial value problems for ordinary differential equations. Are partitions of unity a tool for developing higher order methods for solving partial differential equations?

References

1. CGAL Computational Geometry Algorithms Library, http://www.cgal.org/
2. C. Fefferman, Interpolation by linear programming I. Discret. Contin. Dyn. Syst. **30**, 477–492 (2011)
3. M. Henderson, Multiple parameter continuation: computing implicitly defined k-manifolds. Int. J. Bifurc. Chaos **12**, 451–476 (2002)
4. Matlab, MathWorks, MA, US, http://www.mathworks.com/products/matlab/
5. D, Zorin, P. Schröder, T. DeRose, L. Kobbelt, A. Levin, W. Sweldens, Subdivision for modeling and animation, in *SIGGRAPH 2000 Course Notes*, New Orleans, http://mrl.nyu.edu/publications/subdiv-course2000/coursenotes00.pdf

Bifurcations of Solutions of the 2-Dimensional Navier–Stokes System

Dong Li and Yakov G. Sinai

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract For the 2-dimensional Navier–Stokes System written for the stream functions we construct a set of initial data for which initial critical points bifurcate into three critical points. This can be interpreted as the birth of new viscous vortices from a single one. In another class of solutions vortices merge, i.e. the number of critical points decrease.

1 Introduction

We are very glad to dedicate this paper to Professor S. Smale. The works of Smale in the theory of dynamical systems played a great role in the development of this important field and led to the appearance of new concepts and methods. We wish Professor Smale a very good health and many new important results.

The usual bifurcation theory deals with one-parameter families of smooth maps or vector fields. In this situation fixed points or periodic orbits become functions of this parameter. Bifurcations appear when their linearized spectrum changes its structure. The main role in the theory is played by the so-called versal deformations, i.e. by special families such that arbitrary families can be represented as some

D. Li (✉)
Deparment of Mathematics, University of British Columbia, Vancouver, BC, Canada, V6T 1Z2
e-mail: mpdongli@gmail.com

Y.G. Sinai
Mathematics Department, Princeton University, Princeton, NJ 08544, USA

Landau Institute of Theoretical Physics, Moscow, Russia
e-mail: sinai@math.princeton.edu

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_10, © Springer-Verlag Berlin Heidelberg 2012

projections of versal deformations (see, for example [1]). In this approach the positions of bifurcating orbits and their dependence on the parameter are known.

In this paper we consider a dynamical system generated by the 2-dimensional Navier–Stokes System and deformations are produced by solutions of this system. Certainly, this is a very special case of a much more general problem in which Navier–Stokes System is replaced by linear or non-linear PDE for which strong existence and uniqueness results are known. The next step is to choose fixed points or periodic orbits and sometimes this can be a difficult problem. In our case this is done under the assumption of an additional symmetry of the problem.

We write Navier–Stokes System for the stream function $\psi = \psi(\tilde{x}_1, \tilde{x}_2, t)$ on the 2-dimensional square $0 \le \tilde{x}_1,\ \tilde{x}_2 \le \pi$:

$$\frac{\partial \psi}{\partial t} + \Delta^{-1}\left(\frac{\partial \psi}{\partial \tilde{x}_1} \cdot \frac{\partial \Delta \psi}{\partial_{\tilde{x}_2}} - \frac{\partial \psi}{\partial \tilde{x}_2} \cdot \frac{\partial \Delta \psi}{\partial \tilde{x}_1}\right) = \Delta \psi. \tag{1}$$

In (1) the viscosity is taken to be 1 and the external forcing terms are absent. The velocity of the fluid $u = (u_1, u_2)$ is expressed from ψ through the relations

$$u_1 = -\frac{\partial \psi}{\partial \tilde{x}_2}, \quad u_2 = \frac{\partial \psi}{\partial \tilde{x}_1} \tag{2}$$

which show that u is a local function of ψ. This is one of the advantages of ψ. Moreover, the velocity u given by (2) always satisfies incompressibility condition

$$\operatorname{div}(u) = \frac{\partial u_1}{\partial \tilde{x}_1} + \frac{\partial u_2}{\partial x_2} = 0.$$

We consider the space of functions ψ written as a series

$$\psi(\tilde{x}_1, \tilde{x}_2, t) = \sum_{m^2+n^2 \neq 0} f_{mn} \sin m\tilde{x}_1 \sin n\tilde{x}_2. \tag{3}$$

The coefficients f_{mn} are odd functions of its arguments and decay fast enough so that all appearing series converge. In Sect. 2 we reproduce the proof of the theorem from [4] in which we show that the space of such ψ is invariant under the dynamics generated by (1).

In (1) the operator Δ^{-1} has the form

$$\Delta^{-1}\psi = -\sum_{m^2+n^2 \neq 0} \frac{1}{m^2+n^2} \sin m\tilde{x}_1 \sin n\tilde{x}_2.$$

The formulas (2) and (3) show that on the boundary the velocity vector u is directed along the boundary. This situation is called the slip boundary condition. From the physical point of view it is not so natural but it is quite satisfactory as a mathematical model.

Let us write down an infinite-dimensional system of ODE for the coefficients f_{mn} which follows from (1) and actually is equivalent to (1)

$$\frac{df_{mn}}{dt} - \frac{1}{m^2+n^2} \sum_{\substack{m'+m''=m \\ n'+n''=n}} f_{m'n'} f_{m''n''} \cdot ((m'')^2 + (n'')^2) \cdot (m'n'' - m''n') \\ = -(m^2+n^2) f_{mn}. \tag{4}$$

Introduce the vorticity

$$\omega = \Delta\psi = \sum_{m,n} \omega_{mn} \sin m\tilde{x}_1 \sin n\tilde{x}_2$$

which shows that $\omega_{mn} = -(m^2+n^2) f_{mn}$. For the coefficients ω_{mn} we have even a simpler system of ODE equivalent to (4)

$$\frac{d\omega_{mn}}{dt} + \sum_{\substack{m'+m''=m \\ n'+n''=n}} \omega_{m'n'}\omega_{m''n''} \cdot \frac{m'n''-m''n'}{(m')^2+(n')^2} \\ = -(m^2+n^2)\omega_{mn}. \tag{5}$$

In [2–4], the following theorem was proven

Theorem 1 (Global wellposedness and decay). *Let* $\gamma > 1$, $A > 0$ *and*

$$|\omega_{mn}(0)| \le \frac{A}{(m^2+n^2)^{\frac{\gamma}{2}}}, \tag{6}$$

for all m, n, $m^2+n^2 \neq 0$. *Then for some absolute constant* $K_1 > 0$ *and all* $t > 0$,

$$|\omega_{mn}(t)| \le \frac{AK_1}{(m^2+n^2)^{\frac{\gamma}{2}}}. \tag{7}$$

The proof of Theorem 1 is given in Sect. 2. In the periodic case it was given in [3] and [4] and extended to other boundary conditions in [2]. The inequality (7) implies that for the stream function

$$|f_{mn}(t)| \le \frac{AK_1}{(m^2+n^2)^{\frac{\gamma}{2}+1}}, \quad \forall\, m,n.$$

We shall take γ to be so large that the decay of f_{mn} will be sufficient for our purposes. Actually the decay of f_{mn} is much faster but we do not dwell on this here.

Remark 1. Our flow (1)–(3) is closely connected with a special class of 2π-periodic flows on the whole plane. Namely suppose $\tilde{\psi} = \tilde{\psi}(\tilde{x}_1, \tilde{x}_2, t)$ is a solution to the Navier–Stokes equation with 2π*-periodic boundary condition*, and satisfy

$$\tilde{\psi}(-\tilde{x}_1, \tilde{x}_2, t) = -\tilde{\psi}(\tilde{x}_1, \tilde{x}_2, t) = \tilde{\psi}(\tilde{x}_1, -\tilde{x}_2, t), \quad \forall\, \tilde{x}_1, \tilde{x}_2. \tag{8}$$

It is not difficult to check that the special symmetry (8) is preserved under the dynamics of the Navier–Stokes flow. Furthermore if we write

$$\tilde{\psi}(\tilde{x}_1, \tilde{x}_2, t) = \sum_{m,n} \tilde{f}_{mn} e^{i(m\tilde{x}_1 + n\tilde{x}_2)},$$

then

$$-\tilde{f}_{mn} = \tilde{f}_{-m,n} = \tilde{f}_{m,-n}, \quad \forall\, m, n.$$

Therefore from a simple computation

$$\tilde{\psi}(\tilde{x}_1, \tilde{x}_2, t) = -\sum_{m,n} \tilde{f}_{mn} \sin m\tilde{x}_1 \sin n\tilde{x}_2 \tag{9}$$

which corresponds exactly to (3) up to a minus sign. This shows that $\tilde{\psi}$ is also a solution to our problem (1)–(3).

We shall call extremal points of the stream function ψ the points of local minima or maxima of ψ. Near these points the velocity u is tangent to the level sets of ψ (or $\tilde{\psi}$) which are closed curves. It is natural to call extremal points of ψ viscous vortices. The main purpose of this paper is to show that these vortices can split or merge.

Now we can formulate our main results of this paper.

Theorem 2 (Existence of bifurcations). *There exists an open set $\mathcal{A}$ in the space of stream functions such that the following holds true:*

For each stream function $\psi_0 \in \mathcal{A}$, there is an open neighborhood U of the point $(\tilde{x}_1, \tilde{x}_2) = (\frac{\pi}{2}, \frac{\pi}{2})$, two moments of times $0 < t_1 < t_2$ such that the corresponding stream function $\psi = \psi(\tilde{x}_1, \tilde{x}_2, t)$ solves (1) with initial data ψ_0 and satisfy

1. *At $t = 0$, $(\frac{\pi}{2}, \frac{\pi}{2})$ is a non-degenerate minimum of ψ in the neighborhood U.*
2. *For any $0 < t \leq t_1$, ψ has only one critical point in U given by $(\tilde{x}_1, \tilde{x}_2) = (\frac{\pi}{2}, \frac{\pi}{2})$.*
3. *At $t = t_1$, $(\tilde{x}_1, \tilde{x}_2) = (\frac{\pi}{2}, \frac{\pi}{2})$ is a degenerate local minimum of ψ in U.*
4. *For $t_1 < t \leq t_2$, ψ has exactly three critical points in U. The point $(\frac{\pi}{2}, \frac{\pi}{2})$ becomes a saddle. Two other critical points are of the form $(\frac{\pi}{2} + x^*, \frac{\pi}{2} + y^*)$, $(\frac{\pi}{2} - x^*, \frac{\pi}{2} - y^*)$ where $x^* \neq 0$, $y^* \neq 0$ and are local minima.*

Remark 2. Under our conditions the point $(\frac{\pi}{2}, \frac{\pi}{2})$ is the extremal point of the stream function for all time. This property plays the same role as the knowledge of fixed points or periodic orbits in the usual theory of bifurcations.

Remark 3. The fact that the extra critical points emerge in the form $(\frac{\pi}{2} + x^*, \frac{\pi}{2} + y^*)$, $(\frac{\pi}{2} - x^*, \frac{\pi}{2} - y^*)$ is not surprising. As we shall see later in Sect. 3, by the inversion symmetry (18), our stream function ψ is invariant under the reflection about the point $(\frac{\pi}{2}, \frac{\pi}{2})$.

Our next result is in some sense the reversal of the process described in Theorem 2. For a class of initial data having three critical points near the special point $(\frac{\pi}{2}, \frac{\pi}{2})$, we show that they merge into one critical point in finite time.

Theorem 3 (Merging of critical points). *There exists an open set $\mathcal{A}$ in the space of stream functions such that the following holds true:*

For each stream function $\psi_0 \in \mathcal{A}$, there is an open neighborhood U of the point $(\tilde{x}_1, \tilde{x}_2) = (\frac{\pi}{2}, \frac{\pi}{2})$, two moments of times $0 < t_1 < t_2$ such that the corresponding stream function $\psi = \psi(\tilde{x}_1, \tilde{x}_2, t)$ solves (1) with initial data ψ_0 and satisfy

1. *For $0 \le t < t_1$, ψ has exactly three critical points in U. The point $(\frac{\pi}{2}, \frac{\pi}{2})$ is a saddle. Two other critical points are of the form $(\frac{\pi}{2} + x^*, \frac{\pi}{2} + y^*)$, $(\frac{\pi}{2} - x^*, \frac{\pi}{2} - y^*)$ where $x^* \neq 0$, $y^* \neq 0$ and are local minima.*
2. *At $t = t_1$, $(\frac{\pi}{2}, \frac{\pi}{2})$ is a degenerate minimum of ψ in the neighborhood U.*
3. *For any $t_1 < t \le t_2$, ψ has only one critical point in U given by $(\tilde{x}_1, \tilde{x}_2) = (\frac{\pi}{2}, \frac{\pi}{2})$.*

This paper is organized as follows. In Sect. 2 we give the proof of Theorem 1. In Sect. 3 we derive the equation for extremal points and formulate sufficient conditions for bifurcations needed in Theorem 2. Section 4 is devoted to the construction of bifurcations in the degenerate case. In Sect. 5 we prove the existence of bifurcation for non-degenerate initial data by using a perturbation argument. In Sect. 6 we give the construction of stream functions satisfying the needed conditions. In Sects. 7 and 8 we describe the proof of Theorem 3 and construction of initial conditions.

2 Proof of Theorem 1

In this section we give the proof of Theorem 1 using the trapping argument from [4].

We shall use the letter C with or without indices to denote different absolute constants whose values may vary from line to line. The actual value of C does not play any role in our arguments.

To simplify notations, denote $Z^2_* = \{(m, n) \in \mathbb{Z}^2, m \neq 0, n \neq 0\}$ and $r = (m, n) \in \mathbb{Z}^2_*$, $r' = (m', n') \in \mathbb{Z}^2_*$, $r'' = (m'', n'') \in \mathbb{Z}^2_*$, and also denote $\omega_r = \omega_{mn}$, $\omega_{r'} = \omega_{m'n'}$ and so on.

By standard enstrophy inequality, we have

$$\|\omega(t)\|_{L^2_{\tilde{x}_1\tilde{x}_2}([0,\pi]\times[0,\pi])} \le \mathcal{E}_0, \quad \forall\, t > 0,$$

where $\mathcal{E}_0 > 0$ is the enstrophy at $t = 0$.

By Fourier transform, this implies

$$\left(\sum_{r\in\mathbb{Z}^2_*} |\omega_r(t)|^2\right)^{\frac{1}{2}} \le C_1\mathcal{E}_0, \forall\, t > 0. \tag{10}$$

Let $K_1 > 0$ be a constant depending on A which will be taken sufficiently large. By (10), we get

$$|\omega_r(t)| \le \frac{C_1K_1\mathcal{E}_0}{|r|^{\frac{\gamma}{2}}}, \quad \forall\, |r| \le K_1^{\frac{2}{\gamma}}, \quad \forall\, t > 0.$$

Define the trapping set

$$\Omega(K_1) = \left\{(\tilde{\omega}_r) : |\tilde{\omega}_r| \le \frac{C_1K_1\mathcal{E}_0}{|r|^{\frac{\gamma}{2}}}, \quad \forall\, |r| \ge K_1^{\frac{2}{\gamma}}\right\}. \tag{11}$$

Now we show that for all $t > 0$ the trajectories of our system remain inside the set $\Omega(K_1)$. Indeed at $t = 0$, by choosing $K_1 > 2A$ (see (6)), we get that our system lies strictly inside $\Omega(K_1)$. Assume $t_1 > 0$ is the first moment of time when our system reaches the boundary $\partial\Omega(K_1)$.[1]

Then for some $|r^*| \ge K_1^{\frac{2}{\gamma}}$,

$$|\omega_{r^*}(t_1)| = \frac{C_1K_1\mathcal{E}_0}{|r^*|^{\gamma}}.$$

WLOG assume

$$\omega_{r^*}(t_1) = \frac{C_1K_1\mathcal{E}_0}{|r^*|^{\gamma}}.$$

The case $\omega_{r^*}(t_1) = -\frac{C_1K_1\mathcal{E}_0}{|r^*|^{\gamma}}$ is similar and therefore its discussion is omitted.

We then aim to show that

$$\partial_t\omega_{r^*}(t)\Big|_{t=t_1} < 0.$$

[1] Strictly speaking, we should consider the Galerkin approximations of our system to avoid issues connected with the infinite dimensionality of our system.

This will guarantee that the trajectory of our system cannot exit the trapping set $\Omega(K_1)$ and will remain inside $\Omega(K_1)$.

Recall the vorticity equation

$$\partial_t \omega + \Delta^{-1}\nabla^{\perp}\omega \cdot \nabla\omega = \Delta\omega. \tag{12}$$

By using (12), we have

$$\Delta^{-1}\nabla^{\perp}\omega \cdot \nabla\omega = \sum_{(m,n)\in\mathbb{Z}_*^2} N_{mn} \sin m\tilde{x}_1 \sin n\tilde{x}_2,$$

where

$$|N_{mn}| \leq \sum_{r'+r''=r} \frac{|\omega_{r'}|}{|r'|} \cdot |r''| \cdot |\omega_{r''}|. \tag{13}$$

There are two cases.

Case 1. $|r'| > \frac{1}{3}|r|$. Then

$$\frac{|r''|}{|r'|} \leq \frac{|r| + |r'|}{|r'|} \leq C.$$

Hence

$$\begin{aligned} \text{RHS of (13)} &\leq C \sum_{\substack{r'+r''=r \\ |r'|>\frac{1}{3}|r|}} |\omega_{r'}| \cdot |\omega_{r''}| \\ &\leq C \left(\sum_{|r'|>\frac{1}{3}|r|} |\omega_{r'}|^2 \right)^{\frac{1}{2}} \cdot \left(\sum_{r''} |\omega_{r''}|^2 \right)^{\frac{1}{2}} \\ &\leq \frac{CK_1}{|r|^{\gamma-1}} \mathcal{E}_1. \end{aligned}$$

Case 2. $|r''| > \frac{1}{3}|r|$ and $|r'| \leq \frac{1}{3}|z|$. Then

$$\begin{aligned} \text{RHS of (12)} &\leq \frac{CK_1}{|r|^{\gamma-1}} \sum_{|r'|\leq\frac{1}{3}|r|} \frac{|\omega_{r'}|}{|r'|} \\ &\leq \frac{CK_1}{|r|^{\gamma-1}} \log|r| \cdot \mathcal{E}_1. \end{aligned}$$

Concluding from the above cases, we get

$$|N_{r^*}(t_1)| \le \frac{CK_1\mathcal{E}_1}{|r^*|^{\gamma-1}} \log |r^*|$$

and hence by (12)

$$\begin{aligned}\partial_t \omega_{r^*}(t)\Big|_{t=t_1} &\le \frac{CK_1\mathcal{E}_1}{|r^*|^{\gamma-1}} \log |r^*| - \frac{C_1K_1\mathcal{E}_1}{|r^*|^{\gamma-2}} \\ &< 0,\end{aligned}$$

if K_1 is sufficiently large (recall that by (11), $|r^*| \ge K_1^{\frac{2}{\gamma}}$). This finishes the trapping argument and Theorem 1 is proved.

3 The Equation for Extremal Points

We consider a special class of flows

$$\psi(\tilde{x}_1, \tilde{x}_2, t) = \sum_{m+n \text{ is even}} f_{mn} \sin(m\tilde{x}_1) \sin(n\tilde{x}_2). \tag{14}$$

It is also invariant under the Navier–Stokes dynamics. If this condition is valid, then on the vertical boundaries, for any $0 \le \tilde{x}_2 \le \pi$, the velocity vector at the point $(0, \tilde{x}_2)$ has the same magnitude but opposite direction to the velocity at the point $(0, \pi - \tilde{x}_2)$. In some sense they form a dipole with center at $(\frac{\pi}{2}, \frac{\pi}{2})$. Similar statements also hold for the horizontal boundaries.

In this paper we study bifurcations of the stream function near the point $(\frac{\pi}{2}, \frac{\pi}{2})$.

After the change of variables,

$$\tilde{x}_1 = \frac{\pi}{2} + x, \quad \tilde{x}_2 = \frac{\pi}{2} + y, \tag{15}$$

we shift our coordinate system and define

$$\phi(x, y, t) = \psi(\frac{\pi}{2} + x, \frac{\pi}{2} + y, t). \tag{16}$$

By (16), (14) and (9), we get

$$\begin{aligned}\phi(x, y, t) &= -\sum_{m+n \text{ is even}} f_{mn} e^{i(\frac{m+n}{2}\pi + mx + ny)} \\ &= -\sum_{m+n \text{ is even}} f_{mn} (-1)^{\frac{m+n}{2}} e^{i(mx+ny)}.\end{aligned}$$

Since ϕ and f_{mn} are both real-valued, we get

$$\phi(x,y,t) = -\sum_{m+n \text{ is even}} f_{mn}(-1)^{\frac{m+n}{2}} \cos(mx+ny) \tag{17}$$

$$= \phi(-x,-y,t), \tag{18}$$

i.e. ϕ satisfies the inversion symmetry. It implies that at the point $(x,y) = (0,0)$ the gradient of ϕ vanishes.

Introduce a neighborhood $U_\delta = \{(x,y) : \ x^2 + y^2 \le \delta^2\}$. Later we shall choose δ to be sufficiently small.

For sufficiently small $t_2 > 0$ consider the time interval $[0,t_2]$ and write the following expansion of ϕ in the neighborhood U_δ:

$$\begin{aligned}\phi(x,y,t) &= \phi(0,0,t) + a_1(t)x^2 + a_2(t)y^2 + a_3(t)xy \\ &\quad + b_1(t)x^4 + b_2(t)y^4 + b_3(t)x^3y + b_4(t)x^2y^2 + b_5(t)xy^3 \\ &\quad + \epsilon(x,y,t),\end{aligned} \tag{19}$$

where the remainder term satisfies the inequalities

$$\begin{aligned}\epsilon(x,y,t) &= O(x^6 + y^6), \\ \frac{\partial\epsilon}{\partial x}(x,y,t) &= O(|x|^5 + |y|^5), \\ \frac{\partial\epsilon}{\partial y}(x,y,t) &= O(|x|^5 + |y|^5).\end{aligned} \tag{20}$$

In the expansion (19), terms of odd degree are not present because of the symmetry (18).

The first equation for the critical point takes the form

$$\partial_x \phi = 0.$$

By (19), we get

$$\begin{aligned}&2a_1x + a_3y + 4b_1x^3 + 3b_3x^2y + 2b_4xy^2 \\ &\quad + b_5y^3 + \frac{\partial\epsilon}{\partial x} = 0.\end{aligned} \tag{21}$$

Here and later we occasionally suppress the time dependence and write $a_i(t)$, $b_i(t)$ simply as a_i, b_i when the context is clear.

Assume for $0 \le t \le t_2$

$$a_3(t) \sim O(1). \tag{22}$$

More precisely

$$const \le a_3(t) \le const.$$

The values of constants play some role later. This will be clarified below (see (34)).

Equation 21 takes the form

$$y = -\frac{2a_1}{a_3}x - \frac{4b_1}{a_3}x^3 - \frac{3b_3}{a_3}x^2 y - \frac{2b_4}{a_3}xy^2 - \frac{b_5}{a_3}y^3 - \frac{1}{a_3}\frac{\partial \epsilon}{\partial x}. \tag{23}$$

Assume also that in formula (19)

$$b_2(0) = b_3(0) = b_4(0) = b_5(0) = 0. \tag{24}$$

For sufficiently small t_2 it implies that for $0 \le t \le t_2$,

$$b_i(t) \sim O(t), \quad i = 2, \cdots, 5. \tag{25}$$

Write (23) in the form

$$y = -\frac{2a_1}{a_3}x - \frac{4b_1}{a_3}x^3 + O(t)\cdot O(|x|^3 + |y|^3) + O(|x|^5 + |y|^5). \tag{26}$$

Since $(x, y) \in U_\delta$, we have the rough estimate

$$y = O(x). \tag{27}$$

Consider the other critical point equation

$$\frac{\partial \phi}{\partial y} = 0.$$

By (19), we get

$$2a_2 y + a_3 x + 4b_2 y^3 + b_3 x^3 + 2b_4 x^2 y + 3b_5 xy^2 + \frac{\partial \epsilon}{\partial y} = 0.$$

In view of the assumptions (25) and (20), we obtain

$$2a_2 y + a_3 x + O(t)\cdot O(|x|^3 + |y|^3) + O(|x|^5 + |y|^5) = 0.$$

Using (27), we get

$$2a_2 y + a_3 x + O(t) \cdot O(|x|^3) + O(|x|^5) = 0. \tag{28}$$

Substituting (26) into (28) and using again (27), we have

$$2a_2 \left(-\frac{2a_1}{a_3} x - \frac{4b_1}{a_3} x^3 \right) + a_3 x + O(t) \cdot O(|x|^3) + O(|x|^5) = 0.$$

Or simply,

$$\frac{a_3^2 - 4a_1 a_2}{a_3} x - \frac{8a_2 b_1}{a_3} x^3 + O(t) \cdot O(|x|^3) + O(|x|^5) = 0. \tag{29}$$

It is obvious that (29) has a solution $x = 0$. We now look for other possible solutions in U_δ. Dividing both sides of (29) by $\frac{x}{a_3}$, we obtain

$$(a_3^2 - 4a_1 a_2) - 8a_2 b_1 x^2 + O(t) \cdot O(x^2) + O(x^4) = 0. \tag{30}$$

We shall choose initial data very carefully so that the needed bifurcation happens on the time interval $[0, t_2]$. This will be done in two stages. At the first stage we consider the degenerate case in which the bifurcation happens immediately for $t > 0$. In the second stage we perturb the degenerate data so that the bifurcation is "delayed" to a later time $0 < t_1 < t_2$. In other words, we show that for sufficiently small (and special) perturbations, the desired bifurcation happens at $t = t_1$.

4 Stage 1: The Bifurcation in the Degenerate Case

Rewrite (30) as

$$-(a_3^2 - 4a_1 a_2) + 8a_2 b_1 x^2 + O(t) \cdot O(x^2) + O(x^4) = 0. \tag{31}$$

Choose $\phi_0 = \phi_0(x, y)$ so that

$$\begin{gathered} a_3(0)^2 - 4a_1(0)a_2(0) = 0, \\ \frac{d}{dt}\left(a_3^2(t) - 4a_1(t)a_2(t)\right)|_{t=0} > 0, \\ a_2(0) > 0,\ a_3(0) > 0,\ b_1(0) > 0. \end{gathered} \tag{32}$$

In addition, we also need

$$b_2(0) = b_3(0) = b_4(0) = b_5(0) = 0. \tag{33}$$

The possibility of choosing ϕ_0 with properties (32)–(33) will be shown later (see Sect. 6). Assume for the moment that these conditions are met, then for sufficiently small $t_2 > 0$, we have for $0 < t \le t_2$,

$$\begin{aligned} A_3'' &\ge a_3(t) \ge A_3' > 0, \\ A_2'' &\ge a_2(t) \ge A_2' > 0, \\ B_1'' &\ge \frac{d}{dt}\Big(a_3^2(t) - 4a_1(t)a_2(t)\Big) \ge B_1' > 0, \\ B_2'' &\ge b_1(t) \ge B_2' > 0, \end{aligned} \tag{34}$$

where A_i', A_i'', B_i', B_i'' are constants.

By (32)–(34), we have for $0 < t \le t_2$

$$(a_3^2(t) - 4a_1(t)a_2(t)) \sim t,$$

which means that

$$const \cdot t \le a_3^2(t) - 4a_1(t)a_2(t) \le const \cdot t.$$

Also we have

$$8a_2(t)b_1(t) \sim const.$$

It follows that for $0 < t \le t_2$, the equation (31) is of the form

$$-O(t) + O(1) \cdot x^2 + O(t) \cdot O(x^2) + O(x^4) = 0. \tag{35}$$

For sufficiently small δ and sufficiently small t_2, the equation (35) has two and only two solutions

$$x = \pm O(\sqrt{t})$$

because $O(t)$ is of order of t, $O(1) > 0$ and other terms do not play any essential role. In this sense solutions to (31) bifurcates into two solutions for $0 < t \le t_2$.

Remark that at $t = 0$, the only solution to (31) is $x = 0$ due to the conditions $a_3(0)^2 - 4a_1(0)a_2(0) = 0$ and $a_2(0)b_1(0) \sim const$.

5 Stage 2: Bifurcation from Non-degenerate Initial Data, a Perturbation Argument

In stage 2 we finish our construction of bifurcation from non-degenerate initial data. The main idea is to perturb the initial data considered in Stage 1. The perturbation will be chosen so that initially we will have only one local non-degenerate minimum located at $(x, y) = (0, 0)$.

To this end, consider $\tilde{\phi}_0 = \tilde{\phi}_0(x, y) \in C^\infty$ with the following properties:

$$\tilde{\phi}_0(x, y) = \tilde{\phi}_0(-x, -y), \quad \forall\, x, y,$$

$$\frac{\partial^4 \tilde{\phi}_0}{\partial x^m \partial y^n}\Big|_{(x,y)=(0,0)} = 0, \quad \forall\, m + n = 4, 0 \le m \le 4,$$

$$\frac{\partial^2 \tilde{\phi}_0}{\partial x \partial y}\Big|_{(x,y)=(0,0)} = 0, \quad \frac{\partial^2 \tilde{\phi}_0}{\partial x^2}\Big|_{(x,y)=(0,0)} > 0, \quad \frac{\partial^2 \tilde{\phi}_0}{\partial y^2}\Big|_{(x,y)=(0,0)} > 0. \tag{36}$$

Fix $\phi_0 = \phi_0(x, y)$ taken from Stage 1 which has the properties (32)–(33). We shall consider the perturbation by $\tilde{\phi}_0$ having the form

$$\tilde{\phi}_0^\epsilon(x, y) = \phi_0(x, y) + \epsilon\tilde{\phi}_0(x, y),$$

where $\epsilon > 0$ is sufficiently small.

Denote the corresponding solution of the main equation (1) (in the shifted coordinates) by $\phi^\epsilon = \phi^\epsilon(x, y, t)$. To simplify the notations, we expand $\phi^\epsilon(x, y, t)$ in the form corresponding to (19), i.e. we write

$$\begin{aligned}\phi^\epsilon(x, y, t) = \phi^\epsilon(0, 0, t) &+ a_1^\epsilon(t)x^2 + a_2^\epsilon y^2 + a_3^\epsilon xy \\ &+ b_1^\epsilon(t)x^4 + b_2^\epsilon(t)y^4 + b_3^\epsilon(t)x^3y + b_4^\epsilon(t)x^2y^2 + b_5^\epsilon(t)xy^3 \\ &+ \tilde{\epsilon}(x, y, t)\end{aligned} \tag{37}$$

where $\tilde{\epsilon}$ satisfies an estimate similar to (20).

We now check the properties of $\phi^\epsilon(x, y, t)$.

(a) At $t = 0$, the point $(x, y) = (0, 0)$ is the unique extremum of $\phi^\epsilon(x, y, 0)$ in the neighborhood U_δ. Also $(0, 0)$ is a non-degenerate local minimum.

To prove this, we note that due to (32), (33) and (36), the critical point equation (30) still holds for $\phi^\epsilon(x, y, t)$ for sufficiently small $\epsilon > 0$ with corresponding coefficients a_1, a_2, a_3, b_1 now replaced by a_1^ϵ, a_2^ϵ, a_3^ϵ, b_1^ϵ. In particular this gives us

$$(a_3^\epsilon(0))^2 - 4a_1^\epsilon(0)a_2^\epsilon(0) - 8a_2^\epsilon(0)b_1^\epsilon(0)x^2 + O(x^4) = 0. \tag{38}$$

Denote

$$\tilde{a}_1 = \frac{\partial^2 \tilde{\phi}_0}{\partial x^2}\Big|_{(x,y)=(0,0)} > 0,$$

$$\tilde{a}_2 = \frac{\partial^2 \tilde{\phi}_0}{\partial y^2}\Big|_{(x,y)=(0,0)} > 0.$$

By (32) and (36), we have

$$\begin{aligned}(a_3^\epsilon(0))^2 &- 4a_1^\epsilon(0)a_2^\epsilon(0)\\ &= a_3(0)^2 + O(\epsilon^2) - 4(a_1(0)+\epsilon\tilde{a}_1)(a_2(0)+\epsilon\tilde{a}_2)\\ &= -4(a_1(0)\tilde{a}_2 + a_2(0)\tilde{a}_1)\epsilon + O(\epsilon^2). \qquad (39)\end{aligned}$$

On the other hand, for sufficiently small $\epsilon > 0$, by using (38) and (36), we have

$$\begin{aligned}a_2^\epsilon(0)b_1^\epsilon(0) &= (a_2(0)+O(\epsilon))\cdot(b_1(0)+O(\epsilon^2))\\ &= a_2(0)b_1(0) + O(\epsilon)\\ &\sim const. \qquad (40)\end{aligned}$$

Therefore by (39) and (40), the equation (38) takes the form

$$-O(1)\epsilon - O(1)\cdot x^2 + O(x^4) = 0,$$

or simply

$$O(1)\cdot\epsilon + O(1)\cdot O(x^2) = 0.$$

It is clear that for $\epsilon > 0$ this equation does not have any real-valued solution in U_δ.

To show that $(0,0)$ is a non-degenerate local minimum at $t = 0$, we observe that by (39), for sufficiently small $\epsilon > 0$,

$$(a_3^\epsilon(0))^2 - 4a_1^\epsilon(0)a_2^\epsilon(0) < 0. \qquad (41)$$

Also we have by (32)

$$a_1^\epsilon(0) > 0, \quad a_2^\epsilon(0) > 0. \qquad (42)$$

Equations 41 and 42 show that the Hessian matrix

$$\begin{pmatrix} a_1^\epsilon(0) & \frac{1}{2}a_3^\epsilon(0)\\ \frac{1}{2}a_3^\epsilon(0) & a_2^\epsilon(0)\end{pmatrix}$$

is strictly positive definite. Hence $(0,0)$ is a non-degenerate local minimum.

(b) Consider the function

$$D^\epsilon(t) = (a_3^\epsilon(t))^2 - 4a_1^\epsilon(t)a_2^\epsilon(t).$$

It will be proven below that for sufficiently small $\epsilon > 0$, the following holds: There exists unique $t_1 = t_1(\epsilon) > 0$ such that

$$\begin{aligned} D^{\epsilon}(t) &< 0, \quad \text{for } 0 \leq t < t_1, \\ D^{\epsilon}(t) &= 0, \quad \text{for } t = t_1, \\ D^{\epsilon}(t) &> 0, \quad \text{for } t_1 < t \leq t_2. \end{aligned} \tag{43}$$

Furthermore, the reduced critical-point equation (see (30))

$$D^{\epsilon}(t) - 8a_2^{\epsilon}(t)b_1^{\epsilon}(t)x^2 + O(t) \cdot O(x^2) + O(x^4) = 0 \tag{44}$$

has

- No solution for $0 \leq t < t_1$,
- Exactly one solution given by $x = 0$ for $t = t_1$,
- Two nonzero solutions for $t_1 < t \leq t_2$.

To prove (43), we recall the bound (34) , where for $0 \leq t \leq t_2$

$$B_1'' \geq \frac{d}{dt}\left(a_3^2(t) - 4a_1(t)a_2(t)\right) \geq B_1' > 0. \tag{45}$$

Since our initial data are given by

$$\phi_0^{\epsilon}(x, y) = \phi_0(x, y) + \epsilon\tilde{\phi}_0(x, y),$$

it follows from simple perturbation theory that for sufficiently small $\epsilon > 0$, we have

$$\|\phi^{\epsilon}(x, y, t) - \phi(x, y, t)\|_{H^m_{t,x,y}} \leq \eta(\epsilon, m), \tag{46}$$

where $\eta(\epsilon, m) \to 0$ as $\epsilon \to 0$ and m is fixed.

The notation $H^m_{t,x,y}$ denotes m^{th} Sobolev norms of ψ:

$$\|\psi\|_{H^m_{t,x,y}} = \sum_{0 \leq \alpha+\beta+\gamma \leq m} \left\|\partial_t^{\alpha}\partial_x^{\beta}\partial_y^{\gamma}\psi\right\|_{L^2}.$$

Take m to be sufficiently large and then ϵ sufficiently small. It follows from (45) and (46) that

$$2B_1'' \geq \frac{d}{dt}\left((a_3^{\epsilon}(t))^2 - 4a_1^{\epsilon}(t)a_2^{\epsilon}(t)\right) \geq \frac{B_1'}{2} > 0, \tag{47}$$

for any $0 \leq t \leq t_2$.

This means in particular that $D^{\epsilon}(t)$ is strictly increasing for $0 \leq t \leq t_2$.

By (39), we have for $t = 0$ and ϵ sufficiently small,

$$D^{\epsilon}(0) < 0. \tag{48}$$

On the other hand for $t = t_2$, by using the analysis from Stage 1, we have

$$(a_3(t_2))^2 - 4a_1(t_2)a_2(t_2) > 0.$$

Since

$$D^{\epsilon}(t_2) = (a_3(t_2))^2 - 4a_1(t_2)a_2(t_2) + O(\epsilon),$$

it follows easily that for ϵ sufficiently small

$$D^{\epsilon}(t_2) > 0. \tag{49}$$

Now (47)–(49) easily yield (43).

Finally the conclusion after (44) is a simple corollary of the properties of $D^{\epsilon}(t)$ and perturbation theory. We omit the details.

In summary, we have proved the following:

For sufficiently small $\epsilon > 0$, the function $\phi^{\epsilon}(x, y, t)$ has the following properties in the neighborhood U_{δ}:

There exists $0 < t_1 < t_2$, such that

- For $0 \le t < t_1$, $(x, y) = (0, 0)$ is the only critical point in U_{δ}. Furthermore it is a non-degenerate local minimum.
- For $t = t_1$, $(x, y) = (0, 0)$ is the only critical point in U_{δ}.
- For $t_1 < t \le t_2$, there are three critical points in U_{δ}. The point $(x, y) = (0, 0)$ is a saddle. Two other critical points are of the form (x_*, y_*), $(-x_*, -y_*)$, where $x_* > 0$, $y_* > 0$.

Remark that due to our inversion symmetry (18), if (x_*, y_*) is a critical point with $x_* \neq 0$, then $(-x_*, -y_*)$ is also a critical point.

6 Construction of ϕ_0 Satisfying (32)–(33)

We now demonstrate the existence of $\phi_0 = \phi_0(x, y)$ which satisfies conditions (32)–(33) and also has inversion symmetry (18).

By (17), we choose

$$\phi_0(x, y) = -\sum_{\substack{m+n \text{ is even} \\ |m|\le N, |n|\le N}} \tilde{f}_{mn}(-1)^{\frac{m+n}{2}} \cos(mx + ny). \tag{50}$$

To simplify matters, we impose the following conditions on $\tilde{f}_{mn}$:

- $\tilde{f}_{mn}$ is real-valued;
- $\tilde{f}_{mn} = 0$ if $m = 0$ or $n = 0$;
- $\tilde{f}_{mn}$ are odd in each of its variables m and n.

The above conditions imply that

$$\begin{aligned}\phi_0(x,y) = &- \sum_{\substack{1\le m,n\le N\\ m+n \text{ is even}}} \tilde{f}_{mn} \cdot \Big(2(-1)^{\frac{m+n}{2}} \cos(mx+ny) \\ &\qquad - (-1)^{\frac{-m+n}{2}} \cos(-mx+ny) - (-1)^{\frac{m-n}{2}} \cos(mx-ny)\Big) \\ = &- \sum_{\substack{1\le m,n\le N\\ m+n \text{ is even}}} 2\tilde{f}_{mn}(-1)^{\frac{m+n}{2}} \Big(\cos(mx+ny) - (-1)^n \cos(mx-ny)\Big).\end{aligned} \tag{51}$$

Define

$$f_{mn} = -2\tilde{f}_{mn}(-1)^{\frac{m+n}{2}}.$$

Then we have

$$\phi_0(x,y) = \sum_{\substack{1\le m,n\le N\\ m+n \text{ is even}}} f_{mn}\Big(\cos(mx+ny) - (-1)^n \cos(mx-ny)\Big), \tag{52}$$

where f_{mn} are the coefficients to be determined.

Now recall the conditions (32) and (33) and choose

$$\begin{aligned} a_3(0) &= 2, \\ a_1(0) &= a_2(0) = 1, \\ b_1(0) &= \frac{r_1}{24} > 0, \\ b_2(0) &= b_3(0) = b_4(0) = b_5(0) = 0, \end{aligned} \tag{53}$$

where r_1 is a parameter whose value will be specified later.

We still have to check the second condition in (32). This condition can be simplified a little bit. By (53),

$$\begin{aligned} &\frac{d}{dt}\left(a_3^2(t) - 4a_1(t)a_2(t)\right)\Big|_{t=0} \\ &= 4(\dot{a}_3(0) - \dot{a}_1(0) - \dot{a}_2(0)) \\ &= 2\left(\frac{\partial^3 \phi}{\partial t \partial x \partial y}(0,0,0) - \left(\frac{\partial}{\partial t}\Delta\phi\right)(0,0,0)\right). \end{aligned}$$

By (1), (19), (16) and (53), we have

$$\begin{aligned}\left(\frac{\partial}{\partial t}\Delta\phi\right)(0,0,0) &= \frac{\partial}{\partial t}\Delta\psi\left(\frac{\pi}{2},\frac{\pi}{2},0\right)\\ &= \Delta^2\psi_0\left(\frac{\pi}{2},\frac{\pi}{2}\right) + (\nabla^{\perp}\psi_0)\left(\frac{\pi}{2},\frac{\pi}{2}\right)\cdot(\nabla\Delta\psi_0)\left(\frac{\pi}{2},\frac{\pi}{2}\right)\\ &= r_1.\end{aligned}$$

Similarly

$$\frac{\partial^3\phi}{\partial t\partial x\partial y}(0,0,0) = \left(\partial_{xy}\Delta^{-1}(\nabla^{\perp}\phi_0\cdot\nabla\Delta\phi_0)\right)(0,0).$$

Therefore the condition

$$\frac{d}{dt}\Big(a_3^2(t) - 4a_1(t)a_2(t)\Big)\Big|_{t=0} > 0$$

is equivalent to

$$\partial_{xy}\Delta^{-1}(\nabla^{\perp}\phi_0\cdot\nabla\Delta\phi_0)(0,0) > r_1. \tag{54}$$

Our goal is to find (f_{mn}) in (52) such that both (53) and (54) hold. In our formulae below, the summation is understood to be in the region $\{(m,n): 1\le m,n\le N$ and $m+n$ is even$\}$. In terms of f_{mn}, the conditions (53) now take the form

$$\begin{aligned}&\sum f_{mn}\cdot mn\cdot(1+(-1)^n) = -2,\\ &\sum f_{mn}\cdot m^2\cdot(1-(-1)^n) = -1,\\ &\sum f_{mn}\cdot n^2\cdot(1-(-1)^n) = -1,\\ &\sum f_{mn}\cdot m^4\cdot(1-(-1)^n) = r_1,\\ &\sum f_{mn}\cdot m^3 n\cdot(1+(-1)^n) = 0,\\ &\sum f_{mn}\cdot m^2 n^2\cdot(1-(-1)^n) = 0,\\ &\sum f_{mn}\cdot mn^3\cdot(1+(-1)^n) = 0,\\ &\sum f_{mn}\cdot n^4\cdot(1-(-1)^n) = 0.\end{aligned} \tag{55}$$

Due to the factors $(1\pm(-1)^n)$ which can vanish depending on the parity of n in the summation, we distinguish two types of coefficients. We shall say f_{mn} is even if both m and n are even. Otherwise f_{mn} is called odd. Notice that due to the constraint that $m+n$ is even we shall only have either odd or even coefficients.

We consider first the equations for even coefficients. From (55) we only need

$$\sum_{\substack{m,n\geq 2\\ m,n \text{ are even}}} f_{mn}\cdot mn = -1,$$

$$\sum_{\substack{m,n\geq 2\\ m,n \text{ are even}}} f_{mn}\cdot m^3 n = 0,$$

$$\sum_{\substack{m,n\geq 2\\ m,n \text{ are even}}} f_{mn}\cdot mn^3 = 0, \tag{56}$$

Now we assume that we only have two nonzero even coefficients f_{22} and f_{44}. Then from (56) we get

$$f_{22}\cdot 2^2 + f_{44}\cdot 4^2 = -1,$$

$$f_{22}\cdot 2^4 + f_{44}\cdot 4^4 = 0.$$

A simple computation gives that

$$f_{22} = -1/3, \quad f_{44} = 1/48; \tag{57}$$

Next we turn to odd coefficients.

From (55), we get

$$\sum_{\substack{1\leq m,n\leq N\\ m,n \text{ are odd}}} f_{mn}\cdot m^2 = -\frac{1}{2},$$

$$\sum_{\substack{1\leq m,n\leq N\\ m,n \text{ are odd}}} f_{mn}\cdot n^2 = -\frac{1}{2},$$

$$\sum_{\substack{1\leq m,n\leq N\\ m,n \text{ are odd}}} f_{mn}\cdot m^4 = \frac{r_1}{2},$$

$$\sum_{\substack{1\leq m,n\leq N\\ m,n \text{ are odd}}} f_{mn}\cdot m^2 n^2 = 0,$$

$$\sum_{\substack{1\leq m,n\leq N\\ m,n \text{ are odd}}} f_{mn}\cdot n^4 = 0, \tag{58}$$

To simplify matters, we assume that the only nonzero odd coefficients are f_{11}, f_{31}, f_{33}, f_{15}, f_{51}.

Let r_2 be another parameter whose value will be specified later. We shall choose $f_{51} = r_2$ and add this condition to (58). For the coefficients f_{11}, f_{31}, f_{33}, f_{15}, f_{51} we then have the matrix equation

$$\begin{pmatrix} 1 & 1 & 3^2 & 3^2 & 1 & 5^2 \\ 1 & 3^2 & 1 & 3^2 & 5^2 & 1 \\ 1 & 1 & 3^4 & 3^4 & 1 & 5^4 \\ 1 & 3^2 & 3^2 & 9^2 & 5^2 & 5^2 \\ 1 & 3^4 & 1 & 3^4 & 5^4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f_{11} \\ f_{13} \\ f_{31} \\ f_{33} \\ f_{15} \\ f_{51} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{r_1}{2} \\ 0 \\ 0 \\ r_2 \end{pmatrix} \tag{59}$$

Choose $r_1 = \frac{1}{10}$ and $r_2 = -10$. From (59), we obtain

$$\begin{pmatrix} f_{11} \\ f_{13} \\ f_{31} \\ f_{33} \\ f_{15} \\ f_{51} \end{pmatrix} = \begin{pmatrix} -580.5698 \\ 90.0012 \\ 90.0008 \\ -6.6598 \\ -10.0001 \\ -10.0000 \end{pmatrix} \tag{60}$$

We have completely solved (53). It remains to check the condition (54). To simplify the computation, we rewrite (52) as

$$\begin{aligned} \phi_0(x,y) = & \sum_{\substack{1 \le m,n \le N \\ m+n \text{ is even}}} f_{mn} \cdot \frac{e^{i(mx+ny)} + e^{-i(mx+ny)}}{2} \\ & + \sum_{\substack{1 \le m,n \le N \\ m+n \text{ is even}}} f_{mn} \cdot (-1)^{n+1} \cdot \frac{e^{i(mx-ny)} + e^{-i(mx-ny)}}{2}. \\ = & \sum_{|m| \le N, |n| \le N} g_{mn} e^{i(mx+ny)}, \end{aligned} \tag{61}$$

where the coefficients g_{mn} satisfy

- $g_{mn} = 0$ if $(m+n)$ is not even or $m = 0$ or $n = 0$.
- $g_{mn} = \frac{1}{2} f_{|m|,|n|}$ if $mn > 0$.
- $g_{mn} = \frac{1}{2} f_{|m|,|n|} \cdot (-1)^{n+1}$ if $mn < 0$.

To find the LHS of (54), we use the coefficients g_{mn} and calculate

$$\Delta^{-1}\nabla\phi_0(x,y) = \sum_{|m|\le N, |n|\le N} g_{mn}\cdot i\cdot \frac{\binom{m}{n}}{(-1)(m^2+n^2)} e^{i(mx+ny)},$$

$$\nabla^{\perp}\phi_0(x,y) = \sum_{|\tilde m|\le N, |\tilde n|\le N} g_{\tilde m\tilde n}\cdot i\cdot \begin{pmatrix} -\tilde n \\ \tilde m \end{pmatrix}\cdot e^{i(\tilde m x+\tilde n y)}.$$

Hence

$$(\nabla^{-1}\nabla\phi_0\cdot\nabla^{\perp}\phi_0)(x,y) = \sum_{\substack{|m|\le N, |n|\le N\\ |\tilde m|\le N, |\tilde n|\le N}} g_{mn}g_{\tilde m\tilde n}\cdot\frac{\tilde m n - m\tilde n}{m^2+n^2}\cdot e^{i\left((m+\tilde m)x+(n+\tilde n)y\right)}. \tag{62}$$

Note that in the summation of the RHS of (62), the zero-th mode is not present since if $m=-\tilde m$, $n=-\tilde n$ then $\tilde m n - m\tilde n = 0$.

We then apply the operator $\partial_{xy}\Delta^{-1}$ to both sides of (62) to obtain

$$\begin{aligned}&\partial_{xy}\Delta^{-1}\left(\Delta^{-1}\nabla\phi_0\cdot\nabla^{\perp}\phi_0\right)\Big|_{(x,y)=(0,0)}\\ &\quad = \sum_{\substack{|m|\le N, |n|\le N\\ |\tilde m|\le N, |\tilde n|\le N}} g_{mn}g_{\tilde m\tilde n}\cdot\frac{\tilde m n - m\tilde n}{m^2+n^2}\cdot\frac{(m+\tilde m)(n+\tilde n)}{(m+\tilde m)^2+(n+\tilde n)^2}.\end{aligned} \tag{63}$$

By using (57), (60), (61), and (63) and a tedious calculation, we obtain

$$\text{LHS of (54)} = 0.1436 > 0.1 = r_1.$$

Clearly this gives us all the needed estimates.

We have finished the construction of the desired initial data ϕ_0 needed in Stage 1. The proof of Theorem 2 is now completed.

7 Proof of Theorem 3

In this section we give the proof of Theorem 3. The argument is similar to the proof of Theorem 2 and is again done in two stages. We sketch the details as follows.

- Stage 1: degenerate case. Recall the reduced critical point equation,

$$-(a_3^2-4a_1a_2)+8a_2b_1x^2+O(t)\cdot O(x^2)+O(x^4)=0. \tag{64}$$

Choose $\phi_0 = \phi_0(x, y)$ so that

$$a_3(0)^2 - 4a_1(0)a_2(0) = 0,$$
$$\frac{d}{dt}\left(a_3^2(t) - 4a_1(t)a_2(t)\right)\Big|_{t=0} < 0,$$
$$a_2(0) > 0,\ a_3(0) > 0,\ b_1(0) > 0, \tag{65}$$

and also

$$b_2(0) = b_3(0) = b_4(0) = b_5(0) = 0. \tag{66}$$

The possibility of choosing ϕ_0 with properties (65)–(66) will be shown in Sect. 8. Assume for the moment that these conditions are met, then for sufficiently small $t_2 > 0$, we have for $0 < t \le t_2$,

$$A_3'' \ge a_3(t) \ge A_3' > 0,$$
$$A_2'' \ge a_2(t) \ge A_2' > 0,$$
$$B_1'' \ge \frac{d}{dt}\Big(4a_1(t)a_2(t) - a_3^2(t)\Big) \ge B_1' > 0,$$
$$B_2'' \ge b_1(t) \ge B_2' > 0, \tag{67}$$

where A_i', A_i'', B_i', B_i'' are constants.

By (65)–(67), we have for $0 < t \le t_2$

$$const \cdot t \le 4a_1(t)a_2(t) - a_3^2(t) \le const \cdot t,$$

and also

$$8a_2(t)b_1(t) \sim const.$$

It follows that for $0 < t \le t_2$, the equation (64) is of the form

$$O(t) + O(1) \cdot x^2 + O(t) \cdot O(x^2) + O(x^4) = 0 \tag{68}$$

which clearly has no real-valued solution for $0 < t \le t_2$.

- Stage 2: a perturbation argument. In stage 2 we perturb the initial data considered in Stage 1 so that initially we will have three critical points.

To this end, consider $\tilde{\phi}_0 = \tilde{\phi}_0(x, y) \in C^\infty$ with the following properties:

$$\tilde{\phi}_0(x, y) = \tilde{\phi}_0(-x, -y), \quad \forall\, x, y,$$

$$\frac{\partial^4 \tilde{\phi}_0}{\partial x^m \partial y^n}\Big|_{(x,y)=(0,0)} = 0, \quad \forall\, m + n = 4, 0 \leq m \leq 4,$$

$$\frac{\partial^2 \tilde{\phi}_0}{\partial x \partial y}\Big|_{(x,y)=(0,0)} = 0, \quad \frac{\partial^2 \tilde{\phi}_0}{\partial x^2}\Big|_{(x,y)=(0,0)} > 0, \quad \frac{\partial^2 \tilde{\phi}_0}{\partial y^2}\Big|_{(x,y)=(0,0)} > 0. \tag{69}$$

Fix $\phi_0 = \phi_0(x, y)$ taken from Stage 1 which has the properties (65)–(66) and consider the perturbation by $\tilde{\phi}_0$ having the form

$$\tilde{\phi}_0^\epsilon(x, y) = \phi_0(x, y) - \epsilon\tilde{\phi}_0(x, y), \tag{70}$$

where $\epsilon > 0$ is sufficiently small.

Denote the corresponding solution of the main equation (1) (in the shifted coordinates) by $\phi^\epsilon = \phi^\epsilon(x, y, t)$. Expand $\phi^\epsilon(x, y, t)$ in the form

$$\begin{aligned}\phi^\epsilon(x, y, t) = {} & \phi^\epsilon(0, 0, t) + a_1^\epsilon(t)x^2 + a_2^\epsilon y^2 + a_3^\epsilon xy \\ & + b_1^\epsilon(t)x^4 + b_2^\epsilon(t)y^4 + b_3^\epsilon(t)x^3 y + b_4^\epsilon(t)x^2 y^2 + b_5^\epsilon(t)xy^3 \\ & + \tilde{\epsilon}(x, y, t)\end{aligned} \tag{71}$$

where $\tilde{\epsilon}$ satisfies an estimate similar to (20).

We now check that $\phi^\epsilon(x, y, t)$ has the desired properties needed in Theorem 3.

(a) At $t = 0$, $\phi^\epsilon(x, y, 0)$ has three critical points in the neighborhood U_δ. Also $(0, 0)$ is a saddle point.

To prove this, we note that due to (65), (66) and (69), the reduced critical point equation for $\phi^\epsilon(x, y, t)$ takes the form

$$(a_3^\epsilon(0))^2 - 4a_1^\epsilon(0)a_2^\epsilon(0) - 8a_2^\epsilon(0)b_1^\epsilon(0)x^2 + O(x^4) = 0. \tag{72}$$

Denote

$$\tilde{a}_1 = \frac{\partial^2 \tilde{\phi}_0}{\partial x^2}\Big|_{(x,y)=(0,0)} > 0,$$

$$\tilde{a}_2 = \frac{\partial^2 \tilde{\phi}_0}{\partial y^2}\Big|_{(x,y)=(0,0)} > 0.$$

By (65), (69), and (70), we have

$$\begin{aligned}
&(a_3^\epsilon(0))^2 - 4a_1^\epsilon(0)a_2^\epsilon(0)\\
&\quad = a_3(0)^2 + O(\epsilon^2) - 4(a_1(0) - \epsilon\tilde{a}_1)(a_2(0) - \epsilon\tilde{a}_2)\\
&\quad = 4(a_1(0)\tilde{a}_2 + a_2(0)\tilde{a}_1)\epsilon + O(\epsilon^2). \qquad (73)
\end{aligned}$$

On the other hand, for sufficiently small $\epsilon > 0$, by using (72), (69), and (70), we have

$$\begin{aligned}
a_2^\epsilon(0)b_1^\epsilon(0) &= (a_2(0) - O(\epsilon)) \cdot (b_1(0) + O(\epsilon^2))\\
&= a_2(0)b_1(0) - O(\epsilon)\\
&\sim const. \qquad (74)
\end{aligned}$$

Therefore by (73) and (74), the equation (72) takes the form

$$O(1)\epsilon - O(1) \cdot x^2 + O(x^4) = 0,$$

or simply

$$O(1) \cdot \epsilon - O(1) \cdot O(x^2) = 0.$$

It is clear that for $\epsilon > 0$ sufficiently small this equation has two real-valued solutions in U_δ.

To verify that $(0, 0)$ is a saddle point at $t = 0$, we observe that by (73), for sufficiently small $\epsilon > 0$,

$$(a_3^\epsilon(0))^2 - 4a_1^\epsilon(0)a_2^\epsilon(0) > 0. \qquad (75)$$

Also we have by (65)

$$a_1^\epsilon(0) > 0, \quad a_2^\epsilon(0) > 0. \qquad (76)$$

Equations 75 and 76 show that the Hessian matrix

$$\begin{pmatrix} a_1^\epsilon(0) & \frac{1}{2}a_3^\epsilon(0)\\ \frac{1}{2}a_3^\epsilon(0) & a_2^\epsilon(0) \end{pmatrix}$$

has one positive eigen-value and one negative eigen-value. Hence $(0, 0)$ is a saddle.

(b) Consider the function

$$D^\epsilon(t) = (a_3^\epsilon(t))^2 - 4a_1^\epsilon(t)a_2^\epsilon(t).$$

It will be proven below that for sufficiently small $\epsilon > 0$, the following holds: There exists unique $t_1 = t_1(\epsilon) > 0$ such that

$$\begin{aligned} &D^\epsilon(t) > 0, \quad \text{for } 0 \leq t < t_1, \\ &D^\epsilon(t) = 0, \quad \text{for } t = t_1, \\ &D^\epsilon(t) < 0, \quad \text{for } t_1 < t \leq t_2. \end{aligned} \tag{77}$$

Furthermore, the reduced critical-point equation

$$D^\epsilon(t) - 8a_2^\epsilon(t)b_1^\epsilon(t)x^2 + O(t) \cdot O(x^2) + O(x^4) = 0 \tag{78}$$

has

- Two nonzero solutions for $0 \leq t < t_1$,
- Exactly one solution given by $x = 0$ for $t = t_1$,
- No solutions for $t_1 < t \leq t_2$.

To prove (77), we recall the bound (67), where for $0 \leq t \leq t_2$

$$B_1'' \geq \frac{d}{dt}\left(a_3^2(t) - 4a_1(t)a_2(t)\right) \geq B_1' > 0. \tag{79}$$

Since our initial data are given by

$$\phi_0^\epsilon(x, y) = \phi_0(x, y) + \epsilon\tilde{\phi}_0(x, y),$$

it follows from simple perturbation theory that

$$2B_1'' \geq \frac{d}{dt}\left(4a_1^\epsilon(t)a_2^\epsilon(t) - (a_3^\epsilon(t))^2\right) \geq \frac{B_1'}{2} > 0, \tag{80}$$

for any $0 \leq t \leq t_2$.

This means in particular that $D^\epsilon(t)$ is strictly decreasing for $0 \leq t \leq t_2$.

By (73), we have for $t = 0$ and ϵ sufficiently small,

$$D^\epsilon(0) > 0. \tag{81}$$

On the other hand for $t = t_2$, by using the analysis from Stage 1, we have

$$(a_3(t_2))^2 - 4a_1(t_2)a_2(t_2) < 0.$$

Since

$$D^\epsilon(t_2) = (a_3(t_2))^2 - 4a_1(t_2)a_2(t_2) + O(\epsilon),$$

it follows easily that for ϵ sufficiently small

$$D^{\epsilon}(t_2) < 0. \tag{82}$$

Now (80)–(82) easily yield (77).

8 Construction of ϕ_0 Satisfying (65)–(66)

We now demonstrate the existence of $\phi_0 = \phi_0(x, y)$ which satisfies conditions (65)–(66). The construction is similar to the one in Sect. 6 and therefore we shall only sketch the details.

Choose ϕ_0 in the form

$$\phi_0(x, y) = \sum_{\substack{1 \le m,n \le N \\ m+n \text{ is even}}} f_{mn} \left(\cos(mx + ny) - (-1)^n \cos(mx - ny)\right), \tag{83}$$

where f_{mn} are the coefficients to be determined.

Now recall the conditions (65) and (66) and set

$$\begin{aligned}
a_3(0) &= 2, \\
a_1(0) &= a_2(0) = 1, \\
b_1(0) &= \frac{r_1}{24} > 0, \\
b_2(0) &= b_3(0) = b_4(0) = b_5(0) = 0,
\end{aligned} \tag{84}$$

where r_1 is a parameter whose value will be specified later.

The second condition in (65) simplifies to

$$\partial_{xy} \Delta^{-1}(\nabla^{\perp}\phi_0 \cdot \nabla \Delta\phi_0)(0, 0) < r_1. \tag{85}$$

Our goal is to find (f_{mn}) in (83) such that both (84) and (85) hold. In our formulae below, the summation is understood to be in the region $\{(m, n) : 1 \le m, n \le N$ and $m + n$ is even$\}$. In terms of f_{mn}, the conditions (84) now take the form

$$\begin{aligned}
&\sum f_{mn} \cdot mn \cdot (1 + (-1)^n) = -2, \\
&\sum f_{mn} \cdot m^2 \cdot (1 - (-1)^n) = -1, \\
&\sum f_{mn} \cdot n^2 \cdot (1 - (-1)^n) = -1, \\
&\sum f_{mn} \cdot m^4 \cdot (1 - (-1)^n) = r_1,
\end{aligned}$$

$$\begin{aligned}
&\sum f_{mn} \cdot m^3 n \cdot (1 + (-1)^n) = 0,\\
&\sum f_{mn} \cdot m^2 n^2 \cdot (1 - (-1)^n) = 0,\\
&\sum f_{mn} \cdot m n^3 \cdot (1 + (-1)^n) = 0,\\
&\sum f_{mn} \cdot n^4 \cdot (1 - (-1)^n) = 0.
\end{aligned} \tag{86}$$

Due to the factors $(1 \pm (-1)^n)$ which can vanish depending on the parity of n in the summation, we distinguish two types of coefficients. We shall say f_{mn} is even if both m and n are even. Otherwise f_{mn} is called odd. Notice that due to the constraint that $m + n$ is even we shall only have either odd or even coefficients.

Consider first the equations for even coefficients. From (86) we only need

$$\begin{aligned}
&\sum_{\substack{m,n\geq 2\\ m,n \text{ are even}}} f_{mn} \cdot mn = -1,\\
&\sum_{\substack{m,n\geq 2\\ m,n \text{ are even}}} f_{mn} \cdot m^3 n = 0,\\
&\sum_{\substack{m,n\geq 2\\ m,n \text{ are even}}} f_{mn} \cdot mn^3 = 0,
\end{aligned} \tag{87}$$

Now we assume that we only have two nonzero even coefficients f_{22} and f_{44}. Then from (87) we get

$$\begin{aligned}
f_{22} \cdot 2^2 + f_{44} \cdot 4^2 &= -1,\\
f_{22} \cdot 2^4 + f_{44} \cdot 4^4 &= 0.
\end{aligned}$$

A simple computation gives that

$$f_{22} = -1/3, \quad f_{44} = 1/48; \tag{88}$$

Next we turn to odd coefficients.

From (86), we get

$$\begin{aligned}
&\sum_{\substack{1\leq m,n\leq N\\ m,n \text{ are odd}}} f_{mn} \cdot m^2 = -\frac{1}{2},\\
&\sum_{\substack{1\leq m,n\leq N\\ m,n \text{ are odd}}} f_{mn} \cdot n^2 = -\frac{1}{2},
\end{aligned}$$

$$\sum_{\substack{1\le m,n\le N\\ m,n \text{ are odd}}} f_{mn}\cdot m^4 = \frac{r_1}{2},$$

$$\sum_{\substack{1\le m,n\le N\\ m,n \text{ are odd}}} f_{mn}\cdot m^2n^2 = 0,$$

$$\sum_{\substack{1\le m,n\le N\\ m,n \text{ are odd}}} f_{mn}\cdot n^4 = 0, \tag{89}$$

To simplify matters, we assume that the only nonzero odd coefficients are f_{11}, f_{31}, f_{33}, f_{15}, f_{51}.

Let r_2 be another parameter whose value will be specified later. We shall choose $f_{51} = r_2$ and add this condition to (89). For the coefficients f_{11}, f_{31}, f_{33}, f_{15}, f_{51} we then have the matrix equation

$$\begin{pmatrix} 1 & 1 & 3^2 & 3^2 & 1 & 5^2 \\ 1 & 3^2 & 1 & 3^2 & 5^2 & 1 \\ 1 & 1 & 3^4 & 3^4 & 1 & 5^4 \\ 1 & 3^2 & 3^2 & 9^2 & 5^2 & 5^2 \\ 1 & 3^4 & 1 & 3^4 & 5^4 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f_{11} \\ f_{13} \\ f_{31} \\ f_{33} \\ f_{15} \\ f_{51} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{r_1}{2} \\ 0 \\ 0 \\ r_2 \end{pmatrix} \tag{90}$$

Choose $r_1 = r_2 = 1$. From (90), we obtain

$$\begin{pmatrix} f_{11} \\ f_{13} \\ f_{31} \\ f_{33} \\ f_{15} \\ f_{51} \end{pmatrix} = \begin{pmatrix} 57.3646 \\ -8.9883 \\ -8.9922 \\ 0.6727 \\ 0.9987 \\ 1.0000 \end{pmatrix} \tag{91}$$

We have completely solved (84). It remains to check the condition (85). For this purpose, we rewrite (83) as

$$\begin{aligned} \phi_0(x,y) = &\sum_{\substack{1\le m,n\le N\\ m+n \text{ is even}}} f_{mn}\cdot \frac{e^{i(mx+ny)}+e^{-i(mx+ny)}}{2} \\ &+ \sum_{\substack{1\le m,n\le N\\ m+n \text{ is even}}} f_{mn}\cdot(-1)^{n+1}\cdot \frac{e^{i(mx-ny)}+e^{-i(mx-ny)}}{2}. \\ = &\sum_{|m|\le N,|n|\le N} g_{mn}e^{i(mx+ny)}, \end{aligned} \tag{92}$$

where the coefficients g_{mn} satisfy

- $g_{mn} = 0$ if $(m + n)$ is not even or $m = 0$ or $n = 0$.
- $g_{mn} = \frac{1}{2} f_{|m|,|n|}$ if $mn > 0$.
- $g_{mn} = \frac{1}{2} f_{|m|,|n|} \cdot (-1)^{n+1}$ if $mn < 0$.

In terms of the coefficients g_{mn}, the LHS of (85) takes the form

$$(\nabla^{-1}\nabla\phi_0 \cdot \nabla^{\perp}\phi_0)(x, y) = \sum_{\substack{|m|\le N,|n|\le N\\ |\tilde{m}|\le N,|\tilde{n}|\le N}} g_{mn}g_{\tilde{m}\tilde{n}} \cdot \frac{\tilde{m}n - m\tilde{n}}{m^2 + n^2} \cdot e^{i\left((m+\tilde{m})x+(n+\tilde{n})y\right)}.$$

By a tedious calculation, we obtain

$$\text{LHS of (85)} = -0.1420 < 1 = r_1.$$

Clearly this gives us all the needed estimates.

We have finished the construction of the desired initial data ϕ_0 needed in Stage 1 of Sect. 7. The proof of Theorem 3 is now completed.

Acknowledgements The authors thank V. Yakhot for useful remarks and discussions. The first author is supported in part by a start-up fund from University of British Columbia. The financial support from NSF, grant DMS 0908032, given to the first author and grant DMS060096, given to the second author are highly appreciated.

References

1. V.I. Arnold, Lectures on bifurcations and versal families. A series of articles on the theory of singularities of smooth mappings. Uspehi Mat. Nauk 27 **5**(167), 119–184 (1972)
2. E. Dinaburg, D. Li, Ya.G. Sinai, Navier–Stokes system on the flat cylinder and unit square with slip boundary conditions. Commun. Contemp. Math. **12**(2), 325–349 (2010)
3. C. Foias, R. Temam, Gevrey class regularity for the solutions of the Navier–Stokes equations. J. Funct. Anal. **87**(2), 359–369 (1989)
4. J.C. Mattingly, Ya.G. Sinai, An elementary proof of the existence and uniqueness theorem for the Navier–Stokes equations. Commun. Contemp. Math. **1**(4), 497–516 (1999)

Arnold Diffusion by Variational Methods

John N. Mather

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract In this paper, we correct results announced in Mather (J Math Sci NY 124:5275–5289, 2003) and make some observations on the proofs of these results. The principal result, Theorem 1, is a strong form of Arnold diffusion in two and one half degrees of freedom, under suitable genericity hypotheses. After (Mather, J Math Sci NY 124:5275–5289, 2003) appeared, we realized that there is an oversight in our planned proof. Because of the oversight, the genericity conditions that we imposed on U in Mather (J Math Sci NY 124:5275–5289, 2003) are not enough. In this paper, we state further genericity conditions, which are enough for our revised proof. In addition, we note that a slightly stronger differentiability hypothesis than we stated in Mather (J Math Sci NY 124:5275–5289, 2003) is needed. In the later sections of this paper, we make some observations related to the proof of Theorem 1. The complete (revised) proof will appear elsewhere.

1 Introduction

In [4], I announced results about Arnold diffusion. It has taken me far longer than I ever imagined possible to write up the proofs and I am still not finished. In addition, I have found some errors in the results that I announced in [4]. In Sect. 2 of this paper, I will correct the errors that I found and then make some observations about the proofs in later sections.

I have circulated several versions of a preprint "Arnold diffusion, II" to whoever asked me for a copy. Each version of the preprint contains part of the proof of the

J.N. Mather (✉)
Princeton University, Princeton, NJ 08544
e-mail: jnm@math.princeton.edu

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_11, © Springer-Verlag Berlin Heidelberg 2012

results announced in [4]. Each of the later versions is the previous version with more material added at the end, together with minor reorganization of the earlier material. All results stated in any of these versions are proved in that version.

When I wrote [4], my plan was to write a paper "Arnold Diffusion, II," giving the proofs of the results announced in [4]. At any time, the version of the preprint that I made available was the part of the planned paper written up to that time.

I have now changed my plan because it was taking me much too long to write the complete proof. My new plan is to publish the complete proof as a series of papers, "Arnold Diffusion, II, III," etc. I plan to publish parts of the preprint as "Arnold Diffusion, II," etc. The later parts of this series will contain material not yet written.

The reader will need to have [4] before him to read Sect. 2. I have tried to write the rest of this paper so that it can be read without reference to [4].

2 Errata for [4]

Since I wrote [4], I have discovered mistakes in the proof that necessitate errata for it.

The first mistake can be remedied by replacing "3" by " 4" in any hypothesis involving $\mathscr{L}^r$, e.g. in § 2, replace $\mathscr{L}^3$ by $\mathscr{L}^4$; $\widehat{\mathscr{L}}^3$ by $\widehat{\mathscr{L}}^4$; $\mathscr{L}^r$, $r \geq 3$ by $\mathscr{L}^r$, $r \geq 4$; and in Theorems 1c and 2c replace "Let r be ω, ∞, or an integer $\geq 3''$ by "Let r be ω, ∞, or an integer $\geq$ 4." This slightly changes the statements of the main theorems (Theorems 1 and 2 in § 2). No other changes in the statements of the main theorems are needed.

The other mistakes can be remedied by changing Definition 2 in § 12, which defines $U_\Gamma^r = U_{\Gamma,\ell_0}^r$. For Theorem 1a to be true, it is required that U_Γ^r be open and dense in $\mathscr{P}^r$ for $r \geq 3$. The argument I had in mind to prove that U_Γ^r is open contains an error. This can be remedied by changing the conditions $(C4)_\omega - (C8)_\omega$ listed in Definition 2 in § 12.

Making these changes does not change the statements of the results in § 2. Each of the theorems in § 2 asserts the existence of a function δ with certain properties. The "partial definition" of $U_{\ell_0}^r$ in § 12 amounts to asserting that there exists a function δ having not only the properties stated in Theorem 1 but also the properties given in § 12 by conditions $(C1) - (C3)$ and $(C4)_\omega - (C8)_\omega$ for rational $\omega \in \Gamma$ with small denominator. These are stated as properties of $U_\Gamma^r = U_{\Gamma,\ell_0}^r$, where Γ stands for one of Γ_1 , $\cdots$, Γ_n. They may be interpreted as conditions on δ since $U_{\ell_0}^r = U_{\Gamma_1,\ell_0}^r \cap \cdots \cap U_{\Gamma_n,\ell_0}^r = \{\epsilon P : \epsilon > 0,\ P \in \mathscr{P}^r, \text{ and } \delta(P, \ell_0) > 0\}$.

One problem is that the shortest geodesic γ referred to in $(C8)_\omega$ need not be simple. In the case that it is not simple, the conditions in § 12 do not define an open set. There are also other problems, described below. These problems can be remedied by replacing the conditions $(C4)_\omega - (C8)_\omega$ by the conditions $(C4)'_\omega - (C10)'_\omega$ listed below.

To correctly state what I can prove, I need to strengthen $(C4)_\omega$. In § 10, I defined a function $K_\omega : T(\mathbb{T}^2_\omega) \to \mathbb{R}$ whose restriction to each fiber $T(\mathbb{T}^2_\omega)_\varphi$ of the tangent bundle $T(\mathbb{T}^2_\omega)$ of $\mathbb{T}^2_\omega$ is quadratic and positive definite. The physical interpretation of K_ω is that it is the "kinetic energy" associated to a Riemannian metric $g_\omega := 2K_\omega$ on $\mathbb{T}^2_\omega$.

In § 10, I also defined a function $P_\omega : \mathbb{T}^2_\omega \to \mathbb{R}$. I recall the definitions of $\mathbb{T}^2_\omega$, K_ω and P_ω in § 3 below. In § 12, I stated $(C4)_\omega$ as follows:

$(C4)_\omega$ The function P_ω on $\mathbb{T}^2_\omega$ has only one global minimum n_ω and is non-degenerate in the sense of Morse at n_ω, i.e. the quadratic form $d^2P_\omega(n_\omega)$ is non-singular.

Both $g(n_\omega)$ and $d^2P_\omega(n_\omega)$ are positive definite quadratic forms on the two dimensional vector space $T(\mathbb{T}_\omega)_{n_\omega}$. A basic result in linear algebra states that these quadratic forms can be simultaneously diagonalized. In the present context, it is convenient to state this result as follows: *There exists a C^∞ local coordinate system x, y defined on an open neighborhood of n_ω in $T(\mathbb{T}^2_\omega)$ such that $g_\omega(n_\omega) = dx^2+dy^2$ and $d^2P_\omega(n_\omega) = \lambda dx^2+\mu dy^2$, where* $0 < \lambda \leq \mu$. The numbers λ and μ are called the *eigenvalues of $d^2P_\omega(n_\omega)$ with respect to* $g_\omega(n_\omega)$. The fact that these eigenvalues are positive is equivalent to the condition that n_ω is non-degenerate in the sense of Morse. The condition $\lambda \leq \mu$ is a labeling convention, i.e. if the two eigenvalues are different, we label the smaller one λ and the larger one μ.

To have an Arnold diffusion result that I can prove, I need to assume the following strengthened condition:

$(C4)'_\omega$ $0 < \lambda < \mu$.

In [6], I set $E_0 := -P_\omega(n_\omega)$ and $g_E := (P_\omega + E)g_\omega$, for $E \geq E_0$. For $E > E_0$, g_E is a Riemannian metric. On the other hand, g_{E_0} vanishes at n_ω, although it is a Riemannian metric on the complement of n_ω in $\mathbb{T}^2_\omega$. For $E > E_0$, a g_E-shortest closed curve in h_0 is simple, where h_0 is an indivisible element of $H_1(\mathbb{T}^2_\omega; \mathbb{R})$, whose definition (from [4]). I recall in § 5 below. When $E = E_0$, this is no longer true, although I thought it is true when I wrote [4]. Although I did not state there that g_{E_0}-shortest closed curve in h_0 is simple, my belief that this is so led to several errors.

To remedy these errors, I replace $(C8)_\omega$ with conditions $(C5)'_\omega - (C7)'_\omega$ below.

$(C5)'_\omega$ There is only one g_{E_0}-shortest closed curve γ in h_0.

Under these conditions, there are three possibilities:

- γ is simple and does not pass through n_ω,
- γ is simple and passes through n_ω.
- $\gamma = p\gamma_1 + q\gamma_2$, where p and q are relatively prime positive integers; γ_1 and γ_2 are simple closed curves that pass through n_ω but do not meet elsewhere; and γ_1 and γ_2 are g_{E_0}-shortest closed curves in homology classes h_1 and h_2, which generate the abelian group $H_1(\mathbb{T}^2_\omega; \mathbb{Z})$.

That γ has one of these forms is a consequence of Theorem 2 in [6], together with the assumption that γ is unique. Lemma 1 in [6] implies that if θ is an element of $\mathbb{T}^2_\omega$ sufficiently close to n_ω then there is a unique g_{E_0}-shortest curve Γ_θ in $\mathbb{T}^2_\omega$

connecting θ to n_ω. Lemma 2 in [6] implies that there exists a C^1 curve C in $\mathbb{T}^2_\omega$ passing through n_ω and tangent to the y-axis at n_ω such that the following holds: If $\theta \in C$ and is sufficiently close to n_ω then $\Gamma_\theta \in C$. If $\theta \notin C$ and is sufficiently close to n_ω the Γ_ω is tangent to the x-axis at n_ω.

From this, it follows that the next condition holds generically:

$(C6)'_\omega$ If the g_{E_0}-shortest curve γ in h_0 is simple and passes through n_ω then it is tangent to the $x-$axis at n_ω. If γ is not simple and so has the form $p\gamma_1 + q\gamma_2$ then γ_1 and γ_2 are tangent to the x-axis at n_ω.

In the definition of "nondegenerate in the sense of Morse" following the statement of $(C8)_\omega$ in [4], I asserted that the function $P \mapsto \ell_E(\gamma_P)$ is C^r. (I repeat the relevant definitions in § 5.) I meant to say that this is true for P sufficiently close to $\mu \cap \gamma$. My belief that I could prove this was mistaken. I can, however, prove that it is C^2 for P close enough to $\mu \cap \gamma$ in the case that $(C6)'_\omega$ holds. When $(C6)'_\omega$ holds, I can thus define what it means for γ to be nondegenerate in the sense of Morse in the case that γ is simple and passes through n_ω and what it means for γ_1 and γ_2 to be nondegenerate in the sense of Morse when $\gamma = p\gamma_1 + q\gamma_2$. For example, when γ is simple, I define γ to be non-degenerate in the sense of Morse if the second derivative of $P \mapsto \ell_E(\gamma_P)$ at $\gamma \cap \mu$ does not vanish. The same definition applies to γ_1 and γ_2 when $\gamma = p\gamma_1 + q\gamma_2$. Of course, when γ does not pass through n_ω, the usual definition applies.

This permits the formulation of the next condition:

$(C7)'_\omega$ If the g_{E_0}-shortest curve γ is simple, it is nondegenerate in the sense of Morse. If $\gamma = p\gamma_1 + p\gamma_2$, both γ_1 and γ_2 are nondegenerate in the sense of Morse.

The remaining conditions in the revised partial definition of U^r_Γ are slight modifications of the conditions $(C5)_\omega, (C6)_\omega$, and $(C7)_\omega$ in the original definitions. In the case that the g_{E_0}-shortest curve γ in h_0 is simple, the condition $E > E_0$ is replaced with $\widehat{E}_0 \geq E > E_0$, where $\widehat{E}_0$ is a large positive number. In the case that γ is not simple, the condition $E > E_0$ is replaced with a condition $\widehat{E}_0 > E \geq E^*_0$, where E^*_0 is a little larger than E_0. Moreover, E^*_0 and $\widehat{E}_0$ depend only on $\Gamma,\ \ell_0,\ P$, and ω. The definition of E^*_0 and $\widehat{E}_0$ is a quantitative aspect of the definition of U^r_Γ that will be postponed to a later paper.

Thus, $(C5)_\omega$ is replaced by:

$(C8)'_\omega$ Each g_E-shortest closed curve in h_0 is nondegenerate in the sense of Morse, for $\widehat{E}_0 \geq E > E_0$ in the case that γ is simple and for $\widehat{E}_0 \geq E \geq E^*_0$ in the case that γ is not simple.

Likewise, $(C6)_\omega$ is replaced by:

$(C9)'_\omega$ There are at most two g_E-shortest closed curves in h_0 for $\widehat{E}_0 \geq E > E_0$ in the case that γ is simple and for $\widehat{E}_0 \geq E \geq E^*_0$ in the case that γ is not simple.

Finally, $(C7)_\omega$ is replaced by a condition concerning an E_1 for which there are two g_{E_1}-shortest curves γ and γ' in h_0:

$(C10)'_\omega$ $d(\ell_E(\gamma_E))/dE|_{E=E_1} \neq d(\ell_E(\gamma'_E))/dE|_{E=E_1}$ for $\widehat{E}_0 \geq E_1 > E_0$ in the case that γ is simple and for $\widehat{E}_0 \geq E_1 \geq E^*_0$ in the case that γ is not simple.

Here, γ_E and γ'_E are g_E-locally shortest curves in h_0 that continue γ and γ' for E in a neighborhood of E_1.

3 Statement of the Main Theorem

In order to make it possible to read this and later sections without having read [4], we will repeat some material from there.

We recall that we announced Arnold diffusion-type results for a periodic Lagrangian in two degrees of freedom whose Lagrangian has the form $L(\theta, \dot{\theta}, t) = \ell_0(\dot{\theta}) + \epsilon P(\theta, \dot{\theta}, t)$, where $\theta = (\theta_1, \theta_2) \in \mathbb{T}^2 := \mathbb{R}^2/\mathbb{Z}^2$, $\dot{\theta} = (\dot{\theta}_1, \dot{\theta}_2)$ is a member of a 2-ball B^2 in $\mathbb{R}^2$, and $t \in \mathbb{T} := \mathbb{R}/\mathbb{Z}$.

We assumed that ℓ_0 and P are three times continuously differentiable. As we pointed out in Sect. 2, we actually need to assume that ℓ_0 is four times continuously differentiable for our proofs of Arnold diffusion to work. We assumed that $\|P\|_{C^3} = 1$ and that the Hessian matrix $d^2\ell_0(\dot{\theta})$ of second partial derivatives of ℓ_0 is positive definite for every $\dot{\theta} \in B^2$. Our announced results were for $\epsilon > 0$ small, i.e. the case when L is a small perturbation of the integrable system ℓ_0.

We let $\mathscr{L}^r$ denote the topological space of C^r functions $\ell_0 : B^2 \to \mathbb{R}$ such that $d^2\ell_0 > 0$, provided with the C^r topology. We let $\mathscr{P}^r$ denote the topological space of C^r functions $P : \mathbb{T}^2 \times B^2 \times \mathbb{T} \to \mathbb{R}$ such that $\|P\|_{C^3} = 1$, provided with the C^r topology. We let $\mathbb{R}_{++}$ denote the set of positive numbers and $\mathbb{R}_+$ the set of non-negative numbers, both provided with the usual topology.

The corrected (per Sect. 2) version of Theorem 1 can be formulated in terms of three sets U, V, and W, which satisfy:

(a) $W \subset V \subset \mathbb{R}_{++} \times U$;
(b) $U \subset \mathscr{L}^4 \times \mathscr{P}^3$;
(c) U is open and dense relative to $\mathscr{L}^4 \times \mathscr{P}^3$;
(d) there exists a continuous function $\delta : \mathscr{L}^4 \times \mathscr{P}^3 \to \mathbb{R}_+$ that is positive on U such that if $\ell_0 \in \mathscr{L}^4,\ P \in \mathscr{P}^3$, and $0 < \epsilon < \delta(\ell_0, P)$ then $(\epsilon, \ell_0, P) \in V$;
(e) W is open and dense relative to V;
(f) For any $\ell_0 \in \mathscr{L}^4$ and any $r \geq 3,\ \ U \cap (\ell_0 \times \mathscr{P}^r)$ is dense relative to $\mathscr{P}^r$, and
(g) For any $\ell_0 \in \mathscr{L}^4$ and $r \geq 3,\ \ \{(\epsilon, P) \in \mathbb{R}_{++} \times \mathscr{P}^r : (\epsilon, \ell_0, P) \in W\}$ is dense relative to $V \cap (\mathbb{R}_{++} \times \ell_0 \times \mathscr{P}^r)$.

The corrected version (per Sect. 2) of Theorem 1 may be formulated as follows:

Theorem 1. *Let $\Omega_1, \cdots, \Omega_k$ be open, nonvoid subsets of B^2. There exist sets U, V, and W that satisfy a)-g) such that if $(\epsilon, \ell_0, P) \in W$ then there exists a trajectory θ of $L = \ell_0 + \epsilon P$ that visits the Ω_i's in any pre-assigned order.*

Remark 1. Let $\theta : J \to \mathbb{T}^2$ be a trajectory of L, where J is an open subset of $\mathbb{R}$. In saying that θ visits Ω_i, we mean that there exists $t \in \Omega_i$ such that $\dot{\theta}(t) \in \Omega_i$. In saying that it visits the Ω_i's in any pre-assigned order, we mean:

Let K be a (finite or infinite) set of integers and let $\varphi : K \to \{1, \cdots, k\}$ be a mapping. Then there exists an order-preserving mapping $t : K \to J$ such that $t(\kappa) \in J$ and $\dot{\theta}(t(\kappa)) \in \Omega_{\varphi(\kappa)}$ for $\kappa \in K$.

Remark 2. The sets U, V, and W depend on the choice of sets $\Omega_1, \cdots, \Omega_k$. They are independent of the choice of pre-assigned order.

Remark 3. We can also prove a corrected version of Theorem 2 of [4], but we will not state it here. The corrections to Theorem 2 are similar to the corrections to Theorem 1 given in Sect. 2.

4 The Averaged Lagrangian Associated to a Rational Frequency

In this section, we repeat material from [4], § 10.

We consider $\omega = (\omega_1, \omega_2) \in \mathbb{Q}^2 \cap B^2$, and set $\mathbb{T}^1_\omega := \{(\lambda\omega_1,\ \lambda\omega_2,\ \lambda) \in \mathbb{T}^2 \times \mathbb{T}^1 : \lambda \in \mathbb{R}\}$, $\mathbb{T}^2_\omega := \left(\mathbb{T}^2 \times \mathbb{T}\right)/\mathbb{T}^1_\omega$, and

$$P_\omega(\varphi) = \int\limits_{(\theta,t)\in\varphi} P(\theta,\omega,t)\,d\mathscr{H}$$

Here, φ denotes an element of $\mathbb{T}^2_\omega$, i.e. a coset of T^1_ω in $\mathbb{T}^2 \times \mathbb{T}$ and $d\mathscr{H}$ denotes Haar measure on φ, normalized to have total mass 1.

The projection $\pi : \mathbb{T}^2 \times \mathbb{T} \to \mathbb{T}^2_\omega$ restricted to $\mathbb{T}^2 \times 0$ induces a vector space isomorphism $d\pi : \mathbb{R}^2 = T\mathbb{T}^2_{(\theta,t)} \to T(\mathbb{T}^2_\omega)_\varphi$, for any $\varphi \in \mathbb{T}^2_\omega$ and any $(\theta, t) \in \varphi$. This isomorphism is independent of the choice of $(\theta, t) \in \varphi$. Since ℓ_0 is a C^4 real valued function on B^2, its total second derivative $d^2\ell_0(\omega)$ is a quadratic form on $\mathbb{R}^2$. We set $K_\omega := d^2\ell_0(\omega)/2$. Thus, K_ω may be regarded as a quadratic form on $T(\mathbb{T}^2_\omega)_\varphi$ in view of the identification of $\mathbb{R}^2$ with $T(\mathbb{T}^2_\omega)_\varphi$ given by $d\pi$. Since this is defined for any $\varphi \in \mathbb{T}^2_\omega$, we have that K_ω is a real valued function on $T(\mathbb{T}^2_\omega)$.

We set $L_\omega := K_\omega + P_\omega \circ pr : T(\mathbb{T}^2_\omega) \to \mathbb{R}$, where $pr : T(\mathbb{T}^2_\omega) \to \mathbb{T}^2_\omega$ denotes the projection. This is the *averaged Lagrangian associated to* $\omega \in \mathbb{Q}^2 \cap B^2$.

5 The Averaged Lagrangian Associated to a Frequency and a Resonance of It

In this section, we repeat material from [4], § 11.

We consider $\omega \in B^2$ and $k = (k_0,\ k_1,\ k_2) \in \mathbb{Z}^3$ such that $(k_1,\ k_2) \neq (0, 0)$ and $k_0 + k_1\omega_1 + k_2\omega_2 = 0$.

We set $\Lambda = \Lambda_k := \{(\omega_1,\ \omega_2) \in \mathbb{R}^2 : k_0 + k_1\omega_1 + k_2\omega_2 = 0\}$. Thus, Λ is the line of all $\omega \in \mathbb{R}^2$ for which k is a resonance. We set $\mathbb{T}^2_\Lambda := \{(\theta_1, \theta_2, t) \in \mathbb{T}^2 \times \mathbb{T} : k_0 t + k_1\omega_1 + k_2\omega_2 = 0(\text{mod } 1)\}$, $T^1_\Lambda := (\mathbb{T}^2 \times \mathbb{T})/\mathbb{T}^2_\Lambda$, and

$$P_{\omega,\Lambda}(\varphi) := \int\limits_{(\theta,t)\in\varphi} P(\theta,\omega,t)d\mathscr{H}.$$

Here, φ denotes an element of $\mathbb{T}^1_\Lambda$, i.e. a coset of $\mathbb{T}^2_\Lambda$ in $\mathbb{T}^2 \times \mathbb{T}$ and $d\mathscr{H}$ denotes normalized Haar measure on φ.

The projection $\pi : \mathbb{T}^2 \times \mathbb{T} \to \mathbb{T}^1_\Lambda$ restricted to $\mathbb{T}^2 \times 0$ induced a vector space epimorphism $d\pi : \mathbb{R}^2 = T\mathbb{T}^2_{(\theta,t)} \to T(\mathbb{T}^1_\Lambda)_\varphi$ for any $\varphi \in \mathbb{T}^1_\Lambda$ and any $(\theta,t) \in \varphi$. This epimorphism is independent of the choice of $(\theta,t) \in \varphi$. We let $N \subset \mathbb{R}^2$ denote the null space of $d\pi$ and $N^\perp_\omega$ its orthogonal complement with respect to $d^2\ell_0(\omega)$. We set $K_{\omega,\Lambda} := (d^2\ell_0(\omega)/2)|N^\perp_\omega$. Thus, $K_{\omega,\Lambda}$ may be regarded as a quadratic form on $T(\mathbb{T}^1_\Lambda)_\varphi$ in view of the identification of $N^\perp_\omega$ with $T(\mathbb{T}^1_\Lambda)_\varphi$ given by $d\pi$. Since $K_{\omega,\Lambda}$ is defined on $T(\mathbb{T}^1_\Lambda)_\varphi$ for any $\varphi \in \mathbb{T}^1_\Lambda$, we have defined $K_{\omega,\Lambda} : T(\mathbb{T}^1_\Lambda) \to \mathbb{R}$.

We set $L_{\omega,\Lambda} := K_{\omega,\Lambda} + P_{\omega,\Lambda} \circ \; pr : T(\mathbb{T}^1_\Lambda) \to \mathbb{R}$, where $pr : T(\mathbb{T}^1_\Lambda) \to \mathbb{T}^1_\Lambda$ is the projection. This is the *averaged Lagrangian associated to $\omega \in B^2$ and the resonance k of it.*

6 Definition of *U*

In this section, we define the set U of § 2. Theorem 1 may be strengthened as follows: the set U whose existence is asserted there may be taken to be the set U defined here, provided that certain conditions stated at the end of this section hold.

We consider a resonance, i.e. $(k_0,\ k_1,\ k_2) \in \mathbb{Z}^3$ satisfying $(k_1,\ k_2) \neq (0,0)$. We let Γ be a compact line segment in $\Lambda_k \cap int\ B^2$, where $int\ B^2$ denotes the interior of B^2. We consider a positive number q_0 and define $\widehat{U}_\Gamma = \widehat{U}_{\Gamma,\ell_0}$ to be the set of $P \in \mathscr{P}^3$ such that the conditions $(C1) - (C3)$ below hold and the conditions $(C4)'_\omega - (C10)'_\omega$ in Sect. 2 hold when $\omega \in \Gamma \cap \mathbb{Q}^2$ and ω has small denominator in the sense that $\omega = (p_1/q,\ p_2/q)$, where $(p_1, p_2) \in \mathbb{Z}^2$ and $q \in \mathbb{Z},\ \ 0 < q \leq q_0$.

The number q_0 depends on $\Gamma,\ \ell_0$, and P, continuously on $(\ell_0, P) \in \mathscr{L}^4 \times P^3$, where $\mathscr{L}^4$ is provided with the C^4 topology and $\mathscr{P}^3$ is provided with the C^3 topology.

$(C1)$ For each $\omega \in \Gamma$, each global minimum m_ω of $P_{\omega,\Lambda}$ is non-degenerate, i.e. $P''_{\omega,\Lambda}(m_\omega) > 0$.

$(C2)$ For each $\omega \in \Gamma$, there are at most two global minima of $P_{\omega,\Lambda}$.

We consider $\omega_0 \in \Gamma$ and suppose that $P_{\omega_0,\Lambda}$ has two global minima m_{ω_0} and m'_{ω_0}. We may continue these to local minima m_ω and m'_ω of $P_{\omega,\Lambda}$ for $\omega \in \Gamma$ near ω_0, in view of $(C1)$. Thus, m_ω and m'_ω depend continuously on ω and are the given minima for $\omega = \omega_0$.

$(C3)$ $\ dP_{\omega,\Lambda}(m_\omega)/d\omega\,|_{\omega=\omega_0} \neq dP_{\omega,\Lambda}(m'_\omega)/dw|_{\omega=\omega_0}$.

We call this the *first transversality condition.* We call $(C10)'_\omega$ the *second transversality condition.*

These are the same as the conditions $(C1) - (C3)$ in [4].

We consider $\omega \in \mathbb{Q} \cap \Gamma$. The subset $\mathbb{T}^2_\Lambda/\mathbb{T}^1_\omega$ of $\mathbb{T}^2_\omega := \mathbb{T}^2 \times \mathbb{T}/\mathbb{T}^1_\omega$ is a circle. We choose a generator h_0 of $H_1\left(\mathbb{T}^2_\Lambda/\mathbb{T}^1_\omega;\mathbb{Z}\right)$ and regard it is an element of $H_1\left(\mathbb{T}^2_\omega;\mathbb{R}\right)$ *via* the inclusions $H_1\left(\mathbb{T}^2_\Lambda/\mathbb{T}^1_\omega;\mathbb{Z}\right) \subset H_1(\mathbb{T}^2_\omega;\mathbb{Z}) \subset H_1(\mathbb{T}^2_\omega;\mathbb{R})$. This repeats the definition of h_0 that we gave in [4].

These definitions should suffice for the reader to understand conditions $(C4)'_\omega - (C6)'_\omega$ in Sect. 2 without reading [4]. Next, we explain what we mean when we say that a g_{E_0}-shortest curve is non-degenerate in the sense of Morse.

If μ is a curve in $\mathbb{T}^2_\omega$ and $E \geq E_0$, we let $\ell_E(\mu)$ denote the g_E-length of μ.

We consider a g_E-shortest curve γ in h_0. We consider a transversal μ to γ, intersecting γ in one point, not n_ω in the case that $E = E_0$. For $\theta \in \mu$, we let γ_θ be the g_E-shortest curve through θ.

If $E > E_0$, we have that g_E is a Riemannian metric. It is well known that in this circumstance $\theta \mapsto \ell_E(\gamma_\theta)$ is C^2 in a neighborhood of $\mu \cap \gamma$ and g_E is non-degenerate in the sense Morse if and only if the second derivative of this function on μ is positive at $\mu \cap \gamma$. This is also well known in the case that $E = E_0$ and γ does not pass through n_ω, in view of the fact that g_{E_0} is a Riemannian metric in the complement of n_ω.

In the case that $E = E_0, \gamma$ is simple and passes through n_ω, and $(C6_\omega)'$ holds, it is still true that $\theta \mapsto \ell_E(\theta)$ is C^2 for $\theta \in \mu$ in a neighborhood of $\mu \cap \gamma$. This is a consequence of Lemma 3 in § 7 and Proposition 2 in § 8 of [6], as we will explain in detail in a subsequent paper.

In this case, we say that γ is *non-degenerate in the sense of Morse* if the second derivative of this function is positive.

These definitions should suffice for the reader to understand the remaining conditions $(C7)'_\omega - (C10)'_\omega$ in Sect. 2, without reading [4].

We have thus defined $\widehat{U}_\Gamma = \widehat{U}_{\Gamma,\ell_0} \subset \mathscr{P}^3$. This definition depends, however, on the choice of q_0 and, for each $\omega \in \Gamma$ of the form $\omega = (p_1/q,\ p_2/q)$ with $p_1,\ p_2,\ q \in \mathbb{Z}$ and $0 < q \leq q_0$, numbers $\widehat{E}_0$ and E_0^*. (We can omit E_0^* for those ω for which the g_{E_0}-shortest curve γ in h_0 is simple. We note that E_0 and γ, as well as $\widehat{E}_0$ and E_0^*, depend on $\Gamma,\ \ell_0,\ P$, and ω. The dependence on $(\ell_0, P) \in \mathscr{L}^4 \times \mathscr{P}^3$ is continuous.)

The set U^r_Γ that we discussed in Sect. 2 is related to the set $\widehat{U}_\Gamma$ that we just defined as follows: $U^r_\Gamma = \{\epsilon P : \epsilon > 0 \text{ and } P \in \widehat{U}_\Gamma \cap \mathscr{P}^r\}$, for $r \geq 3$.

We say that a line segment $\Gamma \subset \mathbb{R}^2$ is *rational* if there exists a resonance k such that $\Gamma \subset \Lambda_k$. To define U, we choose a finite number of rational line segments $\Gamma_1, \cdots, \Gamma_n$ in $int\ B^2$ such that $\Gamma_1 \cup \cdots \cup \Gamma_n$ is connected and meets each Ω_i. We set $\widehat{U}_{\ell_0} := \widehat{U}_{\Gamma_1,\ell_0} \cap \cdots \cap \widehat{U}_{\Gamma_n,\ell_0}$. We set $U := \left\{(\ell_0, P) : \ell_0 \in \mathscr{L}^4 \text{ and } P \in \widehat{U}_{\ell_0}\right\}$.

Here is the strengthened version of Theorem 1, mentioned at the beginning of this section: The set U in Theorem 1 may be taken to be the set U defined in this section, provided that for each i, the function q_0 associated to Γ_i is large enough and for each $\omega \in \Gamma_i$ of the form $\omega = (p_1/q,\ p_2/q)$ with $(p_1, p_2) \in \mathbb{Z}^2$ and $q \in \mathbb{Z},\ \ 0 < q \leq q_0$, the function $\widehat{E}_0$ is large enough and (in the relevant cases) the function E_0^* is small enough, subject to the condition $E_0^* > E_0$.

7 Properties of U

In this section, we discuss conditions (b), (c), and (f), stated before Theorem 1.

By definition, U satisfies (b).

For a rational line segment $\Gamma \subset int\ B^2$ and $q_0 > 0$, there are only a finite number of $\omega \in B^2$ that may be expressed in the form $(p_1/q,\ p_2/q)$ with p_1, p_2, $q \in \mathbb{Z}$ and $0 < q < q_0$. It follows that the number of the conditions $(C1) - (C3)$ and $(C4)'_\omega - (C10)'_\omega$ is *locally finite*, in the following sense: For $(\ell_0, P) \in \mathscr{L}^4 \times P^3$, there exists a neighborhood $\mathscr{N}$ of (ℓ_0, P) relative to $\mathscr{L}^4 \times P^3$ and for $i = 1, \cdots, n$, only a finite number of $\omega \in \Gamma_i$ of the form $\omega = (p_1/q,\ p_2/q)$ with $p_1, p_2, q \in \mathbb{Z}$ and $0 < q \leq q_0(\Gamma_i, \ell_0^*, P^*)$ for some $(\ell_0^*, P^*) \in \mathscr{N}$. This is true because q_0 depends continuously on ℓ_0 and P.

Consequently, to prove that U is open, it is enough to prove that the conditions $(C1) - (C3)$ define an open set, that $(C4)'_\omega$ defines an open set, that $(C5)'_\omega - (C7)'_\omega$ define an open set, and that $(C8)'_\omega - (C10)'_\omega$ define an open subset of the set defined by $(C5)'_\omega - (C7)'_\omega$.

That $(C1) - (C3)$ define an open set and that $(C4)'_\omega$ defines an open set are well known.

The condition $(C5)'_\omega$ by itself does not define an open set. Moreover, the formulations of $(C6)'_\omega$ and $(C7)'_\omega$ are meaningful only when $(C5)'_\omega$ holds. For this reason, we formulate more general conditions $(C6)^*_\omega$ and $(C7)^*_\omega$, which reduce to $(C6)'_\omega$ and $(C7)'_\omega$ when $(C5)'_\omega$ holds.

According to Theorem 2 in [6], if $n_\omega \in \gamma$ then, even if there is more than one g_{E_0}-shortest curve γ in h_0, every g_{E_0}-shortest curve γ in h_0 has the form $\gamma = \sum_i p_i \gamma_i$, where p_i is a positive integer, γ_i is a simple closed curve, and $\gamma_i \cap \gamma_j = n_\omega$ for $i \neq j$.

$(C6)^*_\omega$ If γ is a g_{E_0}-shortest curve in h_0 and $n_\omega \in \gamma$, then each γ_i is tangent to the x-axis at n_ω.

$(C7)^*_\omega$ Under the same conditions, each γ_i is non-degenerate in the sense of Morse.

These conditions define an open set. Moreover, the set of $(\ell_0, P) \in \mathscr{L}^4 \times \mathscr{P}^3$ such that $(C5)'_\omega, (C6)^*_\omega$, and $(C7)^*_\omega$ hold is open in the set of (ℓ_0, P) such that $(C6)^*_\omega$ and $(C7)^*_\omega$ hold. Since $(C6)^*_\omega$ is the same as $(C6)'_\omega$ and $(C7)^*_\omega$ is the same as $(C7)'_\omega$ when $(C5)'_\omega$ holds, it follows that the set of (ℓ_0, P) for which $(C5)'_\omega$, $(C6)'_\omega$, and $(C7)'_\omega$ hold is open.

In the case that the g_{E_0}-shortest curve in h_0 is not simple, the fact that $(C8)'_\omega - (C10)'_\omega$ define an open set follows by well-known arguments. In the case that γ is simple, an additional argument is need, viz. conditions $(C4)'_\omega - (C7)'_\omega$ imply that $(C8)'_\omega - (C10)'_\omega$ hold for sufficiently small $E > E_0$. We will prove this in a subsequent paper.

This finishes our sketch of a proof that U is open in $\mathscr{L}^4 \times \mathscr{P}^3$.

In defining the conditions $(C8)'_\omega - (C10)'_\omega$, we treated the cases when the g_{E_0}-shortest curve γ in h_0 is simple and when γ is not simple differently. This seems necessary. On the one hand, when γ is not simple, the conditions $(C4)'_\omega - (C7)'_\omega$

do not imply that $(C8)'_\omega - (C10)'_\omega$ hold for sufficiently small $E > E_0$, and there does not seem to be any reason that the modified conditions $(C8)'_\omega - (C10)'_\omega$, where $\widehat{E}_0 \geq E \geq E_0^*$ is replaced by $\widehat{E}_0 \geq E > E_0$, would define an open set. This is why we restricted E by $\widehat{E}_0 \geq E \geq E_0^*$ in the case γ is not simple. On the other hand, if we were to restrict E in this way when γ is simple, we would not know how to prove our Arnold diffusion result.

Since U is open in $\mathscr{L}^4 \times \mathscr{P}^3$, we have that for any $\ell_0 \in \mathscr{L}^4$ and any $r \geq 3$, $U \cap (\ell_0 \times \mathscr{P}^r)$ is open relative to $\mathscr{P}^r$. Since the intersection of open and dense sets is open and dense, to prove (f) it is enough to prove that each of the conditions $(C1) - (C3)$ and $(C4)'_\omega - (C10)'_\omega$ defines a dense subset of $\mathscr{P}^r$. For each of these conditions, this is well known. This finishes our sketch of the proof of conditions (f) and also of condition (c), since the denseness of $U \cap (\ell_0 \times \mathscr{P}^r)$ in $\mathscr{P}^r$ for $\ell_0 \in \mathscr{L}$ and $r \geq 3$ implies the denseness of U in $\mathscr{L}^4 \times \mathscr{P}^3$.

8 The Legendre–Fenchel Transform

In [2], we associated to a Tonelli Lagrangian $L : TM \times \mathbb{T} \to \mathbb{R}$ a convex function $\beta_L : H_1(M;\mathbb{R}) \to \mathbb{R}$ with superlinear growth. For $h \in H_1(M;\mathbb{R})$, we called $\beta_L(h)$ the *minimal average action* of h. We defined $\beta_L(h)$ to be the minimum of $A(\mu)$ where $A(\mu)$ is the *average action* associated to an invariant Borel probability measure μ on $TM \times \mathbb{T}$ and the minimum is taken over all such μ whose *rotation number* $\rho(\mu)$ is h. Here, M is an arbitrary closed (i.e. compact and boundaryless) manifold. By *Tonelli Lagrangian*, we mean an L as above satisfying the conditions introduced in [2]. We explained the terminology and results that we are using here in [5] § 2 and we will not repeat this explanation here. The results mostly come from [2] and [3], but the terminology is not always the same as in those papers.

The Lagrangian $L = \ell_0 + \epsilon P$ introduced in § 2 is not a Tonelli Lagrangian, since it is defined only on $\mathbb{T}^2 \times B^2 \times \mathbb{T}$ and not all of $T\mathbb{T}^2 \times \mathbb{T} = \mathbb{T}^2 \times \mathbb{R}^2 \times \mathbb{T}$. We may, however, extend it to all of $T\mathbb{T}^2 \times \mathbb{T}$ so as to be a Tonelli Lagrangian, in the manner described in [4] § 3. In the subsequent discussion, we will use the results of [2] and [3] (as recalled in [4] § 2) applied to this extension of L.

Since β_L is convex and has superlinear growth, its *Fenchel conjugate* $\alpha_L : H^1(M;\mathbb{R}) \to \mathbb{R}$ is convex and has superlinear growth. It is defined by $\alpha_L(c) := -min\,\{\beta_L(h) - \langle h, c\rangle : h \in H_1(M;\mathbb{R})\}$, where $\langle h, c\rangle$ is the canonical pairing between $h \in H_1(M;\mathbb{R})$ and $c \in H^1(M;\mathbb{R})$. The *Fenchel inequality* $\beta_L(h) + \alpha_L(c) \geq \langle h, c\rangle$ for $h \in H_1(M;\mathbb{R})$ and $c \in H^1(M;\mathbb{R})$ follows immediately. The *Legendre–Fenchel transform* $\mathscr{L}\mathscr{F}(h) = \mathscr{L}\mathscr{F}_{\beta_L}(h)$ of $h \in H_1(M;\mathbb{R})$ is defined to be $\{c \in H^1(M;\mathbb{R}) : \beta_L(h) + \alpha_L(c) = \langle h, c\rangle\}$. It is a compact, convex, non-empty subset of $H^1(M;\mathbb{R})$. More generally, we define the *Legendre–Fenchel transform* $\mathscr{L}\mathscr{F}(S)$ of a subset S of $H_1(M;\mathbb{R})$ as $\mathscr{L}\mathscr{F}(S) = \bigcup_{h \in S} \mathscr{L}\mathscr{F}(h)$.

In this section, we consider a rational line segment $\Gamma \subset int\ B^2$ and state properties of $\mathscr{L}\mathscr{F}_{\beta_L}(\Gamma)$ that hold when $(\epsilon,\ \ell_0,\ P) \in V$. Here, $L : T\mathbb{T}^2 \times \mathbb{T} \to \mathbb{R}$

is a Tonelli Lagrangian that extends $\ell_0 + \epsilon P$ in the fashion described in [4] § 3. Proving these properties are a step in our proof of Theorem 1, but we will not prove them in this paper.

We let (k_0, k_1, k_2) generate the group of resonances of Γ. For $c \in \mathbb{R}^2$, the function $\dot\theta \mapsto \ell_0(\dot\theta) - \langle \dot\theta, c\rangle : \Gamma \to \mathbb{R}$ has everywhere positive second derivative, so it has a unique minimum $\omega_c \in \Gamma$. If ω_c is in the relative interior $int\ \Gamma$ of Γ, then $c - d\ell_0(\omega_c)$ is in the one dimensional subspace of $\mathbb{R}^2$ spanned by (k_1, k_2). In this case, the line parallel to this one dimensional subspace and passing through c meets $d\ell_0(\Gamma)$ in exactly one point $d\ell_0(\omega_c)$, and crosses $d\ell_0(\Gamma)$ transversely there. We note that since B^2 is convex, and $d^2\ell_0(\dot\theta)$ is positive definite for every $\dot\theta \in B^2$, it follows that the mapping $\dot\theta \mapsto d\ell_0(\dot\theta) : B^2 \to \mathbb{R}^2$ is injective and its derivative at any point in B^2 is an isomorphism.

The diagram

$$\mathbb{T}^2 \times 0 \subset \mathbb{T}^2 \times \mathbb{T} \overset{pr}{\to} \mathbb{T}^1_\Lambda,$$

where pr denotes the projection of $\mathbb{T}^2 \times \mathbb{T}$ on $\mathbb{T}^1_\Lambda := \left(\mathbb{T}^2 \times \mathbb{T}\right)/\mathbb{T}^2_\Lambda$, induces the linear mapping

$$pr^* : H^1\left(\mathbb{T}^1_\Lambda; \mathbb{R}\right) \to H^1(\mathbb{T}^2; \mathbb{R}) = \mathbb{R}^2$$

whose image is the one dimensional subspace of $\mathbb{R}^2$ spanned by (k_1, k_2).

We let

$$\mathscr{L}\mathscr{F}_{\omega,\Lambda} : H_1\left(\mathbb{T}^1_\Lambda; \mathbb{R}\right) \to \left\{\text{compact, convex, non-empty subsets of } H^1\left(\mathbb{T}^1_\Lambda; \mathbb{R}\right)\right\}$$

denote the Legendre–Fenchel transform associated to β_L in the case that $L = L_{\omega,\Lambda}$. The assumption that $P_{\omega,\Lambda}$ is not constant implies that $\mathscr{L}\mathscr{F}_{\omega,\Lambda}(0)$ is not reduced to a point. It is easily seen that β_L is an even function in the case that $L = L_{\omega,\Lambda}$. Hence, $\mathscr{L}\mathscr{F}_{\omega,\Lambda}(0)$ is symmetric about 0, i.e. it is a closed interval $[-a_\omega, a_\omega]$ with $a_\omega = a_{\omega,\Lambda} > 0$ where we identify $H^1(\mathbb{T}^1_\Lambda; \mathbb{R})$ with $\mathbb{R}$ by identifying a generator of $H^1\left(\mathbb{T}^1_\Lambda; \mathbb{Z}\right)$ with 1.

In view of $(C3)$, there are at most finitely many $\omega \in \Gamma$ where $P_{\omega,\Lambda}$ has more than one global minimum, and by $(C2)$, it has two global minima at any such point. The function $\omega \mapsto a_\omega : \Gamma \to \mathbb{R}_{++}$ is continuous except at those ω at which $P_{\omega,\Lambda}$ has two global minima. At $\omega_0 \in \Gamma$ where $P_{\omega_0,\Lambda}$ has two global minima, the one-sided limits $a_{\omega_0-} := \lim\limits_{\omega\uparrow\omega_0} a_\omega$ and $a_{\omega_0+} := \lim\limits_{\omega\downarrow\omega_0} a_\omega$ exist and

$$a_{\omega_0} = \max\left(a_{\omega_0-},\ a_{\omega_0+}\right).$$

We let

$$\mathscr{L}\mathscr{F} : H_1(\mathbb{T}^2; \mathbb{R}) \to \left\{\text{compact, convex, non-empty subsets of } H^1\left(\mathbb{T}^2; \mathbb{R}\right)\right\}$$

denote the Legendre–Fenchel transform associated to β_L in the case that L is a Tonelli extension of $\ell_0 + \epsilon P$ in the fashion of [4] § 3. In the case that $\omega \in \Gamma \subset$

int $B^2 \subset \mathbb{R}^2 = H_1(\mathbb{T}^2;\mathbb{R})$ and ϵ is sufficiently small, we have that $\mathscr{L}\mathscr{F}(\omega) \subset H^1(\mathbb{T}^2;\mathbb{R}) = \mathbb{R}^2$ is independent of the choice of extension L of $\ell_0 + \epsilon P$.

We describe the sets $\mathscr{L}\mathscr{F}(\omega)$ for $\omega \in \Gamma$ by means of three functions $b^{\pm} : \Gamma \to \mathbb{R}$ and $z : \Gamma \to \Lambda$. We choose an orientation for Λ and order Λ accordingly. The functions $b^{\pm}$ and z have the following properties:

- z is monotone increasing;
- $b^- \leq b^+$;
- The one-sided limits $b^{\pm}(\omega_0-) := \lim_{\omega\uparrow\omega_0} b^{\pm}(\omega)$ and $b^{\pm}(\omega_0+) := \lim_{\omega\downarrow\omega_0} b^{\pm}(\omega)$ exist for all $\omega_0 \in \Gamma$, with the exception of the endpoints, where only one one-sided limit exists (from below for the top endpoint and from above for the bottom endpoint);
- $b^+(\omega) = \max(b^+(\omega-), b^+(\omega+))$ and $b^-(\omega) = \min(b^-(\omega-), b^-(\omega+))$; and
- $\mathscr{L}\mathscr{F}(\omega) = \{d\ell_0(\omega^*) + \sqrt{\epsilon}\; pr^*\widehat{c} : z(\omega-) \leq \omega^* \leq z(\omega+) \text{ and } b^-(\omega) \leq \widehat{c} \leq b^+(\omega)\}$.

In the discussion above, we have held ϵ, ℓ_0, and P fixed. Next, we discuss convergence properties of $b^{\pm}$ and z as ϵ goes to zero. For z, we have $z(\omega\pm) \to \omega$ as $\epsilon \downarrow 0$. For $b^{\pm}$, we introduce the quantity $B(\omega) := \limsup_{\epsilon\downarrow 0}\{\min[\max(\,|b^-(\omega) + a_{\omega-}|,\ |b^+(\omega) - a_{\omega-}|\,),\ \max(|b^-(\omega) + a_{\omega+}|\,,\ |b^+(\omega) - a_{\omega+}|)]\}$. We will show in a subsequent paper that $B(\omega) = 0$ if ω is irrational, and that $\omega \mapsto B(\omega)$ is continuous at all irrational ω in Γ.

This says that the interval $[b^-(\omega),\ b^+(\omega)]$ is approximated by one of the intervals $[-a_{\omega-}, a_{\omega-}]$ or $[-a_{\omega+}, a_{\omega+}]$ under the conditions that ϵ is small and ω is irrational or rational with large denominator. This result is incomplete, however, since it is not uniform in ϵ and it says nothing about rational ω with small denominator.

For our proof of Arnold diffusion, what we would like to show is that there exists a connected component of the interior of $\mathscr{L}\mathscr{F}(\Gamma)$ that intersects both $\mathscr{L}\mathscr{F}(\omega_0)$ and $\mathscr{L}\mathscr{F}(\omega_1)$ where ω_0 and ω_1 are the endpoints of Γ. We can show this under the assumptions $(C1) - (C3)$ and $(C4)_\omega - (C10)_\omega$ for rational $\omega \in \Gamma$ with small denominator and the additional assumption that for rational $\omega \in \Gamma$ with small denominator the unique g_{E_0}-shortest curve in h_0 is simple.

We cannot show this, however, without the additional assumption. Fortunately, for our proof of Arnold diffusion there is a way to overcome this difficulty, which we will discuss in a subsequent paper. Nevertheless, it is useful to first explain some results used in showing that this can be proved under the additional assumption.

To show that there is a connected component of the interior of $\mathscr{L}\mathscr{F}(\Gamma)$ that meets both $\mathscr{L}\mathscr{F}(\omega_0)$ and $\mathscr{L}\mathscr{F}(\omega_1)$, it is enough to obtain a uniform (in ϵ) lower bound for $b_+ - b_-$. This can be done under the additional assumption. Next, we state some more results relevant to obtaining such a lower bound, to explain some of the ideas in our proof of Arnold diffusion.

We can show that $[b^-(\omega),\ b^+(\omega)]$ can be approximated uniformly in ϵ by one of the intervals $[-a_{\widehat{\omega}},\ a_{\widehat{\omega}}]$ with $|\omega - \widehat{\omega}| \leq \delta\sqrt{\epsilon}$, provided that there is no ω_0 with small denominator such that $|\omega - \omega_0| \leq C_1\sqrt{\epsilon}$. Here, $\delta > 0$ is arbitrary and C_1 is a suitably large number. More precisely, if $\delta > 0$ is given, then there exist $C_1, q_0, \epsilon_0 > 0$, which depend only on ℓ_0, P, and δ, such that if $0 < \epsilon \leq \epsilon_0$ and $\omega \in \Gamma$ and there is no $\omega_0 = (p_1/q,\ p_2/q) \in \Gamma$ with $p_1,\ p_2,\ q \in \mathbb{Z},\quad 0 < q \leq q_0$, and $|\omega - \omega_0| \leq C_1\sqrt{\epsilon}$, then there exists $\widehat{\omega} \in \Gamma$ with $|\omega - \widehat{\omega}| \leq \delta\sqrt{\epsilon},\ \ |b^-(\omega) + a_{\widehat{\omega}}| < \delta$, and $|b^+(\omega) - a_{\widehat{\omega}}| < \delta$.

Since q_0 is independent of ϵ, the conditions $|\omega - \omega_0| \leq C_1\sqrt{\epsilon}$ with ω_0 as above exclude a finite number of intervals each of length $2C_1\sqrt{\epsilon}$, and the number of intervals excluded is independent of ϵ. Under the conditions $(C1) - (C3)$, the function $\omega \mapsto a_\omega : \Gamma \to \mathbb{R}_{++}$ has a positive lower bound B that depends only on ℓ_0 and P. We may take $\delta = B/2$, so q_0 depends only on ℓ_0 and P. Then $b_+ - b_-$ has the lower bound B on Γ outside of the intervals of length $2C_1\sqrt{\epsilon}$ centered at points of the form $(p_1/q,\ p_2/q) \in \Gamma$ with $p_1,\ p_2,\ q \in \mathbb{Z}$ and $0 < q \leq q_0$.

We consider $\omega_0 \in int\ \Gamma \cap \mathbb{Q}$. We let

$$\mathscr{L}\mathscr{F}_{\omega_0} : H_1\left(\mathbb{T}^2_{\omega_0}; \mathbb{R}\right) \to \{\text{compact, convex, non-empty subsets of } H^1\left(\mathbb{T}^2_{\omega_0}; \mathbb{R}\right)\}$$

denote the Legendre–Fenchel transform associated to β_L in the case that $L = L_{\omega_0}$. We let h_0 be as in § 5, i.e. a generator of $H_1\left(\mathbb{T}^2_\Lambda/\mathbb{T}^1_{\omega_0}; \mathbb{Z}\right)$, regarded as an element of $H_1\left(\mathbb{T}^2_{\omega_0}; \mathbb{R}\right)$. The diagram

$$\mathbb{T}^2 \times 0 \subset \mathbb{T}^2 \times \mathbb{T} \xrightarrow{pr} \mathbb{T}^2_\omega,$$

where pr denotes the projection of $\mathbb{T}^2 \times \mathbb{T}$ on $\mathbb{T}^2_\omega := \left(\mathbb{T}^2 \times \mathbb{T}\right)/\mathbb{T}^1_\omega$, induces an isomorphism

$$pr_* : \mathbb{R}^2 = H_1(\mathbb{T}^2; \mathbb{R}) \to H_1\left(\mathbb{T}^2_\omega; \mathbb{R}\right).$$

For $\lambda \in \mathbb{R}$, we set $\omega_\lambda := \omega_0 + \sqrt{\epsilon}\lambda pr_*^{-1}(h_0) \in \Lambda$.

We choose $h_1 \in H_1\left(\mathbb{T}^2_{\omega_0}; \mathbb{Z}\right)$ such that h_0 and h_1 generate this abelian group. In particular, (h_0, h_1) is a basis of the vector space $H_1\left(\mathbb{T}^2_{\omega_0}; \mathbb{R}\right)$. We let (c_0, c_1) be a dual basis of $H^1\left(\mathbb{T}^2_{\omega_0}; \mathbb{R}\right)$, i.e., we suppose that $\langle h_i, c_j\rangle = \delta_{ij}$ (Kronecker delta symbol) for $i, j = 0, 1$.

We set $\beta_{\omega_0} := \beta_L$ where $L = L_{\omega_0}$ and set $\widehat{\beta}_{\omega_0} := \beta_{\omega_0}|\mathbb{R} \cdot h_0$. Then $\widehat{\beta}_{\omega_0} : \mathbb{R} \to \mathbb{R}$ is an even, convex function with superlinear growth. It is differentiable except at the origin, where it is not differentiable under the assumption that condition $(C4)'_{\omega_0}$ holds. There exist functions $c^\pm : \mathbb{R} \to \mathbb{R}$ such that the following hold:

- If $\lambda \neq 0$ then $\mathscr{L}\mathscr{F}_{\omega_0}(\lambda h_0) = \left\{\widehat{\beta}'_{\omega_0}(\lambda)c_0 + \zeta c_1 : c^-\left(\widehat{\beta}'_{\omega_0}(\lambda)\right) \leq \zeta \leq c^+\left(\widehat{\beta}'_{\omega_0}(\lambda)\right)\right\}$;
- $\mathscr{L}\mathscr{F}_{\omega_0}(0) = \left\{\tau c_0 + \zeta c_1 : \widehat{\beta}'_{\omega_0}(0-) \leq \tau \leq \widehat{\beta}'_{\omega_0}(0+) \text{ and } c^-(\tau) \leq \zeta \leq c^+(\tau)\right\}$; and
- For $\tau \in \mathbb{R}$, $\quad c^\pm(-\tau) = -c^\mp(\tau)$.

There exists an ergodic action minimizing probability measure μ for L_{ω_0} with rotation vector $\rho(\mu) = \lambda h_0$ in the case when λ is an extremal point of the epigraph of $\widehat{\beta}_{\omega_0}$. If furthermore $\lambda \neq 0$, then μ is evenly supported on an L_{ω_0}-invariant simple closed curve γ_λ on $\mathbb{T}^2_{\omega_0}$. Since the Lagrangian L_{ω_0} is autonomous, the corresponding Hamiltonian H_{ω_0} is invariant, so the curve is contained in an energy level $\{H_{\omega_0} = E\}$. According to a result of Dias Carneiro [1], $E = \lambda\widehat{\beta}'_{\omega_0}(\lambda) - \beta_{\omega_0}(\lambda) \geq E_0$ and the curve γ_λ (apart from parameterization) is a g_E-shortest geodesic on $\mathbb{T}^2_{\omega_0}$.

The assumptions $(C8)'_{\omega_0} - (C10)'_{\omega_0}$ imply that for E in the range

- $E_0 < E \leq \widehat{E}_0$ if the g_{E_0}-shortest curve in h_0 is simple, or
- $E_0^* \leq E \leq \widehat{E}_0$ if the g_{E_0}-shortest curve in h_0 is not simple,

we have the following

- If there is only one g_E-shortest curve in h_0 then there is a unique $\lambda > 0$ such that $\lambda\widehat{\beta}'_{\omega_0}(\lambda) - \widehat{\beta}_{\omega_0}(\lambda) = E$ and λh_0 is an extremal point of β_{ω_0}, and
- If there are two g_E-shortest curves in h_0 then the set of $\lambda > 0$ such that $\lambda\widehat{\beta}'_{\omega_0}(\lambda) - \widehat{\beta}_{\omega_0}(\lambda) = E$ is a closed interval $[\lambda_0, \lambda_1]$ and $\lambda_0 h_0$ and $\lambda_1 h_0$ are extremal points of β_{ω_0}.

In the first case, there is a unique action minimizing probability measure for L_{ω_0} with rotation vector λh_0. In the second case, this is true for $\lambda = \lambda_0$ and $\lambda = \lambda_1$, but there does not exist an action minimizing probability measure for L_{ω_0} with rotation vector λh_0 when $\lambda_0 < \lambda < \lambda_1$. In the first case, the measure is supported in the g_E-shortest curve in h_0. In the second case, each of the measures is supported in one of the two g_E-shortest curves.

In the case that the g_{E_0}-shortest curve in h_0 is simple, the functions c^- and c^+ are continuous on the set J of $\widehat{\beta}'_{\omega_0}(\lambda)$ such that $\lambda\widehat{\beta}'_{\omega_0}(\lambda) - \widehat{\beta}'_{\omega_0}(\lambda) \leq \widehat{E}_0$, with the exception of the finite set of discontinuities. These occur only at $\widehat{\beta}'_{\omega_0}(\lambda)$ for which there are two g_E-shortest curve in h_0 with $E = \lambda\widehat{\beta}'_{\omega_0}(\lambda) - \widehat{\beta}_{\omega_0}(\lambda)$, and $\widehat{\beta}'_{\omega_0}(0\pm)$. In the other case, the same statement is true for K in place of J, where K is the set of $\widehat{\beta}'_{\omega_0}(\lambda)$ such that $E_0^* \leq \lambda\widehat{\beta}'_{\omega_0}(\lambda) - \widehat{\beta}_{\omega_0}(\lambda) \leq \widehat{E}_0$.

We note that J is a symmetric closed interval about the origin and K is the union of a closed interval in $\mathbb{R}_{++}$ and its reflection about the origin. Moreover, $c_+ - c_-$ has a lower bound $B \in \mathbb{R}_{++}$ on J in the first case and on K in the second case.

If λ is in J or K (according to the case) we can show that $\left[b^-(\omega_\lambda),\ b^+(\omega_\lambda)\right]$ can be approximated by $\left[c^-\left(\widehat{\beta}'_{\omega_0}(\lambda')\right),\ c^+\left(\widehat{\beta}'_{\omega_0}(\lambda')\right)\right]$, with λ' close to λ. More precisely, if $\delta > 0$ is given, then there exists $\epsilon_0 > 0$, which depends only on ℓ_0, P, and δ such that if $0 < \epsilon \leq \epsilon_0$ and $\lambda \in J$ or $\lambda \in K$ (depending on the case) then there exists λ' with $|\lambda - \lambda'| < \delta$ such that $\left|b^-(\omega_\lambda) - c^-\left(\widehat{\beta}'_{\omega_0}(\lambda')\right)\right| < \delta$ and $\left|b^+(\omega_\lambda) - c^+\left(\widehat{\beta}'_{\omega_0}(\lambda')\right)\right| < \delta$.

This finishes our sketch of why there exists a connected component of the interior of $\mathscr{L}\mathscr{F}(\Gamma)$ that intersects both $\mathscr{L}\mathscr{F}(\omega_0)$ and $\mathscr{L}\mathscr{F}(\omega_1)$ in the case that for any rational ω_0 in Γ with small denominator the g_{E_0}-shortest curve in h_0 is simple.

References

1. M.J. Dias Carneiro, On minimizing measures of the actions of autonomous Lagrangians. Nonlinearity **8**, 1077–1085 (1995)
2. J.N. Mather, Action minimizing invariant measures for positive definite Lagrangian systems. Math. Z. **207**(2), 169–207 (1991)
3. J.N. Mather, Variational construction of connecting orbits. Ann. Inst. Fourier (Grenoble) **43**(5), 1349–1386 (1993)
4. J.N. Mather, Arnold diffusion I. Announcement of results. J. Math. Sci. NY. **124**(5), 5275–5289 (2004); Russian translation in Sovrem. Mat. Fundam. Napravi **2**, 110–130 (2003) (electronic)
5. J.N. Mather, Order structure on action minimizing orbits, in *Symplectic Topology and Measure Preserving Dynamical Systems*, ed. by A. Fathi, Y.-G. Oh, C. Viterbo. Contemporary Mathematics, vol. 512 (American mathematical society, Providence, 2010), pp. 41–125
6. J.N. Mather, Shortest curves associated to a degenerate Jacobi metric on $\mathbb{T}^2$. in *Progress in Variational Methods*, ed. by C. Liu, Y. Long. Nankai Series in Pure, Applied Mathematics and Theoretical Physics, vol. 7, pp. 126–168 (World Scientific, New Jersey, 2011)

Turning Washington's Heuristics in Favor of Vandiver's Conjecture

Preda Mihăilescu

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract A famous conjecture bearing the name of Vandiver states that $p \nmid h_p^+$ in the p – cyclotomic extension of $\mathbb{Q}$. Heuristics arguments of Washington, which have been briefly exposed in Lang (Cyclotomic fields I and II, Springer, New York, 1978/1980, p 261) and Washington (Introduction to cyclotomic fields, Springer, New York/London, 1996, p 158) suggest that the Vandiver conjecture should be false if certain conditions of statistical independence are fulfilled. In this note, we assume that Greenberg's conjecture is true for the $p-$th cyclotomic extensions and prove an elementary consequence of the assumption that Vandiver's conjecture fails for a certain value of p: the result indicates that there are deep correlations between this fact and the defect $\lambda^- > i(p)$, where $i(p)$ is like usual the irregularity index of p, i.e. the number of Bernoulli numbers $B_{2k} \equiv 0 \bmod p, 1 < k < (p-1)/2$. As a consequence, this result could turn Washington's heuristic arguments, *in a certain sense* into an argument in favor of Vandiver's conjecture.

1 Introduction

Let p be an odd prime and $\mathbb{K} = \mathbb{Q}[\zeta]$ be the $p-$th cyclotomic field and $G = \text{Gal}(\mathbb{K}/\mathbb{Q})$. If X is a finite abelian group, we denote by X_p its p – Sylow group; let $A = \text{id}C(\mathbb{K})_p$, the p – Sylow subgroup of the class group $\text{id}C(\mathbb{K})$ and h^+, h^- the

P. Mihăilescu (✉)
Mathematisches Institut der Universität Göttingen, Göttingen, Germany
e-mail: preda@uni-math.gwdg.de

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_12, © Springer-Verlag Berlin Heidelberg 2012

sizes[1] of A^+ respectively A^-. Kummer explained to Kronecker in a letter from 1849 his ideas for approaching Fermat's Last Theorem ([4], Chap. IX, § 8) on base of two assumptions. One was still unsolved in 1853 when he referred to it in a new letter to Kronecker as a theorem "still to be established" (*noch zu beweisender Satz*): this is what we presently call the Vandiver conjecture. The second assumption, which implies that units of $\mathbb{K}$ which are local p^2-th powers must be global $p-$th powers, is neither proved nor disproved, but it is not known as a conjecture with a particular name.

In [3, p. 261], Washington gives a heuristic argument which suggests that there might be an asymptotic amount of $O(\log\log(N))$ of primes $p \leq N$ for which $\lambda(A^-_\infty) = i(p) + 1$, where $i(p)$ is the irregularity index of p, i.e. the number of Bernoulli numbers $B_{2k}, 1 < k < (p-1)/2$ that vanish modulo p. In [5, p. 158], Washington starts with a naive argument, on base of which the cyclotomic unit $\eta_{2k} := e_{2k}(1-\zeta)^{\sigma-1}$ (see below for the definition of the idempotents $e_j \in \mathbb{F}_p[G]$) may be a $p-$th power with probability $1/p$: this yields a probability of almost one half, for the failure of Vandiver's conjecture, so the argument is obviously too crude. Washington assumes then that the conditional probability that a cyclotomic unit η_{2k} is a p-power, given that $B_{2k} \equiv 0 \bmod p$ is $1/p$: this heuristic leads to a frequence of $O(\log\log(N))$ primes $p < N$ for which the conjecture fails. As a consequence, various specialists in the field expect that the conjecture should not always hold. Our result in this note shows that if Vandiver's conjecture fails, then one has the additional condition $\lambda^- > i(p)$. If one considers this condition also as statistically independent (!) from the two conditions in Washington's heuristics, then the same argument suggests that there may be $O(1)$ primes $p < N$ for which Vandiver's conjecture fails, thus possibly none. As a consequence of this result, the failing of Vandiver appears to require for a sequence of conditions that might have been considered previously as statistically independent. The foundation of statistical heuristics becomes herewith more uncertain. It is thus preferable to state that the conditional probability in Washington's heuristics should be sensibly smaller than $1/p$.

No direct consequence can be drawn as to the truth of the Kummer – Vandiver conjecture; however we have an explicit theorem which indicates an unknown dependence, and also a method of investigation which may be extended for the purpose of investigating more possible consequences of the assumption that the Kummer-Vandiver conjecture is false. The central idea of our proof can be extended, with additional detail, to the general case, and this shall be done in a subsequent paper.

The result of this paper is the following:

Theorem 1. *Let p be an odd prime with irregularity index $i(p) = 1$, for which Greenberg's conjecture holds. If $p \mid h_p^+$, then $\lambda^- \geq 2$.*

[1] One may encounter also the notation $h^+ = |\mathrm{id}C(\mathbb{K}^+)|$, but we refer strictly to the p-part in this paper.

Since $\lambda^- > 1$ is an implication of $h_p^+ > 1$, the two events cannot be considered as independent events, each one with probability $1/p$. But this implication can also suggest that the probability that η_{2k} is a $p-$th power has rather the probability $1/p^2$ than $1/p$ since it implies the vanishing of a higher order Bernoulli number.

Note that we restrict our analysis, for simplicity, to the case of irregularity index 1. However, this is the critical case in Washington's heuristics, and if the assumption of "statistical independence" is close to reality,[2] then the probability of failure of the conjecture for higher values of $i(p)$ can only be smaller, so the argument stays valid.

2 Proof of the Theorem

We let $\mathbb{K} = \mathbb{Q}[\zeta]$ be the $p-$th cyclotomic extension and $\mathbb{K}_n = \mathbb{K}[\zeta^{1/p^n}], n \geq 1$, the p^n-th extension. The galois groups are

$$\begin{aligned} G = & \text{ Gal } (\mathbb{K}/\mathbb{Q}) = \{\sigma_a \ : \ a = 1, 2, \ldots, p-1, \ \zeta \mapsto \zeta^a\} \cong (\mathbb{Z}/p \cdot \mathbb{Z})^*, \\ G_n = & \text{ Gal } (\mathbb{K}_n/\mathbb{Q}) = G \times \langle \tau \rangle, \quad \tau(\zeta_{p^n}) = \zeta_{p^n}^{1+p}, \end{aligned}$$

so τ generates Gal $(\mathbb{K}_n/\mathbb{K})$, in particular. If $g \in \mathbb{F}_p$ is a generator of $(\mathbb{Z}/p \cdot \mathbb{Z})^*$, then $\sigma = \sigma_g$ generates G multiplicatively. We write $\jmath \in G$ for complex multiplication. For $\sigma \in G$ and $R \in \{\mathbb{F}_p, \mathbb{Z}_p, \mathbb{Z}/(p^m \cdot \mathbb{Z})\}$ we let $\varpi(\sigma) \in R$ be the value of the Teichmüller character on σ; for $R = \mathbb{F}_p$ we may also write $\hat{\sigma}$ for this values. The orthogonal idempotents $e_k \in R[G]$ are

$$e_k = \frac{1}{p-1} \sum_{a=1}^{p-1} \varpi^k(\sigma_a) \cdot \sigma_a^{-1}.$$

If X is a finite abelian p – group on which G acts, then $e_k(\mathbb{Z}_p)$ acts via its approximants to the p^m-th order; we shall not introduce additional notations for these approximants. A fortiori, complex conjugation acts on X splitting it in the canonical plus and minus parts: $X = X^+ \oplus X^-$, with $X^+ = X^{1+\jmath}, X^- = X^{1-\jmath}$. The units of $\mathbb{K}$ and $\mathbb{K}_n$ are denoted by E, E_n and the cyclotomic units by C, C_n. If $\mathbb{F}$ is an arbitrary field, we let $E'(\mathbb{F})$ denote the p-units of the number field $\mathbb{F}$, i.e. the units of the smallest ring containing $E(\mathbb{F})$ and in which all the primes above p are invertible. In particular $E_n' = \lambda_n^{\mathbb{Z}} \cdot E_n$, with $\lambda_n = 1 - \zeta_{p^n}$; it is customary to denote by $A'(\mathbb{F})$ the p-part of the ideal class group of the p-integers of $\mathbb{F}$. If $\mathbb{L}/\mathbb{K}$ is an unramified abelian p-extension in which all the primes above p in $\mathbb{K}$ are totally split, then it is known (see e.g. [2], § 4) that

$$\text{Ker } (\iota : A(\mathbb{K}) \to A(\mathbb{L})) = \text{ Ker } (\iota' : A'(\mathbb{K}) \to A'(\mathbb{L})). \tag{1}$$

[2] Washington mentions explicitly that this is the critical point in the various heuristics of this kind.

The Iwasawa invariants $\lambda(\mathbb{K}), \lambda^-(\mathbb{K})$ are related to the cyclotomic $\mathbb{Z}_p$-extension $\mathbb{K}_\infty = \cup_n \mathbb{K}_n$ and $A_n = (\mathrm{id}C(\mathbb{K}_n))_p$ are the p-parts of the ideal class groups of $\mathbb{K}_n$. They form a projective sequence with respect to the relative norms $N_{m,n} = \mathrm{Norm}_{\mathbb{K}_m/\mathbb{K}_n}, m > n \geq 1$ and $\mathbf{A} = \varprojlim_n A_n$. Likewise, $\mathbf{A}' = \varprojlim_n A'_n$ and we also write A, A' for A_1, A'_1. We shall write in general $A(\mathbb{L}) = (\mathrm{id}C(\mathbb{L}))_p$ for the p-part of the class group of an arbitrary number field $\mathbb{L}$, so $A = A(\mathbb{K})$, etc. We also write $\lambda_n = 1 - \zeta_{p^n}$, the primes above p which should not be confounded with the Iwasawa invariants.

We fix now an odd prime p such that

1. Greenberg's conjecture holds for p, so A^+ is finite and $\lambda^+ = 0$. In particular, there is an $n_0 \geq n$, such that for all $n \geq n_0$ we have $|A_n^+| = |A_{n+1}^+|$.
2. Vandiver's conjecture fails for p.
3. There is a unique irregular index $2k$ such that $A_{p-2k} = e_{p-2k}A \neq \{1\}$. Additionally $A_{2k} \neq \{1\}$, as a consequence of 2.

Under these premises, we show that $\mathbb{Z}_p$-rk$(e_{p-2k}\mathbf{A}) > 1$, which is the statement of the theorem. We prove the statement by contraposition, so we assume that $\mathbb{Z}_p$-rk$(e_{p-2k}\mathbf{A}) = 1$. Since there is a unique irregular index, the minimal polynomial of $\mathbf{A}$ is linear. Let $\mathbb{H}_n/\mathbb{K}_n$ be the maximal p-abelian unramified extensions. They split in plus and minus parts according to $A_n = A_n^+ \oplus A_n^-$ and our assumption implies that $\mathbb{H}_n^+/\mathbb{K}_n$ are cyclic extensions of degree

$$d_n := [\mathbb{H}_n^+ : \mathbb{K}_n] = |A_n^+|.$$

We may also consider $\mathbb{H}_n^+$ as the compositum of $\mathbb{K}_n$ with the full p-part of the Hilbert class field of $\mathbb{K}_n^+ \subset \mathbb{K}_n$, the maximal real subextension of $\mathbb{K}_n$: thus $\mathbb{H}_n^+$ is a canonical subfield, with galois group corresponding by the Artin map to A_n^+. It follows that $\mathbb{H}_n^+/\mathbb{K}_n^+$ is an abelian extension, and thus $\mathbb{H}_n^+$ is a CM field (see also [5], Lemma 9.2 for a detailed proof).

There is a canonic construction of radicals from A_n^-, such that $\mathbb{H}_n^- \cdot \mathbb{K}_m \subset \mathbb{K}_m[(A_m^-)^{1/p^m}]$ for sufficiently large m. For such $n > n_0$ like in point 1., let $a_n \in A_n^-$ generate this cyclic group. Let $\mathfrak{Q} \in a_n$ and $\alpha_0 \in \mathbb{K}_n^\times$ with $(\alpha_0) = \mathfrak{Q}^{\mathrm{ord}\,(a_n)}$; there is an $\alpha = \eta \cdot \alpha_0^{1-\jmath}, \eta \in \mu_{p^n}$ which is well defined up to roots of unity, such that $\mathbb{H}_n^+ \subset \mathbb{K}_n[\alpha^{1/p^n}]$. The radical B_n of $\mathbb{H}_n^+$ is then the multiplicative group generated by α and $(\mathbb{K}_n^\times)^{d_n}$.

Since $\mathbb{H}_n^+/\mathbb{K}_n$ is cyclic, a folklore result, which we prove for completeness in Lemma 2 of the Appendix below, implies that

$$(A(\mathbb{H}_n^+))^+ = ((\mathrm{id}C(\mathbb{H}_n^+))_p)^+ = \{1\}. \tag{2}$$

For an arbitrary field $\mathbf{K}$ we denote by $P_\mathbf{K}$ the principal ideals of $\mathbf{K}$. If $\mathbf{L}/\mathbf{K}$ is a galois extension with group G, we denote by $\mathrm{id}A(\mathbf{L}/\mathbf{K}) = P_\mathbf{L}^G / P_\mathbf{K}$ the quotient of the *principal ambig* ideals $P_\mathbf{L}^G \subset P_\mathbf{L}$ by the lift of the principal ideals of $\mathbf{K}$. The first can be either ideals from $\mathbf{K}$ that capitulate in $\mathbf{L}$ or (powers of) ramified primes. A classical result, proved by Iwasawa [2] in a general cohomological language, states that for an arbitrary galois extension $\mathbf{L}/\mathbf{K}$ of finite number fields, there is a canonical isomorphism

$$H^1(\text{ Gal }(\mathbf{L}/\mathbf{K}), E(\mathbf{L})) \cong \mathrm{id}A(\mathbf{L}/\mathbf{K}). \tag{3}$$

As a consequence, if $\mathbf{L}/\mathbf{K}$ is a cyclic unramified p-extension, the factor $\mathrm{id}A$ is a capitulation kernel and we obtain:

$$H^1(\text{ Gal }(\mathbf{L}/\mathbf{K}), E(\mathbf{L})) \cong \text{ Ker }(\iota : A(\mathbf{K}) \to A(\mathbf{L})). \tag{4}$$

This applies in particular to the extensions $\mathbb{H}_n^+/\mathbb{K}_n$.

We shall in the sequel consider the homology groups H^0, H^1 for the unit groups. We are only interested in the real units, so we tacitly assume that $E(\mathbb{L}) = E^+(\mathbb{L})$ for CM extensions $\mathbb{L}/\mathbb{K}$. We may then write, for simplicity $H^i(\mathbb{L}/\mathbb{K}) := H^i(\text{ Gal }(\mathbb{L}/\mathbb{K}), E(\mathbb{L}))$ for $i = 1, 2$. The isomorphism in (3) restricts also to one of p-parts of the respective groups; furthermore, complex conjugation also induces canonical isomorphisms of the plus and minus parts of H^i. The extensions $\mathbb{H}_n^+/\mathbb{K}_n$ being cyclic of degree d_n, the Herbrand quotient is d_n and thus

$$|H^1(\mathbb{H}_n^+/\mathbb{K}_n)| = d_n \cdot |H^0(\mathbb{H}_n^+/\mathbb{K}_n)|.$$

We claim that $\left(H^0(\mathbb{H}_n^+/\mathbb{K}_n)\right)^+ = \{1\}$. Indeed, since $d_n = |A_n^+|$ by definition, we have, in view of (4), exactly $|\mathrm{id}A(\mathbb{H}_n^+/\mathbb{K}_n)| = d_n^2$. This follows from the fact that the plus part capitulates completely, while the minus part is cyclic too and generates the radical of the extension, thus capitulating too. Since $\mathfrak{Q}^{(1-\jmath)\text{ord }(a_n)/d_n} = (\alpha_n^{1/d_n})$ is principal in $\mathbb{H}_n^+$, the claimed size for the group of ambig ideals follows, and thus

$$|\mathrm{id}A(\mathbb{H}_n^+/\mathbb{K}_n)| = d_n^2 \quad \text{ and } |\mathrm{id}A(\mathbb{H}_n^+/\mathbb{K}_n)^-| = |\mathrm{id}A(\mathbb{H}_n^+/\mathbb{K}_n)^+| = d_n.$$

Therefore, $|H^0(\mathbb{H}_n^+/\mathbb{K}_n)| = d_n$.

The roots of unity $\zeta_{p^n} \not\in \text{Norm}_{\mathbb{H}_n^+/\mathbb{K}_n}(E(\mathbb{H}_n^+))$: indeed, if $\zeta_{p^m} = N(\delta)$ for $\delta \in E(\mathbb{H}_n^+)$ and $m \leq n$, then $\varepsilon = \delta/\overline{\delta}$ is well defined in the CM field $\mathbb{H}_n^+$ and it is a root of unity, by Kronecker's unit Theorem – so $\varepsilon \in \mathbb{K}_n$. Moreover, we have $\text{Norm}_{\mathbb{H}_n^+/\mathbb{K}_n}(\varepsilon) = \varepsilon^{d_n} = \zeta_{p^m}^2$. Since p is odd, it follows from the above that $\mu_{p^n}/\mu_{p^n/d_n} \subset H^0(\mathbb{H}_n^+/\mathbb{K}_n)$; by comparing orders of the groups, we conclude that

$$H^0(\mathbb{H}_n^+/\mathbb{K}_n) = \mu_{p^n}/\mu_{(p^n/d_n)} = \left(H^0(\mathbb{H}_n^+/\mathbb{K}_n)\right)^-.$$

We have proved:

Lemma 1. *Notations being like above,*

$$\left(H^0(\mathbb{H}_n^+/\mathbb{K}_n)\right)^+ = \{1\}.$$

In particular

$$Norm_{\mathbb{H}_n^+/\mathbb{K}_n}(E^+(\mathbb{H}_n^+)) = E^+(\mathbb{K}_n), \tag{5}$$

where for a CM field $\mathbb{F}$ *we write* $E^+(\mathbb{F}) = \{e \cdot \overline{e} : e \in E(\mathbb{F})\}$.

In our case, E^+ are the real units and the units of $\mathbb{K}_n^+$, resp. $\mathbb{H}(\mathbb{K}_n^+) \subset \mathbb{H}_n^+$; the prime p is odd and we are interested in p-parts, so the implicit exponent 2 in the above definition has no further consequences: the norm is surjective on the real units in our class field.

From $\mathbb{H}_{n+1}^+ = \mathbb{K}_{n+1} \cdot \mathbb{H}_n^+$ and we have a commutative diagram of fields. For $n > n_0$, the capitulation kernel Ker $(\iota_{n,n+1} : A_n^+ \to A_{n+1}^+)$ has constant size p, since A_n^+ is cyclic. Thus $|H^0(\mathbb{K}_{n+1}/\mathbb{K}_n)| = p$. The idea of the proof will be to show that

$$N_{\mathbb{H}_{n+1}^+/\mathbb{H}_n^+}(E(\mathbb{H}_{n+1}^+)) \cdot \text{ Ker } (N : E(\mathbb{H}_n^+) \to E(\mathbb{K}_n)) = E(\mathbb{H}_n^+). \tag{6}$$

In view of (5), we obtain $N_{\mathbb{H}_{n+1}^+/\mathbb{K}_n}(E(\mathbb{H}_{n+1}^+)) = E(\mathbb{K}_n)$, which contradicts the fact that $|H^0(\mathbb{K}_{n+1}/\mathbb{K}_n)| = p$ established above, for sufficiently large n. The core observation of our proof is

Proposition 1. *Let* $\lambda_n = 1 - \zeta_{p^n}$. *Then the ramified prime* $\wp_n = (\lambda_n) \subset \mathbb{K}_n$ *above* p *splits totally in* $\mathbb{H}_n^+$ *in* p*-principal ideals and there is a* $\pi_n \in \mathbb{H}_n^+$ *with* $Norm_{\mathbb{H}_n^+/\mathbb{K}_n}(\pi_n) = \lambda_n^c, (c, p) = 1$.

Proof. The Lemma 2 implies that

$$(A(\mathbb{H}_n^+))^+ = \{1\}, \tag{7}$$

the plus part of Hilbert class fields being canonical. Since $\wp_n = (\lambda_n)$ is principal, the Principal Ideal Theorem (for the p-part of the Hilbert class field of $\mathbb{K}_n^+$) implies that it splits completely in the unramified extension $\mathbb{H}_n^+/\mathbb{K}_n$ and the primes above $\wp_n$ must be real; but $(A(\mathbb{H}_n^+))^+ = \{1\}$, so they must be p-principal. Let $(\pi_n) = \mathfrak{p}^c$ be a principal prime power with $\mathfrak{p}$ a prime above $\wp_n$ and $(c, p) = 1$. Then $N_{\mathbb{H}_n^+/\mathbb{K}_n}(\pi_n) = \varepsilon\lambda_n^c$ and in view of (5) we may assume that the unit $\varepsilon = 1$, which completes the proof.

The proof of this proposition is made particularly simple by the use of (2). However, a more involved proof shows that the facts hold in more generality and the primes above λ are principal in any subfield of the Hilbert class field $\mathbb{H}_n$.

We complete the proof of the theorem as indicated above, before Proposition 1. Recall that the irregular components are $e_{2k}A$ and $e_{p-2k}A$. Since $\mathbb{H}_{n+1}^{+}/\mathbb{H}_n$ is abelian, τ_n, a generator of Gal $(\mathbb{K}_{n+1}/\mathbb{K}_n)$ lifts naturally to Gal $(\mathbb{H}_{n+1}^{+}/\mathbb{H}_n^{+})$ and we write $T = \tau_n - 1$. We denote by $\nu \in$ Gal $(\mathbb{H}_n^{+}/\mathbb{K}_n)$ a generator of this group.

For n large enough, $H^0(\mathbb{K}_{n+1}/\mathbb{K}_n) = e_{2k} H^0(\mathbb{K}_{n+1}/\mathbb{K}_n) \cong \mathbb{Z}/(p \cdot \mathbb{Z})$. We shall prove (6). Since $N_{\mathbb{H}_n^{+}/\mathbb{K}_n}(E(\mathbb{H}_n^{+})) = E(\mathbb{K}_n)$, there is some unit $e_n \in E(\mathbb{H}_n^{+})$ with non trivial image in $E(\mathbb{H}_n^{+})/E^{(p,(\nu-1))}$ and such that $e_n^{\mathbb{Z}} \cap N_{\mathbb{H}_{n+1}^{+}/\mathbb{H}_n^{+}}(E(\mathbb{H}_{n+1}^{+})) = e_n^{p\mathbb{Z}}$. Since $C(\mathbb{K}_n) \subset \mathbb{N}_{\mathbb{H}_{n+1}^{+}/\mathbb{K}_n}(E(\mathbb{H}_{n+1}^{+}))$ as a consequence of Proposition 1, we must have $\varepsilon_n := N_{\mathbb{H}_n^{+}/\mathbb{K}_n}(e_n) \in E(\mathbb{K}_n) \setminus C(\mathbb{K}_n)$. Let concretely $e_{n+1} \in E(\mathbb{H}_{n+1}^{+})$ with $e_n^p = N_{\mathbb{H}_{n+1}^{+}/\mathbb{H}_n^{+}}(e_{n+1})$ and $e = e_{n+1}/e_n$, so $N_{\mathbb{H}_{n+1}^{+}/\mathbb{H}_n^{+}}(e) = 1$. Let us write

$$\varepsilon_m := N_{\mathbb{H}_m^{+}/\mathbb{K}_m}(e_m), m \in \{n, n+1\},$$

and $\delta := \varepsilon_{n+1}/\varepsilon_n = N_{\mathbb{H}_{n+1}^{+}/\mathbb{K}_{n+1}}(e)$. Note that by construction we have $\varepsilon_m E(\mathbb{K}_m)^p \in e_{2k} E(\mathbb{K}_m)/E^p(\mathbb{K}_m)$. The components $e_{2k} E(\mathbb{K}_m)/E(\mathbb{K}_m)^p$ are cyclic, and due to vertical capitulation, there is an ambig ideal $(\omega) \subset \mathbb{K}_{n+1}$ such that $\delta = \omega^T$.

By construction, we also have $\delta \in E(\mathbb{K}_{n+1}) \setminus C(\mathbb{K}_{n+1})$. It follows from Hilbert 90 that $e = \rho^T, \rho \in E'(\mathbb{H}_{n+1})$. Indeed, since $\rho^T \in E(\mathbb{H}_{n+1})$, it follows that (ρ) is an ambig ideal, so $(\rho) = \mathfrak{R}\mathrm{id}O(\mathbb{H}_{n+1})$ for some ideal $\mathfrak{R} \subset \mathbb{H}_n$, which is either ramified or capitulates in $\mathbb{H}_{n+1}/\mathbb{H}_n$. Since $A(\mathbb{H}_n^{+})^{+} = \{1\}$, the second alternative is not possible, so it remains that $\mathfrak{R}$ is a ramified ideal: but the only primes ramifying in $\mathbb{H}_{n+1}/\mathbb{H}_n$ are the primes above p and we may conclude that $\rho \in E'(\mathbb{H}_{n+1})$. Taking the norm $N_{\mathbb{H}_{n+1}^{+}/\mathbb{K}_{n+1}}$ we obtain $(\omega/\mathrm{Norm}_{\mathbb{H}_{n+1}/\mathbb{K}_{n+1}}(\rho))^T = 1$, hence $\delta = \mathrm{Norm}_{\mathbb{H}_{n+1}/\mathbb{K}_{n+1}}(\rho)^{-T} \in E'(\mathbb{K}_{n+1})^T$. But then $\delta \in E'(\mathbb{K}_{n+1}^T) \cap E(\mathbb{K}_{n+1}) \subset C(\mathbb{K}_{n+1})$ is a cyclotomic unit, in contradiction with the choice of e_n, e_{n+1} and the remark above. This contradiction confirms (6) and completes the proof of Theorem 1.

A Appendix

For the sake of completeness, we give a proof of the following

Lemma 2. *Let $\mathbb{K}$ be a number field and A be the p – part of its class group, while $\mathbb{H}$ is the p – part of its Hilbert class field. If A is cyclic, then $A(\mathbb{H}) := \mathrm{id}C(\mathbb{H})_p = \{1\}$.*

Proof. Since A is cyclic, Gal $(\mathbb{H}/\mathbb{K}) \cong A$ is a cyclic group and the ideals of a generating class $a \in A$ are inert and become principal in $\mathbb{H}$. Let $\sigma = \varphi(a) \in$ Gal $(\mathbb{H}/\mathbb{K})$ be a generator and $s = \sigma - 1$. Suppose that $b \in A(\mathbb{H}) \setminus A(\mathbb{H})^{(s,p)}$ is a non trivial class and let $\mathfrak{Q} \in b$ be an ideal above a rational prime $\mathfrak{q} \subset \mathbb{K}$, which splits completely in $\mathbb{H}/\mathbb{K}$: such a prime must exist, by Tchebotarew's Theorem. Since $\mathbb{H}/\mathbb{K}$ is the p-part of the Hilbert class field of $\mathbb{K}$, and $\mathfrak{q}$ is totally split, it must be a principal ideal, so $b' = [\mathfrak{q}] = 1$. Therefore $\mathrm{Norm}_{\mathbb{H}/\mathbb{K}}(b) = b' = 1$, and Furtwängler's Hilbert 90 Theorem for ideal class groups in unramified cyclic extensions [1] says that $\mathrm{Ker}\ (\mathrm{Norm}: A(\mathbb{H}) \rightarrow A(\mathbb{K})) \subset A(\mathbb{H})^s$, which implies that $b \in A(\mathbb{H})^s$. This contradicts the choice of b and completes the proof.

Acknowledgements I thank the anonymous referee for helpful questions and suggestions, which helped improve the clarity of the exposition.

References

1. P. Furtwängler, Über die Reziprozitätsgesetze zwischen l-ten Potenzresten in algebraischen Zahlkörpern, wenn l eine ungerade Primzahl bedeutet. Math. Ann. (German) **58**, 1–50 (1904)
2. K. Iwasawa, A note on the group of units of an algebraic number fields. J. de Math. Pures et Appl. **35/121**, 189–192 (1956)
3. S. Lang, *Cyclotomic Fields I and II*, 1st edn. (Springer, New York, 1978/1980)
4. P. Ribenboim, *13 Lectures on Fermat's Last Theorem* (Springer, New York, 1979)
5. L. Washington, *Introduction to Cyclotomic Fields*. Graduate Texts in Mathematics, vol. 83, 2nd edn. (Springer, New York/London, 1996)

Schwartzman Cycles and Ergodic Solenoids

Vicente Muñoz and Ricardo Pérez Marco

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract We extend Schwartzman theory beyond dimension 1 and provide a unified treatment of Ruelle-Sullivan and Schwartzman theories via Birkhoff's ergodic theorem for the class of immersions of solenoids with a trapping region.

1 Introduction

This is the second paper of a series of articles [1–5] in which we aim to give a geometric realization of *real* homology classes in smooth manifolds. This paper is devoted to the definition of Schwartzman homology classes and its relationship with the generalized currents associated to solenoids defined in [2].

Let M be a smooth manifold. A closed oriented submanifold $N \subset M$ of dimension $k \geq 0$ determines a homology class in $H_k(M, \mathbb{Z})$. This homology class in $H_k(M, \mathbb{R})$, as dual of de Rham cohomology, is explicitly given by integration of the restriction to N of differential k-forms on M. Unfortunately, because of topological reasons dating back to Thom [8], not all integer homology classes in $H_k(M, \mathbb{Z})$ can be realized in such a way. Geometrically, we can realize any class in $H_k(M, \mathbb{Z})$ by topological k-chains. The real homology $H_k(M, \mathbb{R})$ classes are only realized by formal combinations with real coefficients of k-cells. This is not fully satisfactory.

V. Muñoz (✉)
Facultad de Matemáticas, Universidad Complutense de Madrid, Plaza de Ciencias 3, 28040 Madrid, Spain
e-mail: vicente.munoz@mat.ucm.es

R.P. Marco
Université Paris XIII,99, Avenue J.-B. Clément, 93430-Villetaneuse, France
e-mail: ricardo@math.univ-paris13.fr

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_13, © Springer-Verlag Berlin Heidelberg 2012

For various reasons, it is important to have an explicit realization, as geometric as possible, of real homology classes.

The first contribution in this direction came in 1957 from the work of Schwartzman [7]. Schwartzman showed how, by a limiting procedure, one-dimensional curves immersed in M can define a real homology class in $H_1(M, \mathbb{R})$. More precisely, he proved that this happens for almost all curves solutions to a differential equation admitting an invariant ergodic probability measure. Schwartzman's idea consists on integrating 1-forms over large pieces of the parametrized curve and normalizing this integral by the length of the parametrization. Under suitable conditions, the limit exists and defines an element of the dual of $H^1(M, \mathbb{R})$, i.e. an element of $H_1(M, \mathbb{R})$. This procedure is equivalent to the more geometric one of closing large pieces of the curve by relatively short closing paths. The closed curve obtained defines an integer homology class. The normalization by the length of the parameter range provides a class in $H_1(M, \mathbb{R})$. Under suitable hypothesis, there exists a unique limit in real homology when the pieces exhaust the parametrized curve, and this limit is independent of the closing procedure. In Sects. 4 and 5, we shall study this circle of ideas in great generality. In Sect. 4 we shall define Schwartzman cycles for parametrized and unparametrized curves in M, and study their properties. In Sect. 5, we explore an alternative route to define real homology classes associated to curves in M by using the universal covering $\pi : \tilde{M} \to M$.

It is natural to ask whether it is possible to realize every real homology class using Schwartzman limits. By the result of [4], we can realize any real homology class by the generalized current associated to an immersed oriented uniquely ergodic solenoid. A *solenoid* (see [2]) is an abstract laminated space endowed with a transversal structure. For these oriented solenoids we can consider k-forms that we can integrate, provided that we are given a transversal measure invariant by the holonomy group. An immersion of a solenoid S into M is a regular map $f : S \to M$ that is an immersion in each leaf. If the solenoid S is endowed with a transversal measure $\mu = (\mu_T)$, then any smooth k-form in M can be pulled back to S by f and integrated. The resulting numerical value only depends on the cohomology class of the k-form. Therefore we have defined a closed current that we denote by (f, S_μ) and that we call a generalized current [2]. It defines a homology class $[f, S_\mu] \in H_k(M, \mathbb{R})$. This is reviewed in Sect. 2.

In Sect. 6, we study the relation between the generalized current defined by an immersed oriented measured 1-solenoid S_μ and the Schwartzman measure defined by any one of its leaves. The relationship is best expressed for ergodic and uniquely ergodic solenoids. In the first case, almost all μ_T-leaves define Schwartzman classes which represent $[f, S_\mu]$. In the second case, the property holds for all leaves.

Section 7 is devoted to the generalization of the Schwartzman theory to higher dimensions. For a complete k-dimensional immersed submanifold $N \subset M$ of a Riemannian manifold, we define a Schwartzman class by taking large balls, closing them with small caps, normalizing the homology class thus obtained and finally taking the limit. This process is only possible when such capping exist. If S is a k-solenoid immersed in M, one would naturally expect that there is some relation between the generalized currents and the Schwartzman current (if defined) of the leaves. The main result is that there is such relation for the class of minimal, ergodic solenoids with a trapping region (see Definition 26). For such solenoids, the holonomy group is generated by a single map. Then the bridge between generalized currents and Schwartzman currents of the leaves is provided by Birkhoff's ergodic theorem. We prove the following:

Theorem 1. *Let S_μ be an oriented and minimal solenoid endowed with an ergodic transversal measure μ, and possessing a trapping region W. Let $f : S_\mu \to M$ be an immersion of S_μ into M such that $f(W)$ is contained in a ball. Then for μ_T-almost all leaves $l \subset S_\mu$, the Schwartzman homology class of $f(l) \subset M$ is well defined and coincides with the homology class $[f, S_\mu]$.*

We are particularly interested in uniquely ergodic solenoids, with only one ergodic transversal measure. As is well known, in this situation we have uniform convergence of Birkhoff's sums, which implies the stronger result:

Theorem 2. *Let S_μ be a minimal, oriented and uniquely ergodic solenoid which has a trapping region W. Let $f : S_\mu \to M$ be an immersion of S_μ into M such that $f(W)$ is contained in a ball. Then for all leaves $l \subset S_\mu$, the Schwartzman homology class of $f(l) \subset M$ is well defined and coincides with the homology class $[f, S_\mu]$.*

2 Solenoids and Generalized Currents

Let us review the main concepts introduced in [2], and that we shall use later in this paper.

Definition 1. A k-solenoid, where $k \geq 0$, of class $C^{r,s}$, is a compact Hausdorff space endowed with an atlas of flow-boxes $\mathcal{A} = \{(U_i, \varphi_i)\}$,

$$\varphi_i : U_i \to D^k \times K(U_i),$$

where D^k is the k-dimensional open ball, and $K(U_i) \subset \mathbb{R}^l$ is the transversal set of the flow-box. The changes of charts $\varphi_{ij} = \varphi_i \circ \varphi_j^{-1}$ are of the form

$$\varphi_{ij}(x, y) = (X(x, y), Y(y)), \tag{1}$$

where $X(x, y)$ is of class $C^{r,s}$ and $Y(y)$ is of class C^s.

Let S be a k-solenoid, and $U \cong D^k \times K(U)$ be a flow-box for S. The sets $L_y = D^k \times \{y\}$ are called the (local) leaves of the flow-box. A leaf $l \subset S$ of the solenoid is a connected k-dimensional manifold whose intersection with any flow-box is a collection of local leaves. The solenoid is oriented if the leaves are oriented (in a transversally continuous way).

A transversal for S is a subset T which is a finite union of transversals of flow-boxes. Given two local transversals T_1 and T_2 and a path contained in a leaf from a point of T_1 to a point of T_2, there is a well-defined holonomy map $h : T_1 \to T_2$. The holonomy maps form a pseudo-group.

A k-solenoid S is minimal if it does not contain a proper sub-solenoid. By [2, Sect. 2], minimal solenoids exist. If S is minimal, then any transversal is a global transversal, i.e., it intersects all leaves. In the special case of an oriented minimal 1-solenoid, the holonomy return map associated to a local transversal,

$$R_T : T \to T$$

is known as the Poincaré return map (see [2, Sect. 4]).

Definition 2. Let S be a k-solenoid. A transversal measure $\mu = (\mu_T)$ for S associates to any local transversal T a locally finite measure μ_T supported on T, which are invariant by the holonomy pseudogroup, i.e. if $h : T_1 \to T_2$ is a holonomy map, then $h_* \mu_{T_1} = \mu_{T_2}$.

We denote by S_μ a k-solenoid S endowed with a transversal measure $\mu = (\mu_T)$. We refer to S_μ as a measured solenoid. Observe that for any transversal measure $\mu = (\mu_T)$ the scalar multiple $c\,\mu = (c\,\mu_T)$, where $c > 0$, is also a transversal measure. Notice that there is no natural scalar normalization of transversal measures.

Definition 3 (*Transverse ergodicity*). A transversal measure $\mu = (\mu_T)$ on a solenoid S is ergodic if for any Borel set $A \subset T$ invariant by the pseudo-group of holonomy maps on T, we have

$$\mu_T(A) = 0 \quad \text{or} \quad \mu_T(A) = \mu_T(T).$$

We say that S_μ is an ergodic solenoid.

Definition 4. Let S be a k-solenoid. The solenoid S is uniquely ergodic if it has a unique (up to scalars) transversal measure μ and its support is the whole of S.

Now let M be a smooth manifold of dimension n. An immersion of a k-solenoid S into M, with $k < n$, is a smooth map $f : S \to M$ such that the differential restricted to the tangent spaces of leaves has rank k at every point of S. The solenoid $f : S \to M$ is transversally immersed if for any flow-box $U \subset S$ and chart $V \subset M$, the map $f : U = D^k \times K(U) \to V \subset \mathbb{R}^n$ is an embedding, and the images of the leaves intersect transversally in M. If moreover f is injective, then we say that the solenoid is embedded.

Note that under a transversal immersion, resp. an embedding, $f : S \to M$, the images of the leaves are immersed, resp. injectively immersed, submanifolds.

Let $\mathcal{C}_k(M)$ denote the space of k-dimensional currents on M.

Definition 5. Let S_μ be an oriented measured k-solenoid. An immersion $f : S \to M$ defines a generalized Ruelle-Sullivan current $(f, S_\mu) \in \mathcal{C}_k(M)$ as follows. Let $S = \bigcup_i S_i$ be a measurable partition such that each S_i is contained in a flow-box U_i. For $\omega \in \Omega^k(M)$, we define

$$\langle (f, S_\mu), \omega \rangle = \sum_i \int_{K(U_i)} \left(\int_{L_y \cap S_i} f^*\omega \right) d\mu_{K(U_i)}(y),$$

where L_y denotes the horizontal disk of the flow-box.

In [2] it is proved that (f, S_μ) is a closed current. Therefore, it defines a real homology class

$$[f, S_\mu] \in H_k(M, \mathbb{R}).$$

In their original article [6], Ruelle and Sullivan defined this notion for the restricted class of solenoids embedded in M.

3 Schwartzman Measures

Let S be a Riemannian k-solenoid, that is, a solenoid endowed with a Riemannian metric on each leaf. In some situations, we may define transversal measures associated to S by considering large chunks of a single leaf $l \subset S$. These will be called Schwartzman measures. We start by recalling some notions from [2, Sect. 6].

Definition 6 (*daval measures*). Let μ be a measure supported on S. The measure μ is a daval measure if it disintegrates as volume along leaves of S, i.e. for any flow-box (U, φ) with local transversal $T = \varphi^{-1}(\{0\} \times K(U))$, we have a measure $\mu_{U,T}$ supported on T such that for any Borel set $A \subset U$

$$\mu(A) = \int_T \mathrm{Vol}_k(A_y)\, d\mu_{U,T}(y), \tag{2}$$

where $A_y = A \cap \varphi^{-1}(D^k \times \{y\}) \subset U$.

We denote by $\mathcal{M}_{\mathcal{L}}(S)$ the space of probability daval measures, by $\mathcal{M}_{\mathcal{T}}(S)$ the space of (non-zero) transversal measures on S, and by $\overline{\mathcal{M}}_{\mathcal{T}}(S)$ the quotient of $\mathcal{M}_{\mathcal{T}}(S)$ by positive scalars. The following result is Theorem 6.8 in [2].

Theorem 3 (Transverse measures of the Riemannian solenoid). *There is a one-to-one correspondence between transversal measures* (μ_T) *and finite daval measures* μ. *Furthermore, there is an isomorphism*

$$\overline{\mathcal{M}}_{\mathcal{T}}(S) \cong \mathcal{M}_{\mathcal{L}}(S).$$

The correspondence follows from (2). If S is a uniquely ergodic Riemannian solenoid, then the above result allows to normalize the transversal measure in a unique way, by imposing that the corresponding daval measure has total mass 1.

Now we introduce a subclass of solenoids for which daval measures do exist.

Definition 7 (*Controlled growth solenoids*). Let S be a Riemannian solenoid. Fix a leaf $l \subset S$ and an exhaustion (C_n) by subsets of l. For a flow-box (U, φ) write

$$C_n \cap U = A_n \cup B_n,$$

where A_n is composed by all full disks $L_y = \varphi^{-1}(D^k \times \{y\})$ contained in C_n, and B_n contains those connected components B of $C_n \cap U$ such that $B \neq L_y \cap U$ for any y. The solenoid S has controlled growth with respect to l and (C_n) if for any flow-box U in a finite covering of S

$$\lim_{n\to+\infty} \frac{\mathrm{Vol}_k(B_n)}{\mathrm{Vol}_k(A_n)} = 0.$$

The solenoid S has controlled growth if S contains a leaf l and an exhaustion (C_n) such that S has controlled growth with respect to l and (C_n).

For a Riemannian solenoid S, it is natural to consider the exhaustion by Riemannian balls $B(x_0, R_n)$ in a leaf l centered at a point $x_0 \in l$ and with $R_n \to +\infty$, and test the controlled growth condition with respect to such exhaustions.

The controlled growth condition depends a priori on the Riemannian metric. As we see next, it guarantees the existence of daval measures, hence the existence of transversal measures on S. Indeed the measures we construct are Schwartzman measures defined as:

Definition 8 (*Schwartzman limits and measures*). We say that a measure μ is a Schwartzman measure if it is obtained as the limit

$$\mu = \lim_{n\to+\infty} \mu_n,$$

where the measures (μ_n) are the normalized k-volume of the exhaustion (C_n) (that is, μ_n are normalized to have total mass 1). We denote by $\mathcal{M}_{\mathcal{S}}(S)$ the space of (probability) Schwartzman measures.

Compactness of probability measures show:

Proposition 1. *There are always Schwartzman measures on S,*

$$\mathcal{M}_{\mathcal{S}}(S) \neq \emptyset.$$

Theorem 4. *If S is a solenoid with controlled growth, then any Schwartzman measure is a daval measure,*

$$\mathcal{M}_{\mathcal{S}}(S) \subset \mathcal{M}_{\mathcal{L}}(S).$$

In particular, $\mathcal{M}_{\mathcal{L}}(S) \neq \emptyset$ and S admits transversal measures.

Proof. Let $\mu_n \to \mu$ be a Schwartzman limit as in Definition 8. For any flow-box U we prove that μ disintegrates as volume on leaves of U. Since S has controlled growth, pick a leaf and an exhaustion which satisfy the controlled growth condition. Let

$$C_n \cap U = A_n \cup B_n ,$$

be the decomposition for $C_n \cap U$ described before. The set A_n is composed of a finite number of horizontal disks. We define a new measure ν_n with support in U which is the restriction of μ_n to A_n, i.e. it is proportional to the k-volume on horizontal disks. The measure ν_n disintegrates as volume on leaves in U. The transversal measure is a finite sum of Dirac measures. Moreover the controlled growth condition implies that (ν_n) and $(\mu_{n|U})$ must converge to the same limit. But we know that $\mathcal{M}_{\mathcal{L}}(S)$ is closed, thus the limit measure $\mu_{|U}$ disintegrates on leaves in U. So μ is a daval measure.

For uniquely ergodic solenoids we have:

Corollary 1. *The volume μ of a uniquely ergodic solenoid with controlled growth is the unique Schwartzman measure. Therefore there is only one Schwartzman limit*

$$\mu = \lim_{n\to+\infty} \mu_n ,$$

which is independent of the leaf and the exhaustion.

Proof. There are always Schwartzman limits. Theorem 4 shows that any such limit μ disintegrates as volume on leaves. Thus the measure μ defines the unique (up to scalars) transversal measure (μ_T). But, conversely, the transversal measure determines the measure μ uniquely. Therefore there is only possible limit μ, which is the volume of the uniquely ergodic solenoid.

4 Schwartzman Clusters and Asymptotic Cycles

Let M be a compact C^∞ Riemannian manifold. Observe that since $H_1(M, \mathbb{R})$ is a finite dimensional real vector space, it comes equipped with a unique topological vector space structure.

The map $\gamma \mapsto [\gamma]$ that associates to each loop its homology class in $H_1(M, \mathbb{Z}) \subset H_1(M, \mathbb{R})$ is continuous when the space of loops is endowed with the Hausdorff topology. Therefore, by compactness, oriented rectifiable loops in M of uniformly bounded length define a bounded set in $H_1(M, \mathbb{R})$.

We have a more precise quantitative version of this result.

Lemma 1. *Let (γ_n) be a sequence of oriented rectifiable loops in M, and (t_n) be a sequence with $t_n > 0$ and $t_n \to +\infty$. If*

$$\lim_{n\to+\infty} \frac{l(\gamma_n)}{t_n} = 0,$$

then in $H_1(M, \mathbb{R})$ *we have*

$$\lim_{n \to +\infty} \frac{[\gamma_n]}{t_n} = 0.$$

Proof. Via the map

$$\omega \mapsto \int_\gamma \omega,$$

each loop γ defines a linear map L_γ on $H^1(M, \mathbb{R})$ that only depends on the homology class of γ. We can extend this map to $\mathbb{R} \otimes H_1(M, \mathbb{Z})$ by

$$c \otimes \gamma \mapsto c \cdot L_\gamma .$$

We have the isomorphism

$$H_1(M, \mathbb{R}) = \mathbb{R} \otimes H_1(M, \mathbb{Z}) \cong \left(H^1(M, \mathbb{R})\right)^* .$$

The Riemannian metric gives a C^0-norm on forms. We consider the norm in $H^1(M, \mathbb{R})$ given as

$$||[\omega]||_{C^0} = \min_{\omega \in [\omega]} ||\omega||,$$

and the associated operator norm in $H_1(M, \mathbb{R}) \cong \left(H^1(M, \mathbb{R})\right)^*$.

We have

$$|L_\gamma([\omega])| = \left| \int_\gamma \omega \right| \leq l(\gamma) ||\varnothing||_{C^0} \leq l(\gamma) ||[\varnothing]||_{C^0},$$

so

$$||L_\gamma|| \leq l(\gamma).$$

Hence $l(\gamma_n)/t_n \to 0$ implies $L_{\gamma_n}/t_n \to 0$ which is equivalent to $[\gamma_n]/t_n \to 0$.

Definition 9 (*Schwartzman asymptotic* 1*-cycles*). Let c be a parametrized continuous curve $c : \mathbb{R} \to M$ defining an immersion of $\mathbb{R}$. For $s, t \in \mathbb{R}$, $s < t$, we choose a rectifiable oriented curve $\gamma_{s,t}$ joining $c(s)$ to $c(t)$ such that

$$\lim_{\substack{t \to +\infty \\ s \to -\infty}} \frac{l(\gamma_{s,t})}{t - s} = 0 .$$

The parametrized curve c is a Schwartzman asymptotic 1-cycle if the juxtaposition of $c|_{[s,t]}$ and $\gamma_{s,t}$, denoted $c_{s,t}$ (which is a 1-cycle), defines a homology class $[c_{s,t}] \in H_1(M, \mathbb{Z})$ such that the limit

$$\lim_{\substack{t \to +\infty \\ s \to -\infty}} \frac{[c_{s,t}]}{t - s} \in H_1(M, \mathbb{R}) \tag{3}$$

exists.

We define the Schwartzman asymptotic homology class as

$$[c] := \lim_{\substack{t\to+\infty\\ s\to-\infty}} \frac{[c_{s,t}]}{t-s}.$$

Thanks to Lemma 1 this definition does not depend on the choice of the closing curves $(\gamma_{s,t})$. If we take another choice $(\gamma'_{s,t})$, then as homology classes,

$$[c_{s,t}] = [c'_{s,t}] + [\gamma'_{s,t} - \gamma_{s,t}],$$

and

$$\frac{l(\gamma'_{s,t} - \gamma_{s,t})}{t-s} = \frac{l(\gamma'_{s,t})}{t-s} + \frac{l(\gamma_{s,t})}{t-s} \to 0,$$

as $t \to \infty$, $s \to -\infty$. By Lemma 1,

$$\lim_{\substack{t\to+\infty\\ s\to-\infty}} \frac{[\gamma_{s,t} - \gamma'_{s,t}]}{t-s} = 0,$$

thus

$$[c] = \lim_{\substack{t\to+\infty\\ s\to-\infty}} \frac{[c_{s,t}]}{t-s} = \lim_{\substack{t\to+\infty\\ s\to-\infty}} \frac{[c'_{s,t}]}{t-s}.$$

Note that we do not assume that $c(\mathbb{R})$ is an embedding of $\mathbb{R}$, i.e. $c(\mathbb{R})$ could be a loop. In that case, the Schwartzman asymptotic homology class coincides with a scalar multiple (the scalar depending on the parametrization) of the integer homology class $[c(\mathbb{R})]$. This shows that the Schwartzman homology class is a generalization to the case of immersions $c : \mathbb{R} \to M$. More precisely we have:

Proposition 2. *If $c : \mathbb{R} \to M$ is a loop then it is a Schwartzman asymptotic 1-cycle and the Schwartzman asymptotic homology class is a scalar multiple of the homology class of the loop $[c(\mathbb{R})] \in H_1(M, \mathbb{Z})$.*

If $c : \mathbb{R} \to M$ is a rectifiable loop with its arc-length parametrization, and $l(c)$ is the length of the loop c, then

$$[c] = \frac{1}{l(c)} [c(\mathbb{R})].$$

Proof. Let $t_0 > 0$ be the minimal period of the map $c : \mathbb{R} \to M$. Then

$$[c_{s,t}] = \left[\frac{t-s}{t_0}\right] [c(\mathbb{R})] + O(1).$$

Then

$$\lim_{\substack{t\to+\infty\\ s\to-\infty}} \frac{[c_{s,t}]}{t-s} = \frac{1}{t_0}[c(\mathbb{R})].$$

When $c : \mathbb{R} \to M$ is the arc-length parametrization of a rectifiable loop, the period t_0 coincides with the length of the loop.

Remark 1. In Definition 9, we have required that $(\gamma_{s,t})$ satisfies $l(\gamma_{s,t})/(t-s) \to 0$ uniformly and separately on s and t when $t \to +\infty$ and $s \to -\infty$, i.e., that $l(\gamma_{s,t}) = o(t-s) = o(|t|+|s|)$. Actually, as M is compact, we can choose paths $\gamma_{s,t}$ with length bounded by the diameter of M, i.e., $l(\gamma_{s,t}) = O(1)$. Even more, we can take $\{\gamma_{s,t}; s < t\}$ contained in a compact subset of the space of continua of M. Then the uniform boundedness will hold for any Riemannian metric and the notions defined will not depend on the Riemannian structure. For simplicity, we shall do this in the sequel.

Definition 10 (*Positive and negative asymptotic cycles*). Under the assumptions of Definition 9, if the limit

$$\lim_{t\to+\infty} \frac{[c_{s,t}]}{t-s} \in H_1(M,\mathbb{R}) \tag{4}$$

exists then it does not depend on s, and we say that the parametrized curve c defines a positive asymptotic cycle. The positive Schwartzman homology class is defined as

$$[c_+] = \lim_{t\to+\infty} \frac{[c_{s,t}]}{t-s}.$$

The definition of negative asymptotic cycle and negative Schwartzman homology class is the same but taking $s \to -\infty$,

$$[c_-] = \lim_{s\to-\infty} \frac{[c_{s,t}]}{t-s}.$$

The independence of the limit (4) on s follows from

$$\lim_{t\to+\infty} \frac{[c_{s',t}]}{t-s'} = \lim_{t\to+\infty} \frac{[c_{s,t}]+[c_{s',s}]+O(1)}{t-s} \cdot \frac{t-s}{t-s'} = \lim_{t\to+\infty} \frac{[c_{s,t}]}{t-s}.$$

Proposition 3. *A parametrized curve c is a Schwartzman asymptotic 1-cycle if and only if it is both a positive and a negative asymptotic cycle and*

$$[c_+] = [c_-].$$

In that case we have

$$[c] = [c_+] = [c_-].$$

Proof. If c is a Schwartzman asymptotic 1-cycle, then for $t \to +\infty$ take $s \to -\infty$ very slowly, say satisfying the relation $t = s^2\, l(c_{|[s,0]})$, which defines $s = s(t) < 0$ uniquely as a function of $t > 0$. Then

$$
\begin{aligned}
[c] &= \lim_{\substack{t\to\infty \\ s=s(t)\to-\infty}} \frac{[c_{s,t}]}{t-s} = \lim_{t\to+\infty} \frac{[c_{s,0}]+[c_{0,t}]+O(1)}{t-s} \\
&= \lim_{t\to+\infty} \left(\frac{[c_{s,0}]+O(1)}{t} + \frac{[c_{0,t}]}{t} \right) \frac{t}{t-s} = \lim_{t\to+\infty} \frac{[c_{0,t}]}{t},
\end{aligned}
$$

since $\frac{t}{t-s} \to 1$ because $\frac{s}{t} \to 0$, and $\frac{[c_{s,0}]}{t} \to 0$ by Lemma 1. So c is a positive asymptotic cycle and $[c] = [c_+]$. Analogously, c is a negative asymptotic cycle and $[c] = [c_-]$.

Conversely, assume that c is a positive and negative asymptotic cycle with $[c_+] = [c_-]$. For t large we have

$$\frac{[c_{0,t}]}{t} = [c_+] + o(1).$$

For $-s$ large we have

$$\frac{[c_{s,0}]}{-s} = [c_-] + o(1).$$

Now

$$\frac{[c_{s,t}]}{t-s} = \frac{-s}{t-s} \cdot \frac{[c_{s,0}]}{-s} + \frac{t}{t-s} \cdot \frac{[c_{0,t}]}{t} + \frac{O(1)}{t-s} = \frac{-s}{t-s}[c_+] + \frac{t}{t-s}[c_-] + o(1).$$

As $[c_+] = [c_-]$, we get that this limit exists and equals $[c] = [c_+] = [c_-]$.

Definition 11 (*Schwartzman clusters*). Under the assumptions of Definition 9, we can consider, regardless of whether (3) exists or not, all possible limits

$$\lim_{n\to+\infty} \frac{[c_{s_n,t_n}]}{t_n - s_n} \in H_1(M,\mathbb{R}), \tag{5}$$

with $t_n \to +\infty$ and $s_n \to -\infty$, that is, the derived set of $([c_{s,t}]/(t-s))_{t\to\infty, s\to-\infty}$. The limits (5) are called Schwartzman asymptotic homology classes of c, and they form the Schwartzman cluster of c,

$$\mathcal{C}(c) \subset H_1(M,\mathbb{R}).$$

A Schwartzman asymptotic homology class (5) is balanced when the two limits

$$\lim_{n\to+\infty} \frac{[c_{0,t_n}]}{t_n} \in H_1(M,\mathbb{R}),$$

and

$$\lim_{n\to+\infty} \frac{[c_{s_n,0}]}{-s_n} \in H_1(M,\mathbb{R}),$$

do exist in $H_1(M,\mathbb{R})$, but are not necessarily equal. We denote by $\mathcal{C}_b(c) \subset \mathcal{C}(c) \subset H_1(M,\mathbb{R})$ the set of those balanced Schwartzman asymptotic homology classes. The set $\mathcal{C}_b(c)$ is named the balanced Schwartzman cluster.

We define also the positive and negative Schwartzman clusters, $\mathcal{C}_+(c)$ and $\mathcal{C}_-(c)$, by taking only limits $t_n \to +\infty$ and $s_n \to -\infty$ respectively.

Proposition 4. *The Schwartzman clusters $\mathcal{C}(c)$, $\mathcal{C}_+(c)$ and $\mathcal{C}_-(c)$ are closed subsets of $H_1(M,\mathbb{R})$.*

If $\{[c_{s,t}]/(t-s); s<t\}$ is bounded in $H_1(M,\mathbb{R})$, then the Schwartzman clusters $\mathcal{C}(c)$, $\mathcal{C}_+(c)$ and $\mathcal{C}_-(c)$ are non-empty, compact and connected subsets of $H_1(M,\mathbb{R})$.

Proof. The Schwartzman cluster $\mathcal{C}(c)$ is the derived set of

$$([c_{s,t}]/(t-s))_{t\to\infty, s\to-\infty},$$

in $H_1(M,\mathbb{R})$, hence closed.

Under the boundedness assumption, non-emptiness and compactness follow. Also the oscillation of $([c_{s,t}])_{s,t}$ is bounded by the size of $[\gamma_{s,t}]$. Therefore the magnitude of the oscillation of $([c_{s,t}]/(t-s))_{s,t}$ tends to 0 as $t\to\infty$, $s\to-\infty$. This forces the derived set to be connected under the boundedness assumption, since it is ϵ-connected for each $\epsilon>0$. (A compact metric space is ϵ-connected for all $\epsilon>0$ if and only if it is connected.)

Also $\mathcal{C}_+(c)$, resp. $\mathcal{C}_-(c)$, is closed because it is the derived set of

$$([c_{0,t}]/t)_{t\to\infty},$$

resp.

$$([c_{s,0}]/(-s))_{s\to-\infty},$$

in $H_1(M,\mathbb{R})$. Non-emptiness, compactness and connectedness under the boundedness assumption follow for the cluster sets $\mathcal{C}_\pm(c)$ in the same way as for $\mathcal{C}(c)$.

Note that all these cluster sets may be empty if the parametrization is too fast.

The balanced Schwartzman cluster $\mathcal{C}_b(c)$ does not need to be closed, as shown in the following counter-example.

Counter-example 5 *We consider the torus $M=\mathbb{T}^2$. We identify $H_1(M,\mathbb{R})\cong\mathbb{R}^2$, with $H_1(M,\mathbb{Z})$ corresponding to the lattice $\mathbb{Z}^2\subset\mathbb{R}^2$. Consider a line l in $H_1(M,\mathbb{R}^2)$ of irrational slope passing through the origin, $y=\sqrt{2}\,x$ for example. We can find a sequence of pairs of points $(a_n,b_n)\in\mathbb{Z}^2\times\mathbb{Z}^2$ in the open lower half plane H_l determined by the line l, such that the sequence of segments $[a_n,b_n]$ do converge to the line l, and the middle point $(a_n+b_n)/2\to 0$ (this is an easy exercise in diophantine approximation). We assume that the first coordinate of b_n tends to $+\infty$, and the first coordinate of a_n tends to $-\infty$. Now we can construct a parametrized curve c on $\mathbb{T}^2$ such that for all $n\geq 1$ there are an infinite number of times $t_{n,i}\to+\infty$ with $[c_{0,t_{n,i}}]/t_{n,i}=b_n$, and for an infinite number of times $s_{n,i}\to-\infty$, $[c_{s_{n,i},0}]/(-s_{n,i})=a_n$. Thus in homology the curve c oscillates wildly. We can adjust the velocity of the parametrization so that $-s_{n,i}=t_{n,i}$. Hence for these times*

$$\frac{[c_{s_{n,i},t_{n,i}}]}{t_{n,i}-s_{n,i}}=\frac{a_n(-s_{n,i})+b_n(t_{n,i})+O(1)}{t_{n,i}-s_{n,i}}\to\frac{a_n+b_n}{2},$$

when $i\to+\infty$, and the two ends balance each other. We have great freedom in constructing c, so that we may arrange to have always $[c_{s,t}]\subset H_l$. Then we get that $0\in\mathcal{C}(c)$ and all $(a_n+b_n)/2\in\mathcal{C}_b(c)$ but $0\notin\mathcal{C}_b(c)$.

We have that c is a Schwartzman asymptotic 1-cycle (resp. positive, negative) if and only if $\mathcal{C}(c)$ (resp. $\mathcal{C}_+(c)$, $\mathcal{C}_-(c)$) is reduced to one point. In that case the Schwartzman asymptotic 1-cycle is balanced. The next result generalizes Proposition 3. We need first a definition.

Definition 12. Let $A,B\subset V$ be subsets of a real vector space V. For $a,b\in V$ the segment $[a,b]\subset V$ is the convex hull of $\{a,b\}$ in V. The additive hull of A and B is

$$A\widehat{+}B=\bigcup_{\substack{a\in A\\ b\in B}}[a,b].$$

Proposition 5. *The Schwartzman balanced cluster $\mathcal{C}_b(c)$ is contained in the additive hull of $\mathcal{C}_+(c)$ and $\mathcal{C}_-(c)$*

$$\mathcal{C}_b(c)\subset\mathcal{C}_+(c)\widehat{+}\mathcal{C}_-(c).$$

Moreover, for each $a\in\mathcal{C}_+(c)$ and $b\in\mathcal{C}_-(c)$, we have

$$\mathcal{C}_b(c)\cap[a,b]\neq\emptyset.$$

Proof. Let $x\in\mathcal{C}_b(c)$,

$$x=\lim_{n\to+\infty}\frac{[c_{s_n,t_n}]}{t_n-s_n}.$$

We write

$$\frac{[c_{s_n,t_n}]}{t_n-s_n}=\frac{[c_{s_n,0}]}{-s_n}\cdot\frac{-s_n}{t_n-s_n}+\frac{[c_{0,t_n}]}{t_n}\cdot\frac{t_n}{t_n-s_n}+o(1),$$

and the first statement follows.

For the second, consider

$$a=\lim_{n\to+\infty}\frac{[c_{0,t_n}]}{t_n}\in\mathcal{C}_+(c),$$

and

$$b=\lim_{n\to+\infty}\frac{[c_{s_n,0}]}{-s_n}\in\mathcal{C}_-(c).$$

Then taking any accumulation point $\tau\in[0,1]$ of the sequence $(t_n/(t_n-s_n))_n\subset[0,1]$ and taking subsequences in the above formulas, we get a balanced Schwartzman homology class

$$c=\tau a+(1-\tau)b\in\mathcal{C}_b(c).$$

Corollary 2. *If $\mathcal{C}_+(c)$ and $\mathcal{C}_-(c)$ are non-empty, then $\mathcal{C}_b(c)$ is non-empty, and therefore $\mathcal{C}(c)$ is also non-empty.*

Note that we can have $\mathcal{C}_+(c) = \mathcal{C}_-(c) = \emptyset$ (then $\mathcal{C}_b(c) = \emptyset$) but $\mathcal{C}(c) \neq \emptyset$ (modify appropriately Counter-example 5).

There is one situation where we can assert that the balanced Schwartzman cluster set is closed.

Proposition 6. *If $B = \{[c_{s,t}]/(t-s); s < t\} \subset H_1(M, \mathbb{R})$ is a bounded set, then $\mathcal{C}(c)$, $\mathcal{C}_+(c)$, $\mathcal{C}_-(c)$ and $\mathcal{C}_b(c)$ are all compact sets. More precisely, they are all contained in the convex hull of $\overline{B}$.*

Proof. Obviously $\mathcal{C}(c)$, $\mathcal{C}_+(c)$ and $\mathcal{C}_-(c)$ are bounded as cluster sets of bounded sets, hence compact by Proposition 4.

In order to prove that $\mathcal{C}_b(c)$ is bounded, we observe that the additive hull of bounded sets is bounded, therefore boundedness follows from Proposition 5. We show that $\mathcal{C}_b(c)$ is closed. Since $\mathcal{C}_b(c) \subset \mathcal{C}(c)$ and $\mathcal{C}(c)$ is closed, any accumulation point x of $\mathcal{C}_b(c)$ is in $\mathcal{C}(c)$. Let

$$x = \lim_{n\to+\infty} \frac{[c_{s_n,t_n}]}{t_n - s_n},$$

and write as before

$$\frac{[c_{s_n,t_n}]}{t_n - s_n} = \frac{[c_{s_n,0}]}{-s_n} \cdot \frac{-s_n}{t_n - s_n} + \frac{[c_{0,t_n}]}{t_n} \cdot \frac{t_n}{t_n - s_n} + o(1).$$

Note that $([c_{s_n,0}]/(-s_n))_n$ and $([c_{0,t_n}]/t_n)_n$ stay bounded. Therefore we can extract converging subsequences and also for the sequence $(t_n/(t_n - s_n))_n \subset [0,1]$. The limit along these subsequences $t_{n_k} \to +\infty$ and $s_{n_k} \to -\infty$ give the Schwartzman homology class x, which turns out to be balanced.

The final statement follows from the above proofs.

The situation described in Proposition 6 is indeed quite natural. It arises each time that M is a Riemannian manifold and c is an arc-length parametrization of a rectifiable curve. In the following proposition we make use of the natural norm $||\cdot||$ in the homology of a Riemannian manifold defined in the Appendix A.

Proposition 7. *Let M be a Riemannian manifold and denote by $||\cdot||$ the norm in homology. If c is a rectifiable curve parametrized by arc-length then the cluster sets $\mathcal{C}(c)$, $\mathcal{C}_+(c)$, $\mathcal{C}_-(c)$ and $\mathcal{C}_b(c)$ are compact subsets of $\bar{B}(0,1)$, the closed ball of radius 1 for the norm in homology.*

So $\mathcal{C}(c)$ and $\mathcal{C}_\pm(c)$ are non-empty, compact and connected, and $\mathcal{C}_b(c)$ is non-empty and compact.

Proof. Observe that we have

$$l(c_{s,t}) = l(c|_{[s,t]}) + l(\gamma_{s,t}) = t - s + l(\gamma_{s,t}).$$

Thus

$$l([c_{s,t}]) \leq t - s + l(\gamma_{s,t}).$$

By Theorem 13,

$$||[c_{s,t}]|| \leq t - s + l(\gamma_{s,t}),$$

and

$$\left\| \frac{[c_{s,t}]}{t-s} \right\| \leq 1 + \frac{l(\gamma_{s,t})}{t-s}.$$

Since $\frac{l(\gamma_{s,t})}{t-s} \to 0$ uniformly, we get that $B = \{[c_{s,t}]/(t-s); s < t\} \subset H_1(M, \mathbb{R})$ is a bounded set.

By Proposition 4, $\mathcal{C}(c)$ and $\mathcal{C}_\pm(c)$ are non-empty, compact and connected. By Corollary 2, $\mathcal{C}_b(c)$ is non-empty and by Proposition 6, it is compact.

Obviously the previous notions depend heavily on the parametrization. For a non-parametrized curve we can also define Schwartzman cluster sets.

Definition 13. For a non-parametrized oriented curve $c \subset M$, we define the Schwartzman cluster $\mathcal{C}(c)$ as the union of the Schwartzman clusters for all orientation preserving parametrizations of c. We define the positive $\mathcal{C}_+(c)$, resp. negative $\mathcal{C}_-(c)$, Schwartzman cluster set as the union of all positive, resp. negative, Schwartzman cluster sets for all orientation preserving parametrizations.

For this notion to make sense we are forced to use curves $\gamma_{s,t}$ which satisfy $l(\gamma_{s,t}) = O(1)$ (see Remark 1).

Proposition 8. *For an oriented curve $c \subset M$ the Schwartzman clusters $\mathcal{C}(c)$, $\mathcal{C}_+(c)$ and $\mathcal{C}_-(c)$ are non-empty closed cones of $H_1(M, \mathbb{R})$. These cones are degenerate (i.e. reduced to $\{0\}$) if and only if $\{[c_{s,t}]; s < t\}$ is a bounded subset of $H_1(M, \mathbb{Z})$.*

Proof. We can choose the closing curves $\gamma_{s,t}$ only depending on $c(s)$ and $c(t)$ and not on the parameter values s and t, nor on the parametrization. Then the integer homology class $[c_{s,t}]$ only depends on the points $c(s)$ and $c(t)$ and not on the parametrization. Therefore, we can adjust the speed of the parametrization so that $[c_{s,t}]/(t-s)$ remains in a ball centered at 0. This shows that $\mathcal{C}(c)$ is not empty. Adjusting the speed of the parametrization we equally get that it contains elements that are not 0, provided that the set $\{[c_{s,t}]; s < t\}$ is not bounded in $H_1(M, \mathbb{Z})$. Certainly, if $\{[c_{s,t}]; s < t\}$ is bounded, all the cluster sets are reduced to $\{0\}$. Observe also that if $a \in \mathcal{C}(c)$ then any multiple λa, $\lambda > 0$, belongs to $\mathcal{C}(c)$, by considering the new parametrization with velocity multiplied by λ. So $\mathcal{C}(c)$ is a cone in $H_1(M, \mathbb{R})$.

Now we prove that $\mathcal{C}(c)$ is closed. Let $a_n \in \mathcal{C}(c)$ with $a_n \to a \in H_1(M, \mathbb{R})$. For each n we can choose a parametrization of c, say $c^{(n)} = \tilde{c} \circ \psi_n$ (here $\tilde{c}$ is a fixed parametrization and ψ_n is an orientation preserving homeomorphism of $\mathbb{R}$), and parameters s_n and t_n such that $||[c^{(n)}_{s_n,t_n}] - a|| \leq 1/n$ (considering any fixed norm in $H_1(M, \mathbb{R})$). For each n we can choose t_n as large as we like, and s_n negative as we like. Choose them inductively such that (t_n) and $(\psi_n(t_n))$ are both increasing sequences converging to $+\infty$, and (s_n) and $(\psi_n(s_n))$ are both decreasing sequences converging to $-\infty$. Construct a homeomorphism ψ of $\mathbb{R}$ with $\psi(t_n) = \psi_n(t_n)$ and $\psi(s_n) = \psi_n(s_n)$. It is clear that a is obtained as Schwartzman limit for the parametrization $\tilde{c} \circ \psi$ at parameters s_n, t_n.

The proofs for $\mathcal{C}_+(c)$ and $\mathcal{C}_-(c)$ are similar.

Remark 2. The image of these cluster sets in the projective space $\mathbb{P}H_1(M, \mathbb{R})$ is not necessarily connected: On the torus $M = \mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$, choose a curve in $\mathbb{R}^2$ that oscillates between the half y-axis $\{y > 0\}$ and the half x-axis $\{x > 0\}$, remaining in a small neighborhood of these axes and being unbounded for $t \to +\infty$, and being bounded when $s \to -\infty$. Then its Schwartzman cluster consists of two lines through 0 in $H_1(\mathbb{T}^2, \mathbb{R}) \cong \mathbb{R}^2$, and its projection in the projective space consists of two distinct points.

Remark 3. Let c be a parametrized Schwartzman asymptotic 1-cycle, and consider the unparametrized oriented curve defined by c, denoted by $\bar{c}$. Assume that the asymptotic Schwartzman homology class is $a = [c] \neq 0$. Then

$$\mathcal{C}_\pm(\bar{c}) = \mathcal{C}(\bar{c}) = \mathbb{R}_{\geq 0} \cdot a \,,$$

as a subset of $H_1(M, \mathbb{R})$. This follows since any parametrization of $\bar{c}$ is of the form $c' = c \circ \psi$, where $\psi : \mathbb{R} \to \mathbb{R}$ is a positively oriented homeomorphism of $\mathbb{R}$. Then

$$\frac{c'_{s,t}}{t-s} = \frac{c_{\psi(s),\psi(t)}}{\psi(t) - \psi(s)} \cdot \frac{\psi(t) - \psi(s)}{t-s}. \tag{6}$$

The first term in the right hand side tends to a when $t \to +\infty$, $s \to -\infty$. If the left hand side is to converge, then the second term in the right hand side stays bounded. After extracting a subsequence, it converges to some $\lambda \geq 0$. Hence (6) converges to λa.

We define now the notion of asymptotically homotopic curves.

Definition 14 (*Asymptotic homotopy*). Let $c_0, c_1 : \mathbb{R} \to M$ be two parametrized curves. They are asymptotically homotopic if there exists a continuous family c_u, $u \in [0, 1]$, interpolating between c_0 and c_1, such that

$$c : \mathbb{R} \times [0, 1] \to M, \; c(t, u) = c_u(t),$$

satisfies that $\delta_t(u) = c(t, u)$, $u \in [0, 1]$ is rectifiable with

$$l(\delta_t) = o(|t|). \tag{7}$$

Two oriented curves are asymptotically homotopic if they have orientation preserving parametrizations that are asymptotically homotopic.

Proposition 9. *If c_0 and c_1 are asymptotically homotopic parametrized curves then their cluster sets coincide:*

$$\mathcal{C}_\pm(c_0) = \mathcal{C}_\pm(c_1),$$

$$\mathcal{C}_b(c_0) = \mathcal{C}_b(c_1),$$

$$\mathcal{C}(c_0) = \mathcal{C}(c_1).$$

If c_0 and c_1 are asymptotically homotopic oriented curves then their cluster sets coincide:

$$\mathcal{C}_\pm(c_0) = \mathcal{C}_\pm(c_1),$$
$$\mathcal{C}(c_0) = \mathcal{C}(c_1).$$

Proof. For parametrized curves we have

$$[c_{0,s,t}] = [c_{1,s,t}] + [\delta_s - \gamma_{1,s,t} - \delta_t + \gamma_{0,s,t}].$$

The length of the displacement by the homotopy is bounded by (7), so

$$l(\delta_s - \gamma_{1,s,t} - \delta_t + \gamma_{0,s,t}) = l(\gamma_{1,s,t}) + l(\gamma_{0,s,t}) + o(|t| + |s|),$$

thus

$$\frac{[c_{0,s,t}]}{t-s} = \frac{[c_{1,s,t}]}{t-s} + o(1).$$

An homotopy for non-parametrized curves is an homotopy between two particular parametrizations, but we must require $l(\delta_t) = O(1)$ in place of (7). The homotopy yields a one-to-one correspondence between points in the curves

$$c_0(t) \mapsto c_1(t).$$

Using this correspondence, we have a correspondence of pairs of points $(a, b) = (c_0(s), c_0(t))$ with pairs of points $(a', b') = (c_1(s), c_1(t))$. Thus if the sequence of pairs of points (a_n, b_n) gives a cluster value for c_0, then the corresponding sequence (a'_n, b'_n) gives a proportional cluster value, since (with obvious notation)

$$[c_{0,a_n,b_n}] = [c_{1,a'_n,b'_n}] + O(1).$$

So we can always normalize the speed of the parametrization of c_1 in order to assure that the limit value is the same. This proves that the clusters sets coincide.

5 Calibrating Functions

Let M be a C^∞ smooth compact manifold. We define now the notion of calibrating function.

Let $\pi : \tilde{M} \to M$ be the universal cover of M and let Γ be the group of deck transformations of the cover. Fix a point $\tilde{x}_0 \in \tilde{M}$ and $x_0 = \pi(\tilde{x}_0)$. There is a faithful and transitive action of Γ in the fiber $\pi^{-1}(x_0)$ induced by the action of Γ in $\tilde{M}$, and we have a group isomorphism $\Gamma \cong \pi_1(M, x_0)$. Thus from the group homomorphism

$$\pi_1(M, x_0) \to H_1(M, \mathbb{Z}),$$

we get a group homomorphism

$$\rho : \Gamma \to H_1(M, \mathbb{Z}).$$

Definition 15 (*Calibrating function*). A map $\Phi : \tilde{M} \to H_1(M, \mathbb{R})$ is a calibrating function if the diagram

$$\begin{tikzcd}
\Gamma \cong \pi_1(M, x_0) \arrow[r, hook] \arrow[d, "\rho"'] & \tilde{M} \arrow[d, "\Phi"] \\
H_1(M, \mathbb{Z}) \arrow[r] & H_1(M, \mathbb{R})
\end{tikzcd}$$

is commutative and Φ is equivariant for the action of Γ on $\tilde{M}$, i.e. for any $g \in \Gamma$ and $\tilde{x} \in \tilde{M}$,

$$\Phi(g \cdot \tilde{x}) = \Phi(\tilde{x}) + \rho(g).$$

If $\tilde{x}_0 \in \tilde{M}$ we say that the calibrating function Φ is associated to $\tilde{x}_0$ if $\Phi(\tilde{x}_0) = 0$.

Proposition 10. *There are smooth calibrating functions associated to any point $\tilde{x}_0 \in \tilde{M}$.*

Proof. Fix a smooth non-negative function $\varphi : \tilde{M} \to \mathbb{R}$ with compact support $K = \overline{U}$ with $U = \{\varphi > 0\}$ such that $\pi(U) = M$. Moreover, we can request that $U \cap \pi^{-1}(x_0) = \{\tilde{x}_0\}$.

For any $g_0 \in \Gamma$, define $\varphi_{g_0}(\tilde{x}) = \varphi(g_0^{-1} \cdot \tilde{x})$. The support of φ_{g_0} is $g_0 K$, and $(g_0 K)_{g_0 \in \Gamma}$ is a locally finite covering of $\tilde{M}$, as follows from the compactness of K. Set

$$\psi_{g_0}(\tilde{x}) := \frac{\varphi_{g_0}(\tilde{x})}{\sum_{g \in \Gamma} \varphi_g(\tilde{x})}.$$

Then $\psi_{g_0}(\tilde{x}) = \psi_e(g_0^{-1} \cdot \tilde{x})$ and

$$\sum_{g \in \Gamma} \psi_g \equiv 1.$$

Also ψ_{g_0} has compact support $g_0 K$, and it is a smooth function since the denominator is strictly positive (because $\pi(U) = M$) and it is at each point a finite sum of smooth functions.

We define the map

$$\Phi : \tilde{M} \to H_1(M, \mathbb{R}) ,$$

by

$$\Phi(\tilde{x}) = \sum_{g\in\Gamma} \psi_g(\tilde{x})\,\rho(g).$$

We check that Φ is a calibrating function:

$$\begin{aligned}\Phi(g\cdot\tilde{x}) &= \sum_{h\in\Gamma} \psi_h(g\cdot\tilde{x})\,\rho(h)\\ &= \sum_{h\in\Gamma} \psi_{g^{-1}h}(\tilde{x})\,(\rho(g) + \rho(g^{-1}h))\\ &= \sum_{h'\in\Gamma} \psi_{h'}(\tilde{x})\,\rho(g) \;+\; \sum_{h'\in\Gamma} \psi_{h'}(\tilde{x})\,\rho(h')\\ &= \rho(g) + \Phi(\tilde{x}).\end{aligned}$$

Notice that by construction $\Phi(\tilde{x}_0) = 0$.

We note also that choosing a function ϕ of rapid decay, we may do a similar construction, as long as $\sum_{g\in\Gamma} \phi_g$ is summable (we may need to add a translation to Φ in order to ensure $\Phi(\tilde{x}_0) = 0$).

Observe that the calibrating property implies that for a curve $\gamma : [a,b] \to M$, the quantity $\Phi(\tilde{\gamma}(b)) - \Phi(\tilde{\gamma}(a))$ does not depend on the lift $\tilde{\gamma}$ of γ, because for another choice $\tilde{\gamma}'$, we would have for some $g \in \Gamma$,

$$\tilde{\gamma}'(a) = g\cdot\tilde{\gamma}(a),$$

and

$$\tilde{\gamma}'(b) = g\cdot\tilde{\gamma}(b).$$

Therefore

$$\Phi(\tilde{\gamma}'(b)) - \Phi(\tilde{\gamma}'(a)) = \Phi(g\cdot\tilde{\gamma}(b)) - \Phi(g\cdot\tilde{\gamma}(a)) = \Phi(\tilde{\gamma}(b)) - \Phi(\tilde{\gamma}(a)).$$

This justifies the next definition.

Definition 16. Given a calibrating function Φ, for any curve $\gamma : [a,b] \to M$, we define $\Phi(\gamma) := \Phi(\tilde{\gamma}(b)) - \Phi(\tilde{\gamma}(a))$ for any lift $\tilde{\gamma}$ of γ.

Proposition 11. *For any loop $\gamma \subset M$ we have*

$$\Phi(\gamma) = [\gamma] \in H_1(M, \mathbb{Z}).$$

Proof. Modifying γ, but without changing its endpoints nor $\Phi(\gamma)$ nor $[\gamma]$, we can assume that $x_0 \in \gamma$. Since $\Gamma \cong \pi_1(M, x_0)$, let $h_0 \in \Gamma$ be the element corresponding to γ. Then γ lifts to a curve joining $\tilde{x}_0$ to $h_0 \cdot \tilde{x}_0$, and

$$\Phi(\gamma) = \Phi(h_0 \cdot \tilde{x}_0) - \Phi(\tilde{x}_0) = \rho(h_0) = [\gamma] \in H_1(M, \mathbb{Z}).$$

Proposition 12. *We assume that M is endowed with a Riemannian metric and that the calibrating function Φ is smooth. Then for any rectifiable curve γ we have*

$$|\Phi(\gamma)| \leq C \cdot l(\gamma),$$

where $l(\gamma)$ is the length of γ, and $C > 0$ is a positive constant depending only on the metric.

Proof. The calibrating function Φ is a smooth function on $\tilde{M}$ and Γ-equivariant, hence it is bounded as well as its derivatives. The result follows.

Example 1. For $M = \mathbb{T}$, $\tilde{M} = \mathbb{R}$, $H_1(M, \mathbb{Z}) = \mathbb{Z} \subset \mathbb{R} = H_1(M, \mathbb{R})$, $\Gamma = \mathbb{Z}$ and $\rho : \Gamma \to H_1(M, \mathbb{Z})$ is given (with these identifications) by $\rho(n) = n$. We can take $\varphi(x) = |1 - x|$, for $x \in [-1, 1]$, and $\varphi(x) = 0$ elsewhere. Then

$$\sum_{n=-\infty}^{\infty} \varphi(x - n) = 1,$$

and

$$\psi_n(x) = \varphi_n(x) = \varphi(x - n).$$

Therefore we get the calibrating function

$$\Phi(x) = \sum_{n=-\infty}^{\infty} \varphi(x - n)\, n = x\,.$$

It is a smooth calibrating function (despite that φ is not).

A similar construction works for higher dimensional tori.

Proposition 13. *Let $c : \mathbb{R} \to M$ be a C^1 curve. Consider two sequences (s_n) and (t_n) such that $s_n < t_n$, $s_n \to -\infty$, and $t_n \to +\infty$.*

Then the following conditions are equivalent:

1. *The limit*
$$[c] = \lim_{n \to +\infty} \frac{[c_{s_n, t_n}]}{t_n - s_n} \in H_1(M, \mathbb{R})$$
exists.
2. *The limit*
$$[c]_\Phi = \lim_{n \to \infty} \frac{\Phi(c_{|[s_n, t_n]})}{t_n - s_n} \in H_1(M, \mathbb{R})$$
exists.

3. *For any closed* 1*-form* $\alpha \in \Omega^1(M)$*, the limit*

$$[c](\alpha) = \lim_{n\to\infty} \frac{1}{t_n - s_n} \int_{c([s_n,t_n])} \alpha$$

exists.

4. *For any cohomology class* $[\alpha] \in H^1(M,\mathbb{R})$*, the limit*

$$[c][\alpha] = \lim_{n\to\infty} \frac{1}{t_n - s_n} \int_{c([s_n,t_n])} \alpha$$

exists, and does not depend on the closed 1*-form* $\alpha \in \Omega^1(M)$ *representing the cohomology class.*

5. *For any continuous map* $f : M \to \mathbb{T}$*, let* $\widetilde{f \circ c} : \mathbb{R} \to \mathbb{R}$ *be a lift of* $f \circ c$*, the limit*

$$\rho(f) = \lim_{n\to+\infty} \frac{\widetilde{f \circ c}(t_n) - \widetilde{f \circ c}(s_n)}{t_n - s_n}$$

exists.

6. *For any (two-sided, embedded, transversally oriented) hypersurface* $H \subset M$ *such that all intersections* $c(\mathbb{R}) \cap H$ *are transverse, the limit*

$$[c]\cdot[H] = \lim_{n\to\infty} \frac{\#\{u \in [s_n,t_n]\ ;\ c(u) \in H\}}{t_n - s_n}$$

exists. The notation # means a signed count of intersection points.

When these conditions hold, we have $[c] = [c]_\Phi$ *for any calibrating function* Φ*. If* $\alpha \in \Omega^1(M)$ *is a closed form, then* $[c](\alpha) = [c][\alpha] = \langle [c],[\alpha]\rangle$*. If* $f : M \to \mathbb{T}$ *is a continuous map and* $a = f^*[dx] \in H^1(M,\mathbb{Z})$ *is the pull-back of the generator* $[dx] \in H^1(\mathbb{T},\mathbb{Z})$*, and* H *is a hypersurface such that* $[H]$ *is the Poincaré dual of* a*, then* $\langle [c], a\rangle = \rho(f) = [c]\cdot[H]$*.*

Proof. The equivalence of (1) and (2) follows from the properties of Φ. Let $c : \mathbb{R} \to M$ be a curve. Then

$$\Phi(c_{|[s_n,t_n]}) = \Phi([c_{s_n,t_n}]) - \Phi(\gamma_{s_n,t_n}) = [c_{s_n,t_n}] + O(l(\gamma_{s_n,t_n})).$$

Dividing by $t_n - s_n$ and passing to the limit the equivalence of (1) and (2) follows.

We prove that (1) is equivalent to (3). First note that

$$\left| \int_{\gamma_{s_n,t_n}} \alpha \right| \le C\, l(\gamma_{s_n,t_n})\, ||\alpha||_{C^0}.$$

We have when $t_n - s_n \to +\infty$,

$$\frac{1}{t_n - s_n}\int_{c([s_n,t_n])}\alpha = \frac{1}{t_n - s_n}\int_{c_{s_n,t_n}}\alpha + O\left(\frac{l(\gamma_{s_n,t_n})}{t_n - s_n}\right) = \frac{[c_{s_n,t_n}](\alpha)}{t_n - s_n} + o(1).$$

and the equivalence of (1) and (3) results.

The equivalence of (3) and (4) results from the fact that the limit

$$[c](\alpha) = \lim_{n\to\infty}\frac{1}{t_n - s_n}\int_{c([s_n,t_n])}\alpha$$

does not depend on the representative of the cohomology class $a = [\alpha]$. If $\beta = \alpha + d\phi$, with $\phi : M \to \mathbb{R}$ smooth, then $[c](\alpha) = [c](\beta)$ since

$$[c](d\phi) = \lim_{n\to\infty}\frac{1}{t_n - s_n}\int_{c([s_n,t_n])}d\phi = \lim_{n\to\infty}\frac{\phi(c(t_n)) - \phi(c(s_n))}{t_n - s_n} \to 0,$$

since ϕ is bounded. Also $[c][\alpha] = [c](\alpha)$.

We turn now to (4) implies (5) . First note that there is an identification $H^1(M,\mathbb{Z}) \cong [M, K(\mathbb{Z},1)] = [M,\mathbb{T}]$, where any cohomology class $[\alpha] \in H^1(M,\mathbb{Z})$ is associated to a (homotopy class of a) map $f : M \to \mathbb{T}$ such that $[\alpha] = f^*[\mathbb{T}]$, where $[\mathbb{T}] \in H^1(\mathbb{T},\mathbb{Z})$ is the fundamental class. To prove (5), assume first that f is smooth. With the identification $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, the class $f^*(dx) = df \in \Omega^1(M)$ represents $[\alpha]$. Therefore

$$\begin{aligned}\frac{\widetilde{f \circ c}(t_n) - \widetilde{f \circ c}(s_n)}{t_n - s_n} &= \frac{1}{t_n - s_n}\int_{[s_n,\,t_n]} d(f\circ c) = \\ = \frac{1}{t_n - s_n}\int_{[s_n,\,t_n]}(df)(c') &= \frac{1}{t_n - s_n}\int_{c([s_n,\,t_n])} df,\end{aligned} \tag{8}$$

and from the existence of the limit in (4) we get the limit in (5) that we identify as

$$\rho(f) = [c][df].$$

If f is only continuous, we approximate it by a smooth function, which does not change the limit in (5) .

Conversely, if (5) holds, then any integer cohomology class admits a representative of the form $\alpha = df$, where $f : M \to \mathbb{T}$ is a smooth map. Then using (8) we have

$$\frac{1}{t_n - s_n}\int_{c([s_n,t_n])}\alpha \to \rho(f).$$

So the limit in (4) exists for $\alpha = df$. This implies that the limit in (4) exists for any closed $\alpha \in \Omega^1(M)$, since $H^1(M,\mathbb{Z})$ spans $H^1(M,\mathbb{R})$.

We check the equivalence of (5) and (6). First, let us see that (6) implies (5). As before, it is enough to prove (5) for a smooth map $f : M \to \mathbb{T}$. Let $x_0 \in \mathbb{T}$ be a regular value of f, so that $H = f^{-1}(x_0) \subset M$ is a smooth (two-sided) hypersurface. Then $[H]$ represents the Poincaré dual of $[df] \in H^1(M, \mathbb{Z})$. Choose x_0 such that it is also a regular value of $f \circ c$, so all the intersections of $c(\mathbb{R})$ with H are transverse. Now for any $s < t$,

$$[c_{s,t}] \cdot [H] = \#c([s,t]) \cap H + \#\gamma_{s,t} \cap H,$$

where # denotes signed count of intersection points (we may assume that all intersections of $\gamma_{s,t}$ and H are transverse, by a small perturbation of $\gamma_{s,t}$; also we do not count the extremes of $\gamma_{s,t}$ in $\#\gamma_{s,t} \cap H$ in case that either $c(s) \in H$ or $c(t) \in H$).

Now

$$\#c([s,t]) \cap H = [\widetilde{f \circ c}(t)] + [-\widetilde{f \circ c}(s)] = \widetilde{f \circ c}(t) - \widetilde{f \circ c}(s) + O(1),$$

where $[\cdot]$ denotes the integer part, and $|\#\gamma_{s,t} \cap H|$ is bounded by the total variation of $\widetilde{f \circ \gamma_{s,t}}$, which is bounded by the maximum of df times the total length of $\gamma_{s,t}$, which is $o(t-s)$ by assumption. Hence

$$\lim_{n \to +\infty} \frac{\widetilde{f \circ c}(t_n) - \widetilde{f \circ c}(s_n)}{t_n - s_n} = \lim_{n \to +\infty} \frac{\#c([s_n, t_n]) \cap H}{t_n - s_n}$$

exists.

Conversely, if (5) holds, consider a two-sided embedded topological hypersurface $H \subset M$. Then there is a collar $[0,1] \times H$ embedded in M such that H is identified with $\{\frac{1}{2}\} \times H$. There exists a continuous map $f : M \to \mathbb{T}$ such that $H = f^{-1}(x_0)$ for $x_0 = \frac{1}{2} \in \mathbb{T}$, constructed by sending $[0,1] \times H \to [0,1] \to \mathbb{T}$ and collapsing the complement of $[0,1] \times H$ to 0.

Now if all intersections of $c(\mathbb{R})$ and H are transverse, that means that for any $t \in \mathbb{R}$ such that $c(t) \in H$, we have that $c(t-\epsilon)$ and $c(t+\epsilon)$ are at opposite sides of the collar, for $\epsilon > 0$ small (the sign of the intersection point is given by the direction of the crossing). So $f(c(s))$ crosses x_0 increasingly or decreasingly (according to the sign of the intersection). Hence

$$\frac{\#\{u \in [s_n, t_n] \, ; \, c(u) \in H\}}{t_n - s_n} = \frac{\widetilde{f \circ c}(t_n) - \widetilde{f \circ c}(s_n)}{t_n - s_n} + o(1).$$

The required limit exists.

Remark 4. Proposition 13 holds if we only assume the curve c to be rectifiable.

Corollary 3. *Let $c : \mathbb{R} \to M$ be a C^1 curve. The following conditions are equivalent:*

1. The curve c is a Schwartzman asymptotic cycle.

2. *The limit*

$$\lim_{\substack{t\to+\infty\\ s\to-\infty}} \frac{\Phi(c_{|[s,t]})}{t-s} \in H_1(M,\mathbb{R})$$

 exists.
3. *For any closed* 1*-form* $\alpha \in \Omega^1(M)$*, the limit*

$$\lim_{\substack{t\to+\infty\\ s\to-\infty}} \frac{1}{t-s}\int_{c([s,t])} \alpha$$

 exists.
4. *For any cohomology class* $[\alpha] \in H^1(M,\mathbb{R})$*, the limit*

$$[c][\alpha] = \lim_{\substack{t\to+\infty\\ s\to-\infty}} \frac{1}{t-s}\int_{c([s,t])} \alpha$$

 exists, and does not depend on the closed 1*-form* $\alpha \in \Omega^1(M)$ *representing the cohomology class.*
5. *For any continuous map* $f : M \to \mathbb{T}$*, let* $\widetilde{f\circ c} : \mathbb{R} \to \mathbb{R}$ *be a lift of* $f \circ c$*, we have that the limit*

$$\lim_{\substack{t\to+\infty\\ s\to-\infty}} \frac{\widetilde{f\circ c}(t) - \widetilde{f\circ c}(s)}{t-s}$$

 exists.
6. *For a (two-sided, embedded, transversally oriented) hypersurface* $H \subset M$ *such that all intersections* $c(\mathbb{R}) \cap H$ *are transverse, the limit*

$$\lim_{\substack{t\to+\infty\\ s\to-\infty}} \frac{\#\{u \in [s,t]\,;\, c(u) \in H\}}{t-s}$$

 exists.

When c *is a Schwartzman asymptotic cycle, we have* $[c] = [c]_\Phi$ *for any calibrating function* Φ*. If* $\alpha \in \Omega^1(M)$ *is a closed form then*

$$[c](\alpha) = [c][\alpha] = \langle [\alpha], [c] \rangle.$$

If $f : M \to \mathbb{T}$ *and* $a = f^*[dx] \in H^1(M,\mathbb{Z})$*, where* $[dx] \in H^1(\mathbb{T},\mathbb{Z})$ *is the generator, and* $H \subset M$ *is a hypersurface such that* $[H]$ *is the Poincaré dual of* a*, then we have*

$$\langle [c], [\alpha] \rangle = \rho(f) = [c]\cdot[H].$$

6 Schwartzman 1-Dimensional Cycles

We assume that M is a compact C^∞ Riemannian manifold, with Riemannian metric g.

Definition 17 (*Schwartzman representation of homology classes*). Let $f : S \to M$ be an immersion in M of an oriented 1-solenoid S. Then S is a Riemannian solenoid with the pull-back metric f^*g.

1. If S is endowed with a transversal measure $\mu = (\mu_T) \in \mathcal{M}_\mathcal{T}(S)$, the immersed measured solenoid $f : S_\mu \to M$ represents a homology class $a \in H_1(M, \mathbb{R})$ if for (μ_T)-almost all leaves $c : \mathbb{R} \to S$, parametrized positively and by arc-length, we have that $f \circ c$ is a Schwartzman asymptotic 1-cycle with $[f \circ c] = a$.
2. The immersed solenoid $f : S \to M$ fully represents a homology class $a \in H_1(M, \mathbb{R})$ if for all leaves $c : \mathbb{R} \to S$, parametrized positively and by arc-length, we have that $f \circ c$ is a Schwartzman asymptotic 1-cycle with $[f \circ c] = a$.

Note that if $f : S \to M$ fully represents an homology class $a \in H_1(M, \mathbb{R})$, then for all oriented leaves $c \subset S$, we have that $f \circ c$ is a Schwartzman asymptotic cycle and

$$\mathcal{C}_+(f \circ c) = \mathcal{C}_-(f \circ c) = \mathcal{C}(f \circ c) = \mathbb{R}_{\geq 0} \cdot a \subset H_1(M, \mathbb{R}),$$

by Remark 3.

Observe that contrary to what happens with Ruelle-Sullivan cycles, we can have an immersed solenoid fully representing an homology class without the need of a transversal measure on S.

Definition 18 (*Cluster of an immersed solenoid*). Let $f : S \to M$ be an immersion in M of an oriented 1-solenoid S. The homology cluster of (f, S), denoted by $\mathcal{C}(f, S) \subset H_1(M, \mathbb{R})$, is defined as the derived set of $([(f \circ c)_{s,t}]/(t - s))_{c, t \to \infty, s \to -\infty}$, taken over all images of orientation preserving parametrizations c of all leaves of S, and $t \to +\infty$ and $s \to -\infty$. Analogously, we define the corresponding positive and negative clusters.

The Riemannian cluster of (f, S), denoted by $\mathcal{C}^g(f, S)$, is defined in a similar way, using arc-length orientation preserving parametrizations. Analogously, we define the positive, negative and balanced Riemannian clusters.

As in Sect. 4, we can prove with arguments analogous to those of Propositions 7 and 8 :

Proposition 14. *The homology clusters $\mathcal{C}(f, S)$, $\mathcal{C}_\pm(f, S)$ are non-empty, closed cones of $H_1(M, \mathbb{R})$. If these cones are non-degenerate, their images in $\mathbb{P}H_1(M, \mathbb{R})$ are non-empty and compact sets.*

The Riemannian homology clusters $\mathcal{C}^g(f, S)$, $\mathcal{C}^g_\pm(f, S)$ are non-empty, compact and connected subsets of $H_1(M, \mathbb{R})$.

The following proposition is clear, and gives the relationship with the clusters of the images by f of the leaves of S.

Proposition 15. *Let $f : S \to M$ be an immersion in M of an oriented 1-solenoid S. We have*

$$\bigcup_{c \subset S} \mathcal{C}(f \circ c) \subset \mathcal{C}(f, S),$$

where the union runs over all parametrizations of leaves of S. We also have

$$\bigcup_{c \subset S} \mathcal{C}_{\pm}(f \circ c) \subset \mathcal{C}_{\pm}(f, S),$$

and

$$\bigcup_{c \subset S} \mathcal{C}_{b}(f \circ c) \subset \mathcal{C}_{b}(f, S).$$

And similarly for all Riemannian clusters with $\mathcal{C}_*(f \circ c)$ denoting the Schwartzman clusters for the arc-length parametrization.

We recall that given an immersion $f : S \to M$ of an oriented 1-solenoid, S becomes a Riemannian solenoid and Theorem 3 gives a one-to-one correspondence between the space of transversal measures (up to scalar normalization) and the space of probability daval measures,

$$\overline{\mathcal{M}}_{\mathcal{T}}(S) \cong \mathcal{M}_{\mathcal{L}}(S).$$

Moreover, in the case of 1-solenoids that we consider here, they do satisfy the controlled growth condition of Definition 7. Therefore all Schwartzman measures disintegrate as length on leaves by Theorem 4.

Giving any transversal measure μ we can consider the associated generalized current $(f, S_\mu) \in \mathcal{C}_k(M)$.

Definition 19. We define the Ruelle-Sullivan map

$$\Psi : \mathcal{M}_{\mathcal{T}}(S) \to H_1(M, \mathbb{R})$$

by

$$\mu \mapsto \Psi(\mu) = [f, S_\mu].$$

The Ruelle-Sullivan cluster cone of (f, S) is the image of Ψ

$$\mathcal{C}_{RS}(f, S) = \Psi(\mathcal{M}_{\mathcal{T}}(S)) = \big\{[f, S_\mu]\,;\, \mu \in \mathcal{M}_{\mathcal{T}}(S)\big\} \subset H_1(M, \mathbb{R}).$$

The Ruelle-Sullivan cluster set is

$$\mathbb{P}\mathcal{C}_{RS}(f, S) \cong \big\{[f, S_\mu]\,;\, \mu \in \mathcal{M}_{\mathcal{L}}(S)\big\} \subset H_1(M, \mathbb{R}),$$

i.e. using transversal measures which are normalized (using the Riemannian metric of M).

Proposition 16. *Let $\mathcal{V}_{\mathcal{T}}(S)$ be the set of all signed measures, with finite absolute measure and invariant by holonomy, on the solenoid S. The Ruelle-Sullivan map Ψ extended by linearity to $\mathcal{V}_{\mathcal{T}}(S)$ is a linear continuous operator,*

$$\Psi : \mathcal{V}_{\mathcal{T}}(S) \to H_1(M, \mathbb{R}).$$

Proof. Coming back to the definition of generalized current, it is clear that $\mu \mapsto [f, S_\mu]$ is linear in flow-boxes, therefore globally. It is also continuous because if $\mu_n \to \mu$, then $[f, S_{\mu_n}] \to [f, S_\mu]$ as can be seen in a fixed flow-box covering of S.

Corollary 4. *The Ruelle-Sullivan cluster $\mathcal{C}_{RS}(f, S)$ is a non-empty, convex, compact cone of $H_1(M, \mathbb{R})$. Extremal points of the convex set $\mathcal{C}_{RS}(f, S)$ come from the generalized currents of ergodic measures in $\mathcal{M}_{\mathcal{L}}(S)$.*

Proof. Since $\mathcal{M}_{\mathcal{L}}(S)$ is non-empty, convex and compact set, its image by the continuous linear map Ψ is also a non-empty, convex and compact set. Any extremal point of $\mathcal{C}_{RS}(f, S)$ must have an extremal point of $\mathcal{M}_{\mathcal{L}}(S)$ in its pre-image, and these are the ergodic measures in $\mathcal{M}_{\mathcal{L}}(S)$ (according to the identification of $\mathcal{M}_{\mathcal{L}}(S)$ to $\overline{\mathcal{M}}_{\mathcal{T}}(S)$ and by Proposition 5.10 in [2]).

It is natural to investigate the relation between the Schwartzman cluster and the Ruelle-Sullivan cluster.

Theorem 6. *Let S be a 1-solenoid. For any immersion $f : S \to M$ we have*

$$\bigcup_{c \subset S} \mathcal{C}(f \circ c) \subset \mathcal{C}_{RS}(f, S).$$

Proof. It is enough to prove the theorem for minimal solenoids, since each leaf $c \subset S$ is contained in a minimal solenoid $S_0 \subset S$, and

$$\mathcal{C}(f \circ c) \subset \mathcal{C}_{RS}(f, S_0) \subset \mathcal{C}_{RS}(f, S).$$

The last inclusion holds because if μ is a transversal measure for S_0, then it defines a transversal measure μ' for S, which is clearly invariant by holonomy. Now the generalized currents coincide, $(f, S_{\mu'}) = (f, S_{0,\mu})$, as can be seen in a fixed flow-box covering of S. Therefore, the Ruelle-Sullivan homology classes are the same, $[f, S_{\mu'}] = [f, S_{0,\mu}]$.

The statement for minimal solenoids follows from Theorem 7 below.

Theorem 7. *Let S be a minimal 1-solenoid. For any immersion $f : S \to M$ we have*

$$\mathcal{C}(f, S) \subset \mathcal{C}_{RS}(f, S).$$

Proof. Consider an element $a \in \mathcal{C}(f, S)$ obtained as limit of a sequence $([(f \circ c_n)_{s_n, t_n}])$, where c_n is a positively oriented parametrized leaf of S and $s_n < t_n$, $s_n \to -\infty$, $t_n \to \infty$. The points $(c_n(t_n))$ must accumulate a point $x \in S$, and taking a subsequence, we can assume they converge to it. Choose a small local transversal T of S at this point, such that $f(T) \subset B$ where $B \subset M$ is a contractible ball in M. By minimality, the return map $R_T : T \to T$ is well defined.

Note that we may assume that $\bar{T} \subset T'$, where T' is also a local transversal. By compactness of $\bar{T}$, the return time for $R_{T'} : T' \to T'$ of any leaf, measured with the arc-length parametrization, for any $x \in \bar{T}$, is universally bounded. Therefore we can adjust the sequences (s_n) and (t_n) such that $c_n(s_n) \in \bar{T}$ and $c_n(t_n) \in \bar{T}$, by changing each term by an amount $O(1)$. Now, after further taking a subsequence, we can arrange that $c_n(s_n), c_n(t_n) \in T$.

Taking again a subsequence if necessary we can assume that we have a Schwartzman limit of the measures μ_n which correspond to the arc-length on $c_n([s_n, t_n])$ normalized with total mass 1. The limit measure μ disintegrates on leaves because of Theorem 4, so it defines a transversal measure μ.

The transversal measures corresponding to μ_n are atomic, supported on $T \cap c_n([s_n, t_n])$, assigning equal weights to each point in $T \cap c_n([s_n, t_n])$. The transversal measure corresponding to μ is its normalized limit. For each 1-cohomology class, we may choose a closed 1-form ω representing it and vanishing on B (this is so because $H^1(M, B) = H^1(M)$, since B is contractible). Assume that we have constructed $[(f \circ c_n)_{s_n,t_n}]$ by using γ_{n,s_n,t_n} inside B. So

$$\langle [f, S_{\mu_n}], \omega\rangle = \int_S f^*\omega \, d\mu_n = \int_{f\circ c_n([s_n,t_n])} \omega = \langle [(f \circ c_n)_{s_n,t_n}], [\omega]\rangle,$$

thus

$$\langle [f, S_\mu], [\omega]\rangle = \lim_{n\to\infty} \frac{1}{t_n - s_n} \langle [f, S_{\mu_n}], \omega\rangle = \lim_{n\to\infty} \langle \frac{[(f \circ c_n)_{s_n,t_n}]}{t_n - s_n}, [\omega]\rangle = \langle a, [\omega]\rangle.$$

Thus the generalized current of the limit measure coincides with the Schwartzman limit.

We use the notation $\partial^* C$ for the extremal points of a compact convex set C. For the converse result, we have:

Theorem 8. *Let S be a minimal solenoid and an immersion $f : S \to M$. We have*

$$\partial^* \mathcal{C}_{RS}(f, S) \subset \bigcup_{c\subset S} \mathcal{C}(f \circ c) \subset \mathcal{C}(f, S).$$

Proof. We have seen that the points in $\partial^* \mathcal{C}_{RS}(f, S)$ come from ergodic measures in $\mathcal{M}_{\mathcal{L}}(S)$ by the Ruelle-Sullivan map. Therefore it is enough to prove the following theorem that shows that the Schwartzman cluster of almost all leaves is reduced to the generalized current for an ergodic 1-solenoid.

Theorem 9. *Let S be a minimal 1-solenoid endowed with an ergodic measure $\mu \in \mathcal{M}_{\mathcal{L}}(S)$. Consider an immersion $f : S \to M$. Then for μ-almost all leaves $c \subset S$ we have that $f \circ c$ is a Schwartzman asymptotic 1-cycle and*

$$[f \circ c] = [f, S_\mu] \in H_1(M, \mathbb{R}).$$

Therefore the immersion $f : S_\mu \to M$ represents its Ruelle-Sullivan homology class.

In particular, this homology class is independent of the metric g on M up to a scalar factor.

Proof. The proof is an application of Birkhoff's ergodic theorem. Choose a small local transversal T such that $f(T) \subset B$, where $B \subset M$ is a small contractible ball. Consider the associated Poincaré first return map $R_T : T \to T$. Denote by μ_T the transversal measure supported on T.

For each $x \in T$ we consider $\varphi_T(x)$ to be the homology class in M of the loop image by f of the leaf $[x, R_T(x)]$ closed by a segment in B joining x with $R_T(x)$. In this way we have defined a measurable map

$$\varphi_T : T \to H_1(M, \mathbb{Z}).$$

Also for $x \in S$, we denote by $l_T(x)$ the length of the leaf joining x with its first impact on T (which is $R_T(x)$ for $x \in T$). We have then an upper semi-continuous map

$$l_T : S \to \mathbb{R}_+.$$

Therefore l_T is bounded by compactness of S. In particular, l_T is bounded on T and thus in $L^1(T, \mu_T)$. The boundedness of l_T implies also the boundedness of φ_T by Lemma 1.

Consider $x_0 \in T$ and its return points $x_i = R_T^i(x_0)$. Let $0 < t_1 < t_2 < t_3 < \ldots$ be the times of return for the positive arc-length parametrization. We have

$$t_{i+1} - t_i = l_T(x_i).$$

Therefore

$$t_n = \sum_{i=0}^{n-1} (t_{i+1} - t_i) = \sum_{i=0}^{n-1} l_T \circ R_T^i(x_0),$$

and by Birkhoff's ergodic theorem

$$\lim_{n \to +\infty} \frac{1}{n} t_n = \int_T l_T(x)\, d\mu_T(x) = \mu(S) = 1.$$

Now observe that, by contracting B, we have

$$\begin{aligned}[f \circ c_{0,t_n}] &= [f \circ c_{0,t_1}] + [f \circ c_{t_1,t_2}] + \ldots + [f \circ c_{t_{n-1},t_n}] \\ &= \varphi_T(x_0) + \varphi_T \circ R_T(x_0) + \ldots + \varphi_T \circ R_T^{n-1}(x_0).\end{aligned}$$

We recognize a Birkhoff's sum and by Birkhoff‘s ergodic theorem we get the limit

$$\lim_{n\to+\infty}\frac{1}{n}[f\circ c_{0,t_n}]=\int_T \varphi_T(x)\,d\mu_T(x)\in H_1(M,\mathbb{R}).$$

Finally, putting these results together,

$$\lim_{n\to+\infty}\frac{1}{t_n}[f\circ c_{0,t_n}]=\lim_{n\to+\infty}\frac{[f\circ c_{0,t_n}]/n}{t_n/n}=\frac{\int_T \varphi_T(x)\,d\mu_T(x)}{\int_T l_T(x)\,d\mu_T(x)}=\int_T \varphi_T(x)\,d\mu_T(x).$$

Let us see that this equals the generalized current. Take a closed 1-form $\omega\in\Omega^1(M)$, which we can assume to vanish on B. Then

$$\langle[f,S_\mu],\omega\rangle=\int_T\left(\int_{[x,R_T(x)]}f^*\omega\right)d\mu_T(x)=\int_T\langle\varphi_T(x),\omega\rangle d\mu_T(x),$$

and so

$$[f,S_\mu]=\int_T \varphi_T(x)\,d\mu_T(x).$$

Observe that so far we have only proved that $\mathcal{C}^g_+(f\circ c)=\{[f,S_\mu]\}$ for almost all leaves $c\subset S$. Considering the reverse orientation, the result follows for the negative clusters, and finally for the whole cluster of almost all leaves.

The last statement follows since $[f,S_\mu]$ only depends on $\mu\in\mathcal{M}_{\mathcal{T}}(S)$, which is independent of the metric up to scalar factor, thanks to the isomorphism of Theorem 3.

Therefore for a minimal oriented ergodic 1-solenoid, the generalized current coincides with the Schwartzman asymptotic homology class of almost all leaves. It is natural to ask when this holds for all leaves, i.e. when the solenoid fully represents the generalized current. This indeed happens when the solenoid S is uniquely ergodic (unique ergodicity for a 1-solenoid implies that all orbits are dense and therefore minimality, by Proposition 5.8 in [2]).

Theorem 10. *Let S be a uniquely ergodic oriented 1-solenoid, and let $\mathcal{M}_{\mathcal{L}}(S)=\{\mu\}$. Let $f:S\to M$ be an immersion. Then for each leaf $c\subset S$ we have that $f\circ c$ is a Schwartzman asymptotic cycle with*

$$[f\circ c]=[f,S_\mu]\in H_1(M,\mathbb{R}),$$

and we have

$$\mathcal{C}^g(f\circ c)=\mathcal{C}^g(f,S)=\mathbb{P}\mathcal{C}_{RS}(f,S)=\{[f,S_\mu]\}\subset H_1(M,\mathbb{R}).$$

Therefore $f:S\to M$ fully represents its Ruelle-Sullivan homology class $[f,S_\mu]$.

7 Schwartzman k-Dimensional Cycles

We study in this section how to extend Schwartzman theory to k-dimensional submanifolds of M. We assume that M is a compact C^∞ Riemannian manifold.

Given an immersion $c : N \to M$ from an oriented smooth manifold N of dimension $k \geq 1$, it is natural to consider exhaustions (U_n) of N with $U_n \subset N$ being k-dimensional compact submanifolds with boundary ∂U_n. We close U_n with a k-dimensional oriented manifold Γ_n with boundary $\partial \Gamma_n = -\partial U_n$ (that is, ∂U_n with opposite orientation, so that $N_n = U_n \cup \Gamma_n$ is a k-dimensional compact oriented manifold without boundary), in such a way that $c_{|U_n}$ extends to a piecewise smooth map $c_n : N_n \to M$. We may consider the associated homology class $[c_n(N_n)] \in H_k(M, \mathbb{Z})$. By analogy with Sect. 4, we consider

$$\frac{1}{t_n}[c_n(N_n)] \in H_k(M, \mathbb{R}), \tag{9}$$

for increasing sequences (t_n), $t_n > 0$, and $t_n \to +\infty$, and look for sufficient conditions for (9) to have limits in $H_k(M, \mathbb{R})$. Lemma 1 extends to higher dimension to show that, as long as we keep control of the k-volume of $c_n(\Gamma_n)$, the limit is independent of the closing procedure.

Lemma 2. *Let (Γ_n) be a sequence of closed (i.e. compact without boundary) oriented k-dimensional manifolds with piecewise smooth maps $c_n : \Gamma_n \to M$, and let (t_n) be a sequence with $t_n > 0$ and $t_n \to +\infty$. If*

$$\lim_{n \to +\infty} \frac{\mathrm{Vol}_k(c_n(\Gamma_n))}{t_n} = 0,$$

then in $H_k(M, \mathbb{R})$ we have

$$\lim_{n \to +\infty} \frac{[c_n(\Gamma_n)]}{t_n} = 0.$$

The proof follows the same lines as the proof of Lemma 1. We define now k-dimensional Schwartzman asymptotic cycles.

Definition 20 (*Schwartzman asymptotic k-cycles and clusters*). Let $c : N \to M$ be an immersion from a k-dimensional oriented manifold N into M. For all increasing sequences (t_n), $t_n \to +\infty$, and exhaustions (U_n) of N by k-dimensional compact submanifolds with boundary, we consider all possible Schwartzman limits

$$\lim_{n \to +\infty} \frac{[c_n(N_n)]}{t_n} \in H_k(M, \mathbb{R}),$$

where $N_n = U_n \cup \Gamma_n$ is a closed oriented manifold with

$$\frac{\mathrm{Vol}_k(c_n(\Gamma_n))}{t_n} \to 0. \tag{10}$$

Each such limit is called a Schwartzman asymptotic k-cycle. These limits form the Schwartzman cluster $\mathcal{C}(c, N) \subset H_k(M, \mathbb{R})$ of N.

Observe that a Schwartzman limit does not depend on the choice of the sequence (Γ_n), as long as it satisfies (10). Note that this condition is independent of the particular Riemannian metric chosen for M.

As in dimension 1 we have

Proposition 17. *The Schwartzman cluster $\mathcal{C}(c, N)$ is a closed cone of $H_k(M, \mathbb{R})$.*

The Riemannian structure on M induces a Riemannian structure on N by pulling back by c. We define the Riemannian exhaustions (U_n) of N as exhaustions of the form

$$U_n = \bar{B}(x_0, R_n),$$

i.e. the U_n are Riemannian (closed) balls in N centered at a base point $x_0 \in N$ and $R_n \to +\infty$. If the R_n are generic, then the boundary of U_n is smooth

We define the Riemannian Schwartzman cluster of N as follows. It plays the role of the balanced Riemannian cluster of Sect. 4 for dimension 1.

Definition 21. The Riemann-Schwartzman cluster of $c : N \to M$, $\mathcal{C}^g(c, N)$, is the set of all limits, for all Riemannian exhaustions (U_n),

$$\lim_{n\to+\infty} \frac{1}{\mathrm{Vol}_k(c_n(N_n))}[c_n(N_n)] \in H_k(M, \mathbb{R}),$$

such that $N_n = U_n \cup \Gamma_n$ and

$$\frac{\mathrm{Vol}_k(c_n(\Gamma_n))}{\mathrm{Vol}_k(c_n(N_n))} \to 0. \tag{11}$$

All such limits are called Riemann-Schwartzman asymptotic k-cycles.

Definition 22. The immersed manifold $c : N \to M$ represents a homology class $a \in H_k(M, \mathbb{R})$ if the Riemann-Schwartzman cluster $\mathcal{C}^g(c, N)$ contains only a,

$$\mathcal{C}^g(c, N) = \{a\}.$$

We denote $[c, N] = a$, and call it the Schwartzman homology class of (c, N). We say that (c, N) is a Riemann-Schwartzman asymptotic k-cycle.

Now we can define the notion of representation of homology classes by immersed solenoids extending Definition 17 to higher dimension.

Definition 23 (*Schwartzman representation of homology classes*). Let $f : S \to M$ be an immersion in M of an oriented k-solenoid S. Then S is a Riemannian solenoid with the pull-back metric f^*g.

1. If S is endowed with a transversal measure $\mu = (\mu_T) \in \mathcal{M}_{\mathcal{T}}(S)$, the immersed solenoid $f : S_\mu \to M$ represents a homology class $a \in H_1(M, \mathbb{R})$ if for (μ_T)-almost all leaves $l \subset S$, we have that (f, l) is a Riemann-Schwartzman asymptotic k-cycle with $[f, l] = a$.
2. The immersed solenoid $f : S \to M$ fully represents a homology class $a \in H_1(M, \mathbb{R})$ if for all leaves $l \subset S$, we have that (f, l) is a Riemann-Schwartzman asymptotic k-cycle with $[f, l] = a$.

Definition 24 (*Equivalent exhaustions*). Two exhaustions (U_n) and $(\hat{U}_n)$ are equivalent if

$$\frac{\mathrm{Vol}_k(U_n - \hat{U}_n) + \mathrm{Vol}_k(\hat{U}_n - U_n)}{\mathrm{Vol}_k(U_n)} \to 0.$$

Note that if two exhaustions (U_n) and $(\hat{U}_n)$ are equivalent, then

$$\frac{\mathrm{Vol}_k(\hat{U}_n)}{\mathrm{Vol}_k(U_n)} \to 1.$$

Moreover, if $N_n = U_n \cup \Gamma_n$ are closings satisfying (11), then we may close $\hat{U}_n$ as follows: after slightly modifying $\hat{U}_n$ so that U_n and $\hat{U}_n$ have boundaries intersecting transversally, we glue $F_1 = U_n - \hat{U}_n$ to $\hat{U}_n$ along $F_1 \cap \partial\hat{U}_n$, then we glue a copy of $F_2 = \hat{U}_n - U_n$ (with reversed orientation) to $\hat{U}_n$ along $F_2 \cap \partial\hat{U}_n$. The boundary of $\hat{U}_n \cup F_1 \cup F_2$ is homeomorphic to ∂U_n, so we may glue Γ_n to it, to get $\hat{N}_n = \hat{U}_n \cup F_1 \cup F_2 \cup \Gamma_n$. Note that

$$\mathrm{Vol}_k(\hat{N}_n) = \mathrm{Vol}_k(N_n) + 2\,\mathrm{Vol}_k(\hat{U}_n - U_n) \approx \mathrm{Vol}_k(N_n).$$

Define $\hat{c}_n$ by $\hat{c}_{n|F_1} = c_{|(U_n - \hat{U}_n)}$, $\hat{c}_{n|F_2} = c_{|(\hat{U}_n - U_n)}$ and $\hat{c}_{n|\Gamma_n} = c_{n|\Gamma_n}$. Then

$$[c_n(N_n)] = [\hat{c}_n(\hat{N}_n)]\,,$$

so both exhaustions define the same Schwartzman asymptotic k-cycles.

Definition 25 (*Controlled solenoid*). Let $V \subset S$ be an open subset of a solenoid S. We say that S is controlled by V if for any Riemann exhaustion (U_n) of any leaf of S there is an equivalent exhaustion $(\hat{U}_n)$ such that for all n we have $\partial\hat{U}_n \subset V$.

Definition 26 (*Trapping region*). An open subset $W \subset S$ of a solenoid S is a trapping region if there exists a continuous map $\pi : S \to \mathbb{T}$ such that

1. For some $0 < \epsilon_0 < 1/2$, $W = \pi^{-1}((-\epsilon_0, \epsilon_0))$.
2. There is a global transversal $T \subset \pi^{-1}(\{0\})$.
3. Each connected component of $\pi^{-1}(\{0\})$ intersects T in exactly one point.

4. 0 is a regular value for π, that is, π is smooth in a neighborhood of $\pi^{-1}(\{0\})$ and it $d\pi$ is surjective at each point of $\pi^{-1}(\{0\})$ (the differential $d\pi$ is understood leaf-wise).
5. For each connected component L of $\pi^{-1}(\mathbb{T} - \{0\})$ we have $\overline{L} \cap T = \{x, y\}$, where $\{x\} \in \overline{L} \cap T \cap \pi^{-1}((-\epsilon_0, 0])$ and $\{y\} \in \overline{L} \cap T \cap \pi^{-1}([0, \epsilon_0))$. We define $R_T : T \to T$ by $R_T(x) = y$.

Let C_x be the (unique) component of $\pi^{-1}(\{0\})$ through $x \in T$. By (4), C_x is a smooth $(k-1)$-dimensional manifold. By (5), there is no holonomy in $\pi^{-1}((-\epsilon_0, \epsilon_0))$, so C_x is a compact submanifold. Let L_x be the connected component of $\pi^{-1}(\mathbb{T} - \{0\})$ with $\overline{L}_x \cap T = \{x, y\}$. This is a compact manifold with boundary

$$\partial \overline{L}_x = C_x \cup C_y = C_x \cup C_{R_T(x)}. \tag{12}$$

Proposition 18. *If S has a trapping region W with global transversal T, then holonomy group of T is generated by the map R_T.*

Proof. If γ is a path with endpoints in T, we may homotope it so that each time it traverses $\pi^{-1}(\{0\})$, it does it through T. Then we may split γ into sub-paths such that each path has endpoints in T and no other points in $\pi^{-1}(\{0\})$. Each of this sub-paths therefore lies in some $\overline{L}_x$ and has holonomy R_T, R_T^{-1} or the identity. The result follows.

Theorem 11. *A solenoid S with a trapping region W is controlled by W.*

Proof. Fix a base point $y_0 \in S$ and a exhaustion (U_n) of the leaf l through y_0 of the form $U_n = \bar{B}(y_0, R_n)$, $R_n \to +\infty$. Consider $x_0 \in T$ so that $y_0 \in \overline{L}_{x_0}$. The leaf l is the infinite union

$$l = \bigcup_{n \in \mathbb{Z}} \overline{L}_{R_T^n(x_0)}.$$

If $R_T^n(x_0) = x_0$ for some $n \geq 1$ then l is a compact manifold. Then for some N, we have $U_N = l$, so the controlled condition of Definition 25 is satisfied for l.

Assume that $R_T(x_0) \neq x_0$. Then l is a non-compact manifold. For integers $a < b$, denote

$$\hat{U}_{a,b} := \bigcup_{k=a}^{b-1} \overline{L}_{R_T^k(x_0)}. \tag{13}$$

This is a manifold with boundary

$$\partial \hat{U}_{a,b} = C_{R_T^a(x_0)} \cup C_{R_T^b(x_0)}.$$

Given U_n, pick the maximum $b \geq 1$ and minimum $a \leq 0$ such that $\hat{U}_{a,b} \subset U_n$, and denote $\hat{U}_n = \hat{U}_{a,b}$ for such a and b. Clearly $\partial \hat{U}_n \subset W$. Let us see that (U_n) and $(\hat{U}_n)$ are equivalent exhaustions, i.e. that

$$\frac{\mathrm{Vol}_k(U_n - \hat{U}_n)}{\mathrm{Vol}_k(U_n)} \to 0.$$

Let $b' \geq 1$ the minimum and $a' \leq 0$ the maximum such that $U_n \subset \hat{U}_{a',b'}$. Let us prove that

$$\mathrm{Vol}_k(\hat{U}_{a',b'} - \hat{U}_{a,b})$$

is bounded. This clearly implies the result.

Take $y \in \overline{L}_{R_T^{b'-1}(x_0)} \cap U_n$. Then $d(y_0, y) \leq R_n$. By compactness of T, there is a lower bound $c_0 > 0$ for the distance from C_x to $C_{R_T(x)}$ in L_x, for all $x \in T$. Taking the geodesic path from y_0 to y, we see that there are points in $y_i \in \overline{L}_{R_T^{b'-i}(x_0)}$ with $d(y_0, y_i) \leq R_n - (i-2)\, c_0$, for $2 \leq i \leq b'$.

As $\overline{L}_{R_T^b(x_0)}$ is not totally contained in U_n, we may take $z \in \overline{L}_{R_T^b(x_0)} - U_n$, so $d(y_0, z) > R_n$. Both z and $y_{b'-b}$ are on the same leaf $\overline{L}'_{R_T^b(x_0)}$. By compactness of T, the diameter for a leaf $\overline{L}_x$ is bounded above by some $c_1 > 0$, for all $x \in T$. So

$$R_n - (b' - b - 2)\, c_0 \geq d(y_0, y_{b'-b}) \geq d(y_0, z) - d(y_{b'-b}, z) > R_n - c_1,$$

hence

$$b' - b < \frac{c_1}{c_0} + 2.$$

Analogously,

$$a - a' < \frac{c_1}{c_0} + 2.$$

Again by compactness of T, the k-volumes of $\overline{L}_x$ are uniformly bounded by some $c_2 > 0$, for all $x \in T$. So

$$\mathrm{Vol}_k(\hat{U}_{a',b'} - \hat{U}_{a,b}) \leq (b' - b + a - a')c_2 < 2\left(\frac{c_1}{c_0} + 2\right) c_2\,,$$

concluding the proof.

Theorem 12. *Let S be a minimal oriented k-solenoid endowed with a transversal ergodic measure $\mu \in \mathcal{M}_{\mathcal{L}}(S)$ and with a trapping region $W \subset S$. Consider an immersion $f : S \to M$ such that $f(W)$ is contained in a contractible ball in M. Then $f : S_\mu \to M$ represents its Ruelle-Sullivan homology class $[f, S_\mu]$, i.e. for μ_T-almost all leaves $l \subset S$,*

$$[f, l] = [f, S_\mu] \in H_k(M, \mathbb{R}).$$

If S_μ is uniquely ergodic, then $f : S_\mu \to M$ fully represents its Ruelle-Sullivan homology class.

In particular, this homology class is independent of the metric g on M up to a scalar factor.

Proof. We define a map $\varphi_T : T \to H_k(M, \mathbb{Z})$ as follows: given $x \in T$, consider $f(\overline{L}_x)$. Since $\partial f(\overline{L}_x)$ is contained in a contractible ball B of M, we can close $f(L_x)$ locally as $N_x = f(\overline{L}_x) \cup \Gamma_x$ and define an homology class $\varphi_T(x) = [N_x] \in$

$H_k(M,\mathbb{Z})$. This is independent of the choice of the closing. This map φ_T is measurable and bounded in $H_k(M,\mathbb{Z})$ since the k-volume of Γ_x may be chosen uniformly bounded. Also we can define a map $l_T : T \to \mathbb{R}_+$ by $l_T(x) = \mathrm{Vol}_k(\overline{L}_x)$. It is also a measurable and bounded map.

We have seen that every Riemann exhaustion (U_n) is equivalent to an exhaustion $(\hat{U}_n)$ with $\partial\hat{U}_n \subset W$. Note also that we can saturate the exhaustion $(\hat{U}_n)$ into $(\hat{U}_{n,m})_{n\leq 0\leq m}$, with $\hat{U}_{n,m}$ defined in (13), where $\partial\hat{U}_{n,m} = C_{R_T^n(x_0)} \cup C_{R_T^m(x_0)}$, and $x_0 \in T$ is a base point. Since $f(W)$ is contained in a contractible ball B of M, we can always close $f(\hat{U}_{n,m})$, with a closing inside B, to get $N_{n,m}$ defining an homology class $[N_{n,m}] \in H_k(M,\mathbb{Z})$. Moreover we have

$$[N_{n,m}] = \sum_{i=n}^{m-1} \varphi_T(R_T^i(x_0)).$$

Thus by ergodicity of μ and Birkhoff's ergodic theorem, we have that for μ_T-almost all $x_0 \in T$,

$$\frac{1}{m-n}[N_{n,m}] \to \int_T \varphi_T \, d\mu_T.$$

Also

$$\mathrm{Vol}_k(\hat{U}_{n,m}) = \sum_{i=n}^{m-1} l_T(R_T^i(x_0)),$$

where $\mathrm{Vol}_k(N_{n,m})$ differs from $\mathrm{Vol}_k(\hat{U}_{n,m})$ by a bounded quantity due to the closings. By Birkhoff's ergodic theorem, for μ_T-almost all $x_0 \in T$,

$$\frac{1}{m-n}\,\mathrm{Vol}_k\, f(\hat{U}_{n,m}) \to \int_T l_T \, d\mu_T = \mu(S) = 1.$$

Thus we conclude that for μ_T-almost $x_0 \in T$,

$$\frac{1}{\mathrm{Vol}_k(N_{n,m})}[N_{n,m}] \to \int_T \varphi_T \, d\mu_T,$$

It is easy to see as in Theorem 9 that $\int_T \varphi_T \, d\mu_T$ is the Ruelle-Sullivan homology class $[f, S_\mu]$.

Actually, when $f : S \to M$ is an immersed oriented uniquely ergodic k-solenoid with a trapping region which is mapped to a contractible ball in M, we may prove that $f : S_\mu \to M$ fully represents the Ruelle-Sullivan homology class $[f, S_\mu]$ by checking that the exhaustion $\hat{U}_n$ satisfies the controlled growth condition (see Definition 7) and using Corollary 1 which guarantees that the normalized measures μ_n supported on $\hat{U}_n$ converge to the unique Schwartzman limit μ.

A Appendix: Norm on the Homology

Let M be a compact C^∞ Riemannian manifold. For each $a \in H_1(M, \mathbb{Z})$ we define

$$l(a) = \inf_{[\gamma]=a} l(\gamma),$$

where γ runs over all closed loops in M with homology class a and $l(\gamma)$ is the length of γ,

$$l(\gamma) = \int_\gamma ds_g .$$

By application of Ascoli-Arzela it is classical to get

Proposition 19. *For each $a \in H_1(M, \mathbb{Z})$ there exists a minimizing geodesic loop γ with $[\gamma] = a$ such that*

$$l(\gamma) = l(a).$$

Note that the minimizing property implies the geodesic character of the loop. We also have

Proposition 20. *There exists a universal constant $C_0 = C_0(M) > 0$ only depending on M, such that for $a, b \in H_1(M, \mathbb{Z})$ and $n \in \mathbb{Z}$, we have*

$$l(n \cdot a) \leq |n|\, l(a),$$

and

$$l(a + b) \leq l(a) + l(b) + C_0 .$$

(We can take for C_0 twice the diameter of M.)

Proof. Given a loop γ, the loop $n\gamma$ obtained from γ running through it n times (in the direction compatible the sign of n) satisfies

$$[n\gamma] = n\, [\gamma],$$

and

$$l(n\gamma) = |n|\, l(\gamma).$$

Therefore

$$l(n \cdot a) \leq l(n\gamma) = |n|\, l(\gamma),$$

and we get the first inequality taking the infimum over γ.

Let C_0 be twice the diameter of M. Any two points of M can be joined by an arc of length smaller than or equal to $C_0/2$. Given two loops α and β with $[\alpha] = a$ and $[\beta] = b$, we can construct a loop γ with $[\gamma] = a + b$ by picking a point in α and another point in β and joining them by a minimizing arc which pastes together α and β running through it back and forth. This new loop satisfies

$$l(\gamma) = l(\alpha) + l(\beta) + C_0 ,$$

therefore

$$l(a+b) \leq l(\alpha) + l(\beta) + C_0.$$

and the second inequality follows.

Remark 5. It is not true that $l(n \cdot a) = n\, l(\gamma)$ if $l(a) = l(\gamma)$. To see this take a surface M of genus $g \geq 2$ and two elements $e_1, e_2 \in H_1(M, \mathbb{Z})$ such that

$$l(e_1) + l(e_2) < l(e_1 + e_2).$$

(For instance we can take M to be the connected sum of a large sphere with two small 2-tori at antipodal points, and let e_1, e_2 be simple closed curves, non-trivial in homology, inside each of the two tori.) Let $a = e_1 + e_2$. Then

$$l(n \cdot a) = l(n \cdot (e_1 + e_2)) \leq n\, l(e_1) + n\, l(e_2) + C_0,$$

we get for n large

$$l(n \cdot a) < n\, l(a).$$

Theorem 13 (*Norm in homology*). *Let $a \in H_1(M, \mathbb{Z})$. The limit*

$$||a|| = \lim_{n \to +\infty} \frac{l(n \cdot a)}{n},$$

exists and is finite. It satisfies the properties

(i) For $a \in H_1(M, \mathbb{Z})$, we have $||a|| = 0$ if and only if a is torsion.
(i) For $a \in H_1(M, \mathbb{Z})$ and $n \in \mathbb{Z}$, we have $||n \cdot a|| = |n|\, ||a||$.
(iii) For $a, b \in H_1(M, \mathbb{Z})$, we have

$$||a + b|| \leq ||a|| + ||b||.$$

(iv) $||a|| \leq l(a)$.

Proof. Let $u_n = l(n \cdot a) + C_0$. By the properties proved before, the sequence (u_n) is sub-additive

$$u_{n+m} \leq u_n + u_m,$$

therefore

$$\limsup_{n \to +\infty} \frac{u_n}{n} = \liminf_{n \to +\infty} \frac{u_n}{n}.$$

Moreover, we have also

$$\frac{u_n}{n} \leq l(a) < +\infty,$$

thus the limit exists and is finite. Property (iv) holds.

Property (ii) follows from

$$||n \cdot a|| = \lim_{m \to \infty} \frac{l(mn \cdot a)}{m} = |n| \lim_{m \to \infty} \frac{l(m|n| \cdot a)}{m|n|} = |n|\, ||a||.$$

Property (iii) follows from

$$l(n \cdot (a+b)) \leq l(n \cdot a) + l(n \cdot b) + C_0 \leq n\, l(a) + n\, l(b) + C_0,$$

dividing by n and passing to the limit.

Let us check property (i). If a is torsion then $n \cdot a = 0$, so $||a|| = \frac{1}{n}||n \cdot a|| = 0$. If a is not torsion, then there exists a smooth map $\phi : M \to S^1$ which corresponds to an element $[\phi] \in H^1(M, \mathbb{Z})$ with $m = \langle [\phi], a \rangle > 0$. Then for any loop $\gamma : [0,1] \to M$ representing $n \cdot a$, $n > 0$, we take $\phi \circ \gamma$ and lift it to a map $\tilde{\gamma} : [0,1] \to \mathbb{R}$. Thus

$$\tilde{\gamma}(1) - \tilde{\gamma}(0) = \langle [\phi], n \cdot a \rangle = m\, n.$$

Now let C be an upper bound for $|d\phi|$. Then

$$m\, n = |\tilde{\gamma}(1) - \tilde{\gamma}(0)| = l(\phi \circ \gamma) \leq C\, l(\gamma),$$

so $l(\gamma) \geq m\, n/C$, hence $l(n \cdot a) \geq m\, n/C$ and $||a|| \geq m/C$.

Now we can define a norm in $H_1(M, \mathbb{Q}) = \mathbb{Q} \otimes H_1(M, \mathbb{Z})$ by

$$||\lambda \otimes a|| = |\lambda| \cdot ||a||,$$

and extend it by continuity to $H_1(M, \mathbb{R}) = \mathbb{R} \otimes H_1(M, \mathbb{Z})$.

References

1. V. Muñoz, R. Pérez-Marco, Ergodic solenoidal homology II: density of ergodic solenoids. Aust. J. Math. Anal. Appl. **6**(1), Article 11, 1–8 (2009)
2. V. Muñoz, R. Pérez-Marco, Ergodic solenoids and generalized currents. Revista Matemática Complutense **24**, 493–525 (2011)
3. V. Muñoz, R. Perez-Marco, Intersection theory for ergodic solenoids (2009). Arxiv preprint arXiv:0910.2915
4. V. Muñoz, R. Pérez-Marco, Ergodic solenoidal homology: realization theorem. Commun. Math. Phys. **302**, 737–753 (2011)
5. V. Munoz, R. P. Marco, Hodge theory for Riemannian solenoids. Funct. Equ. Math. Anal. 633–657 (2012). Springer
6. D. Ruelle, D. Sullivan, Currents, flows and diffeomorphisms. Topology **14**, 319–327 (1975)
7. S. Schwartzman, Asymptotic cycles. Ann. Math. **66**(2), 270–284 (1957)
8. R. Thom, Sous-variétés et classes d'homologie des variétés différentiables. I et II. C. R. Acad. Sci. Paris **236**, 453–454, 573–575 (1953)

An Additive Functional Equation in Orthogonality Spaces

Choonkil Park and Themistocles M. Rassias

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract By applying the fixed point method as well as the direct method, we provide a proof of the Hyers-Ulam stability of linear mappings, isometric linear mappings and 2-isometric linear mappings in Banach modules over a unital C^*-algebra and in non-Archimedean Banach modules over a unital C^*-algebra associated with an orthogonally additive functional equation. Moreover, we prove the Hyers-Ulam stability of homomorphisms in C^*-algebras associated with an orthogonally additive functional equation.

1 Introduction and Preliminaries

Assume that X is a real inner product space and $f : X \to \mathbb{R}$ is a solution of the orthogonally Cauchy functional equation $f(x + y) = f(x) + f(y)$ for $x, y \in X$ with $\langle x, y \rangle = 0$. By the Pythagorean theorem $f(x) = \|x\|^2$ is a solution of the conditional equation. Of course, this function does not satisfy the additivity equation everywhere. Thus the orthogonally Cauchy equation is not equivalent to the classic Cauchy equation on the whole inner product space.

C. Park (✉)
Department of Mathematics, Research Institute for Natural Sciences, Hanyang University,
Seoul 133-791, South Korea
e-mail: baak@hanyang.ac.kr

T.M. Rassias
Department of Mathematics, National Technical University of Athens, Zografou Campus,
15780 Athens, Greece
e-mail: trassias@math.ntua.gr

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_14, © Springer-Verlag Berlin Heidelberg 2012

Pinsker [65] characterized orthogonally additive functionals on an inner product space when the orthogonality is the ordinary one in such spaces. Sundaresan [78] generalized this result to arbitrary Banach spaces equipped with the Birkhoff-James orthogonality. The orthogonal Cauchy functional equation

$$f(x+y) = f(x) + f(y), \qquad x \perp y,$$

in which $\perp$ is an abstract orthogonality relation, was first investigated by Gudder and Strawther [27]. They defined $\perp$ by a system consisting of five axioms and described the general semi-continuous real-valued solution of conditional Cauchy functional equation. In 1985, Rätz [75] introduced a new definition of orthogonality by using more restrictive axioms than those of Gudder and Strawther. Moreover, he investigated the structure of orthogonally additive mappings. Rätz and Szabó [76] investigated the problem in a rather more general framework.

Let us recall the orthogonality in the sense of Rätz; cf. [75].

Suppose X is a real vector space (algebraic module) with $\dim X \geq 2$ and $\perp$ is a binary relation on X with the following properties:

(O_1) totality of $\perp$ for zero: $x \perp 0, 0 \perp x$ for all $x \in X$;
(O_2) independence: if $x, y \in X - \{0\}, x \perp y$, then x, y are linearly independent;
(O_3) homogeneity: if $x, y \in X, x \perp y$, then $\alpha x \perp \beta y$ for all $\alpha, \beta \in \mathbb{R}$;
(O_4) the Thalesian property: if P is a 2-dimensional subspace of $X, x \in P$ and $\lambda \in \mathbb{R}_+$, which is the set of nonnegative real numbers, then there exists $y_0 \in P$ such that $x \perp y_0$ and $x + y_0 \perp \lambda x - y_0$.

The pair $(X, \perp)$ is called an orthogonality space (module). By an orthogonality normed space (normed module) we mean an orthogonality space (module) having a normed (normed module) structure.

Some interesting examples are the following:

1. The trivial orthogonality on a vector space X defined by (O_1), and for non-zero elements $x, y \in X$, $x \perp y$ if and only if x, y are linearly independent.
2. The ordinary orthogonality on an inner product space $(X, \langle ., . \rangle)$ given by $x \perp y$ if and only if $\langle x, y \rangle = 0$.
3. The Birkhoff-James orthogonality on a normed space $(X, \|.\|)$ defined by $x \perp y$ if and only if $\|x + \lambda y\| \geq \|x\|$ for all $\lambda \in \mathbb{R}$.

The relation $\perp$ is called symmetric if $x \perp y$ implies that $y \perp x$ for all $x, y \in X$. Clearly examples 1 and 2 are symmetric but example 3 is not. It is remarkable to note, however, that a real normed space of dimension greater than 2 is an inner product space if and only if the Birkhoff-James orthogonality is symmetric. There are several orthogonality notions on a real normed space such as Birkhoff-James, Boussouis, Singer, Carlsson, unitary-Boussouis, Roberts, Phythagorean, isosceles and Diminnie (see [2, 3, 5, 10, 20, 32, 33, 55]).

Assume that if A is a C^*-algebra and X is a module over A and if $x, y \in X, x \perp y$, then $ax \perp by$ for all $a, b \in A$. Then $(X, ||.||)$ is called an *orthogonality module over* A.

The stability problem of functional equations originated from the following question of Ulam [80]: *Under what condition does there exist an additive mapping near an approximately additive mapping?* In 1941, Hyers [29] gave a partial affirmative answer to the question of Ulam in the context of Banach spaces. In 1978, Rassias [67] extended the theorem of Hyers by considering the unbounded Cauchy difference $\|f(x+y)-f(x)-f(y)\| \leq \varepsilon(\|x\|^p+\|y\|^p)$, $(\varepsilon > 0, p \in [0,1))$. The result of Rassias has provided a lot of influence in the development of what we now call *generalized Hyers-Ulam stability* or *Hyers-Ulam-Rassias stability* of functional equations. During the last decades several stability problems of functional equations have been investigated in the spirit of Hyers-Ulam-Rassias. The reader is referred to [17, 30, 35, 73] and references therein for a detailed information on stability properties of functional equations.

Ger and Sikorska [24] investigated the orthogonal stability of the Cauchy functional equation $f(x+y) = f(x)+f(y)$, namely, they showed that if f is a mapping from an orthogonality space X into a real Banach space Y and

$$\|f(x+y)-f(x)-f(y)\| \leq \varepsilon$$

for all $x, y \in X$ with $x \perp y$ and some $\varepsilon > 0$, then there exists exactly one orthogonally additive mapping $g : X \to Y$ such that $\|f(x)-g(x)\| \leq \frac{16}{3}\varepsilon$ for all $x \in X$.

The first author treating the stability of the quadratic equation was Skof [77] by proving that if f is a mapping from a normed space X into a Banach space Y satisfying $\|f(x+y)+f(x-y)-2f(x)-2f(y)\| \leq \varepsilon$ for some $\varepsilon > 0$, then there is a unique quadratic mapping $g : X \to Y$ such that $\|f(x)-g(x)\| \leq \frac{\varepsilon}{2}$. Cholewa [12] extended the Skof's theorem by replacing X by an abelian group G. The Skof's result was later generalized by Czerwik [15] in the spirit of Hyers-Ulam-Rassias. The stability problem of functional equations has been extensively investigated by several mathematicians (see [16, 34, 58, 69–71]).

The orthogonally quadratic equation

$$f(x+y)+f(x-y) = 2f(x)+2f(y), \ x \perp y$$

was first investigated by Vajzović [81] when X is a Hilbert space, Y is the scalar field, f is continuous and $\perp$ means the Hilbert space orthogonality. Later, Drljević [22], Fochi [23], Moslehian [50, 51], Moslehian and Rassias [52] and Szabó [79] provided a generalization of this result.

In 1897, Hensel [28] introduced a normed space which does not satisfy the Archimedean property. It turned out that non-Archimedean spaces have many nice applications (see [18, 43, 44, 54]).

Definition 1. By a non-Archimedean field we mean a field $\mathbb{K}$ equipped with a function (valuation) $|\cdot| : \mathbb{K} \to [0,\infty)$ such that for all $r, s \in \mathbb{K}$, the following conditions hold:

(1) $|r| = 0$ if and only if $r = 0$;
(2) $|rs| = |r||s|$;
(3) $|r + s| \leq max\{|r|, |s|\}$.

Definition 2 ([53]). Let X be a vector space over a scalar field $\mathbb{K}$ with a non-Archimedean non-trivial valuation $|\cdot|$. A function $||\cdot|| : X \to \mathbb{R}$ is a non-Archimedean norm (valuation) if it satisfies the following conditions:

(1) $||x|| = 0$ if and only if $x = 0$;
(2) $||rx|| = |r|||x||$ $(r \in \mathbb{K}, x \in X)$;
(3) The strong triangle inequality (ultrametric); namely,

$$||x + y|| \leq max\{||x||, ||y||\}, \qquad x, y \in X.$$

Then $(X, ||.||)$ is called a non-Archimedean space.

Assume that if A is a C^*-algebra and X is a module over A, which is a non-Archimedean space, and if $x, y \in X, x \perp y$, then $ax \perp by$ for all $a, b \in A$. Then $(X, ||.||)$ is called an *orthogonality non-Archimedean module over* A.

It follows from (3) of Definition 2 that

$$||x_n - x_m|| \leq max\{||x_{j+1} - x_j|| : m \leq j \leq n - 1\} \qquad (n > m).$$

Definition 3. A sequence $\{x_n\}$ is a *Cauchy sequence* if and only if $\{x_{n+1} - x_n\}$ converges to zero in a non-Archimedean space. By a complete non-Archimedean space we mean one in which every Cauchy sequence is convergent.

Let X be a set. A function $d : X \times X \to [0, \infty]$ is called a *generalized metric* on X if d satisfies

1. $d(x, y) = 0$ if and only if $x = y$;
2. $d(x, y) = d(y, x)$ for all $x, y \in X$;
3. $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$.

We recall a fundamental result in fixed point theory.

Theorem 1 ([7, 19]) *Let (X, d) be a complete generalized metric space and let $J : X \to X$ be a strictly contractive mapping with Lipschitz constant $\alpha < 1$. Then for each given element $x \in X$, either*

$$d(J^n x, J^{n+1} x) = \infty$$

for all nonnegative integers n or there exists a positive integer n_0 such that

1. $d(J^n x, J^{n+1} x) < \infty, \qquad \forall n \geq n_0$;
2. the sequence $\{J^n x\}$ converges to a fixed point y^* of J;
3. y^* is the unique fixed point of J in the set $Y = \{y \in X \mid d(J^{n_0} x, y) < \infty\}$;
4. $d(y, y^*) \leq \frac{1}{1-\alpha} d(y, Jy)$ for all $y \in Y$.

In 1996, Isac and Rassias [31] provided applications of stability theory of functional equations for the proof of new fixed point theorems. By using fixed point methods, the stability problems of several functional equations have been extensively investigated by a number of authors (see [8, 9, 37–39, 49, 56, 57, 66]).

Let X and Y be metric spaces. A mapping $f : X \to Y$ is called an isometry if f satisfies

$$d_Y\big(f(x), f(y)\big) = d_X(x, y)$$

for all $x, y \in X$, where $d_X(\cdot,\cdot)$ and $d_Y(\cdot,\cdot)$ denote the metrics in the spaces X and Y, respectively. For some fixed number $r > 0$, suppose that f preserves distance r, i.e., for all x, y in X with $d_X(x, y) = r$, we have $d_Y\big(f(x), f(y)\big) = r$. Then r is called a conservative (or preserved) distance for the mapping f. Aleksandrov [1] posed the following problem:

Remark 1 (Aleksandrov problem). Examine whether the existence of a single conservative distance for some mapping T implies that T is an isometry.

The Aleksandrov problem has been investigated by several mathematicians (see [4, 6, 13, 14, 21, 25, 26, 36, 40, 41, 45–47, 68, 72, 82]). Rassias and Šemrl [74] proved the following theorem for mappings satisfying the strong distance one preserving property (SDOPP), i.e., for every $x, y \in X$ with $\|x - y\| = 1$ it follows that $\|f(x) - f(y)\| = 1$ and conversely.

Theorem 2 ([74]) *Let X and Y be real normed linear spaces such that one of them has dimension greater than one. Suppose that $f : X \to Y$ is a Lipschitz mapping with Lipschitz constant $\kappa \le 1$. Assume that f is a surjective mapping satisfying* (SDOPP). *Then f is an isometry.*

Definition 4 ([11]). Let X be a real linear space with $\dim X \ge N$ and $\|\cdot, \cdots, \cdot\| : X^N \to \mathbb{R}$ a function. Then $(X, \|\cdot, \cdots, \cdot\|)$ is called a *linear N-normed space* if

(N$_1$) $\|x_1, \cdots, x_N\| = 0 \Longleftrightarrow x_1, \cdots, x_N$ are linearly dependent

(N$_2$) $\|x_1, \cdots, x_N\| = \|x_{j_1}, \cdots, x_{j_N}\|$ for every permutation $(j_1, \cdots j_N)$ of $(1, \cdots, N)$

(N$_3$) $\|\alpha x_1, \cdots, x_N\| = |\alpha|\|x_1, \cdots, x_N\|$

(N$_4$) $\|x + y, x_2, \cdots, x_N\| \le \|x, x_2, \cdots, x_n\| + \|y, x_2, \cdots, x_N\|$

for all $\alpha \in \mathbb{R}$ and all $x, y, x_1, \cdots, x_N \in X$. The function $\|\cdot, \cdots, \cdot\|$ is called the *N-norm on X*.

Note that the notion of *1-norm* is the same as that of *norm*.

In [59], it was defined the notion of N-isometry and proved the Rassias and Šemrl's theorem in linear N-normed spaces.

Definition 5 ([59]). We call $f : X \to Y$ an *N-Lipschitz mapping* if there is a $\kappa \ge 0$ such that

$$\|f(x_1)-f(y_1),\cdots,f(x_N)-f(y_N)\| \le \kappa\|x_1-y_1,\cdots,x_N-y_N\|$$

for all $x_1,\cdots,x_N,y_1,\cdots,y_N \in X$. The smallest such κ is called the *N-Lipschitz constant*.

Definition 6 ([59]). Let X and Y be linear N-normed spaces and $f : X \to Y$ a mapping. We call f an *N-isometry* if

$$\|x_1-y_1,\cdots,x_N-y_N\| = \|f(x_1)-f(y_1),\cdots,f(x_N)-f(y_N)\|$$

for all $x_1,\cdots,x_N,y_1,\cdots,y_N \in X$.

Park and Rassias [60–64] investigated the Hyers-Ulam stability problems for N-isometric linear mappings in linear N-normed Banach modules over a C^*-algebra and for isometric linear mappings in Banach modules over a C^*-algebra.

This paper is organized as follows: In Sect. 2, we prove the Hyers-Ulam stability of the following orthogonally additive functional equation

$$2f\left(\frac{x}{2}+y\right) = f(x)+f(2y), \qquad x \perp y, \tag{1}$$

in Banach modules over a unital C^*-algebra by using the fixed point method. In Sect. 3, we prove the Hyers-Ulam stability of the orthogonally additive functional equation (1) in non-Archimedean Banach modules over a unital C^*-algebra by using the fixed point method. In Sect. 4, we prove the Hyers-Ulam stability of the orthogonally additive functional equation (1) in Banach modules over a unital C^*-algebra by using the direct method. In Sect. 5, we prove the Hyers-Ulam stability of homomorphisms in unital C^*-algebras associated with the orthogonally additive functional equation (1) by using the fixed point method. In Sect. 6, we prove the Hyers-Ulam stability of homomorphisms in unital C^*-algebras associated with the orthogonally additive functional equation (1) by using the direct method. In Sect. 7, we prove the Hyers-Ulam stability of isometric linear mappings in Banach modules over a unital C^*-algebra associated with the orthogonally additive functional equation (1) by using the fixed point method. In Sect. 8, we prove the Hyers-Ulam stability of isometric linear mappings in non-Archimedean Banach modules over a unital C^*-algebra associated with the orthogonally additive functional equation (1) by using the direct method. In Sect. 9, we prove the Hyers-Ulam stability of isometric linear mappings in Banach modules over a unital C^*-algebra associated with the orthogonally additive functional equation (1) by using the direct method. In Sect. 10, we prove the Hyers-Ulam stability of 2-isometric linear mappings in Banach modules over a unital C^*-algebra associated with the orthogonally additive functional equation (1) by using the fixed point method. In Sect. 11, we prove the Hyers-Ulam stability of 2-isometric linear mappings in non-Archimedean Banach modules over a unital C^*-algebra associated with the orthogonally additive functional equation (1) by using the fixed point method. In Sect. 12, we prove the Hyers-Ulam stability of 2-isometric linear mappings in Banach modules over a unital C^*-algebra associated with the orthogonally additive functional equation (1) by using the direct method.

Furthermore, applying some ideas from [24, 30], we deal with the stability problem for the orthogonally additive functional equation (1).

2 Stability of the Orthogonally Additive Functional Equation (1) in Banach Modules Over a C^*-Algebra: Fixed Point Method

Throughout this section, assume that A is a unital C^*-algebra with unit e and unitary group $U(A) := \{u \in A \mid u^*u = uu^* = e\}$, $(X, \perp)$ is an orthogonality normed module over A and $(Y, \|.\|_Y)$ is a Banach module over A.

Definition 7. An orthogonally additive mapping $f : X \to Y$ is called an *orthogonally additive A-linear mapping* if $f(ax) = af(x)$ for all $a \in A$ and all $x \in X$.

Lemma 1. *Let V and W be vector spaces. A mapping $f : V \to W$ satisfies $f(0) = 0$ and*

$$2f\left(\frac{x}{2} + y\right) = f(x) + f(2y)$$

for all $x, y \in V$ if and only if the mapping $f : V \to W$ is Cauchy additive, i.e.,

$$f(x + y) = f(x) + f(y)$$

for all $x, y \in V$.

Proof. The proof is obvious.

Theorem 3 *Let $\varphi : X \times X \to [0, \infty)$ be a function such that there exists an $\alpha < 1$ with*

$$\varphi(x, y) \leq 2\alpha\varphi\left(\frac{x}{2}, \frac{y}{2}\right) \tag{2}$$

for all $x, y \in X$ with $x \perp y$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$ and

$$\left\|2uf\left(\frac{x}{2} + y\right) - f(ux) - f(2uy)\right\|_Y \leq \varphi(x, y) \tag{3}$$

for all $u \in U(A)$ and all $x, y \in X$ with $x \perp y$. If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ such that

$$\|f(x) - L(x)\|_Y \leq \frac{\alpha}{1-\alpha}\varphi(x, 0) \tag{4}$$

for all $x \in X$.

Proof. Putting $y = 0$ and $u = e$ in (3), we get

$$\left\|2f\left(\frac{x}{2}\right) - f(x)\right\|_Y \leq \varphi(x, 0) \tag{5}$$

for all $x \in X$, since $x \perp 0$. So

$$\left\|f(x) - \frac{1}{2}f(2x)\right\|_Y \leq \frac{1}{2}\varphi(2x, 0) \leq \alpha \cdot \varphi(x, 0) \tag{6}$$

for all $x \in X$.

Consider the set

$$S := \{h : X \to Y\}$$

and introduce the generalized metric on S:

$$d(g, h) = \inf\{\mu \in \mathbb{R}_+ : \|g(x) - h(x)\|_Y \leq \mu\varphi\,(x, 0)\,, \ \ \forall x \in X\},$$

where, as usual, $\inf\varphi = +\infty$. It is easy to show that the space (S, d) is complete (see [38, Theorem 3.1] or [48, Lemma 2.1]).

Now we consider the linear mapping $J : S \to S$ such that

$$Jg(x) := \frac{1}{2}g\,(2x)$$

for all $x \in X$.

Let $g, h \in S$ be given such that $d(g, h) = \varepsilon$. Then

$$\|g(x) - h(x)\|_Y \leq \varepsilon\varphi\,(x, 0)$$

for all $x \in X$. Hence

$$\|Jg(x) - Jh(x)\|_Y = \left\|\frac{1}{2}g\,(2x) - \frac{1}{2}h\,(2x)\right\|_Y \leq \alpha\varepsilon\varphi\,(x, 0)$$

for all $x \in X$. So $d(g, h) = \varepsilon$ implies that $d(Jg, Jh) \leq \alpha\varepsilon$. This means that

$$d(Jg, Jh) \leq \alpha d(g, h)$$

for all $g, h \in S$.

It follows from (6) that $d(f, Jf) \leq \alpha$.

By Theorem 1, there exists a mapping $L : X \to Y$ satisfying the following:

1. L is a fixed point of J, i.e.,

$$L\,(2x) = 2L(x) \tag{7}$$

for all $x \in X$. The mapping L is a unique fixed point of J in the set

$$M = \{g \in S : d(f, g) < \infty\}.$$

This implies that L is a unique mapping satisfying (7) such that there exists a $\mu \in (0, \infty)$ for which

$$\| f(x) - L(x)\|_Y \leq \mu\varphi (x, 0)$$

for all $x \in X$;

2. $d(J^n f, L) \to 0$ as $n \to \infty$. This implies the equality

$$\lim_{n\to\infty} \frac{1}{2^n} f\ (2^n x) = L(x)$$

for all $x \in X$;

3. $d(f, L) \leq \frac{1}{1-\alpha} d(f, Jf)$, which yields the inequality

$$d(f, L) \leq \frac{\alpha}{1-\alpha}.$$

Thus (4) holds true.

Let $u = e$ in (3). It follows from (2) and (3) that

$$\begin{aligned}
&\left\|2L\left(\frac{x}{2} + y\right) - L(x) - L(2y)\right\|_Y \\
&\quad = \lim_{n\to\infty} \frac{1}{2^n} \|2f(2^{n-1}x + 2^n y) - f(2^n x) - f(2^{n+1} y)\|_Y \\
&\quad \leq \lim_{n\to\infty} \frac{1}{2^n}\varphi(2^n x, 2^n y) \leq \lim_{n\to\infty} \frac{2^n\alpha^n}{2^n}\varphi(x, y) = 0
\end{aligned}$$

for all $x, y \in X$ with $x \perp y$. So

$$2L\left(\frac{x}{2} + y\right) - L(x) - L(2y) = 0$$

for all $x, y \in X$ with $x \perp y$. Hence $L : X \to Y$ is an orthogonally additive mapping.

Let $y = 0$ in (3). It follows from (2) and (3) that

$$\begin{aligned}
\left\|2uL\left(\frac{x}{2}\right) - L(ux)\right\|_Y &= \lim_{n\to\infty} \frac{1}{2^n} \|2uf(2^{n-1}x) - f(2^n ux)\|_Y \\
&\leq \lim_{n\to\infty} \frac{1}{2^n}\varphi(2^n x, 0) \leq \lim_{n\to\infty} \frac{2^n\alpha^n}{2^n}\varphi(x, 0) = 0
\end{aligned}$$

for all $x \in X$. Therefore,

$$2uL\left(\frac{x}{2}\right) - L(ux) = 0$$

for all $x \in X$. Hence

$$L(ux) = 2uL\left(\frac{x}{2}\right) = uL(x) \tag{8}$$

for all $u \in U(A)$ and all $x \in X$.

By the same reasoning as in [67], we can show that $L : X \to Y$ is $\mathbb{R}$-linear, since the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$ for each $x \in X$ and $L : X \to Y$ is additive.

Since L is $\mathbb{R}$-linear and each $a \in A$ is a finite linear combination of unitary elements (see [42, Theorem 4.1.7]), i.e., $a = \sum_{j=1}^{m} \lambda_j u_j$ ($\lambda_j \in \mathbb{C}, u_j \in U(A)$), it follows from (8) that

$$\begin{aligned} L(ax) &= L(\sum_{j=1}^{m} \lambda_j u_j x) = L\left(\sum_{j=1}^{m} |\lambda_j| \cdot \frac{\lambda_j}{|\lambda_j|} u_j x\right) = \sum_{j=1}^{m} |\lambda_j| L\left(\frac{\lambda_j}{|\lambda_j|} u_j x\right) \\ &= \sum_{j=1}^{m} |\lambda_j| \cdot \frac{\lambda_j}{|\lambda_j|} u_j L(x) = \sum_{j=1}^{m} \lambda_j u_j L(x) = aL(x) \end{aligned}$$

for all $x \in X$. It is obvious that $\frac{\lambda_j}{|\lambda_j|} u_j \in U(A)$. Thus $L : X \to Y$ is a unique orthogonally additive A-linear mapping satisfying (4).

Corollary 1 *Let θ be a positive real number and p a real number with $0 < p < 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$ and*

$$\left\| 2uf\left(\frac{x}{2} + y\right) - f(ux) - f(2uy) \right\|_Y \le \theta(\|x\|^p + \|y\|^p) \tag{9}$$

for all $u \in U(A)$ and all $x, y \in X$ with $x \perp y$. If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ such that

$$\|f(x) - L(x)\|_Y \le \frac{2^p \theta}{2 - 2^p} \|x\|^p \tag{10}$$

for all $x \in X$.

Proof. The proof follows from Theorem 3 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$ with $x \perp y$. Then we can choose $\alpha = 2^{p-1}$ and the result follows.

Theorem 4 *Let $f : X \to Y$ be a mapping satisfying (3) and $f(0) = 0$ for which there exists a function $\varphi : X \times X \to [0, \infty)$ such that there exists an $\alpha < 1$ with*

$$\varphi(x, y) \leq \frac{\alpha}{2}\varphi\left(2x, 2y\right) \tag{11}$$

for all $x, y \in X$ with $x \perp y$. If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ such that

$$\|f(x) - L(x)\|_Y \leq \frac{1}{1-\alpha}\varphi\left(x, 0\right) \tag{12}$$

for all $x \in X$.

Proof. Let (S, d) be the generalized metric space defined in the proof of Theorem 3.

Now we consider the linear mapping $J : S \to S$ such that

$$Jg(x) := 2g\left(\frac{x}{2}\right)$$

for all $x \in X$.

It follows from (5) that $d(f, Jf) \leq 1$.

The rest of the proof is similar to the proof of Theorem 3.

Corollary 2 *Let θ be a positive real number and p a real number with $p > 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$ and (9). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ such that*

$$\|f(x) - L(x)\|_Y \leq \frac{2^p\theta}{2^p - 2}\|x\|^p \tag{13}$$

for all $x \in X$.

Proof. The proof follows from Theorem 4 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$ with $x \perp y$. Then we can choose $\alpha = 2^{1-p}$ and the result follows.

3 Stability of the Orthogonally Additive Functional Equation (1) in Non-Archimedean Banach Modules Over a C^*-Algebra: Fixed Point Method

Throughout this section, assume that A is a unital C^*-algebra with unit e and unitary group $U(A)$, $(X, \perp)$ is an orthogonality non-Archimedean normed module over A and $(Y, \|.\|_Y)$ is a non-Archimedean Banach module over A. Assume that $|2| \neq 1$.

Theorem 5 *Let $\varphi : X \times X \to [0, \infty)$ be a function such that there exists an $\alpha < 1$ with*

$$\varphi(x, y) \le |2|\alpha\varphi\left(\frac{x}{2}, \frac{y}{2}\right) \tag{14}$$

for all $x, y \in X$ with $x \perp y$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$ and (3). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ satisfying (4).

Proof. It follows from (5) that

$$\left\| f(x) - \frac{1}{2} f(2x) \right\|_Y \le \frac{1}{|2|}\varphi(2x, 0) \le \alpha \cdot \varphi(x, 0) \tag{15}$$

for all $x \in X$.

Let (S, d) be the generalized metric space defined in the proof of Theorem 3.

We consider the linear mapping $J : S \to S$ such that

$$Jg(x) := 2g\left(\frac{x}{2}\right)$$

for all $x \in X$.

It follows from (15) that $d(f, Jf) \le \alpha$.

By Theorem 1, there exists a mapping $L : X \to Y$ satisfying the following:

1. $d(J^n f, L) \to 0$ as $n \to \infty$. This implies the equality
$$\lim_{n\to\infty} \frac{1}{2^n} f\left(2^n x\right) = L(x)$$
for all $x \in X$;
2. $d(f, L) \le \frac{1}{1-\alpha} d(f, Jf)$, which yields the inequality
$$d(f, L) \le \frac{\alpha}{1-\alpha}.$$

Thus (4) holds true.

It follows from (14) and (3) that

$$\begin{aligned}
&\left\| 2uL\left(\frac{x}{2} + y\right) - L(ux) - L(2uy) \right\|_Y \\
&\quad = \lim_{n\to\infty} \frac{1}{|2|^n} \| 2uf(2^{n-1}x + 2^n y) - f(2^n ux) - f(2^{n+1}uy) \|_Y \\
&\quad \le \lim_{n\to\infty} \frac{1}{|2|^n} \varphi(2^n x, 2^n y) \le \lim_{n\to\infty} \frac{|2|^n \alpha^n}{|2|^n} \varphi(x, y) = 0
\end{aligned}$$

for all $u \in U(A)$ and all $x, y \in X$ with $x \perp y$. So

$$2uL\left(\frac{x}{2}+y\right)-L(ux)-L(2uy)=0$$

for all $u \in U(A)$ and all $x, y \in X$ with $x \perp y$.

The rest of the proof is similar to the proof of Theorem 3.

Corollary 3 *Let θ be a positive real number and p a real number with $p > 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$ and (9). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ such that*

$$\|f(x)-L(x)\|_Y \le \frac{|2|^p\theta}{|2|-|2|^p}\|x\|^p \tag{16}$$

for all $x \in X$.

Proof. The proof follows from Theorem 5 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$ with $x \perp y$. Then we can choose $\alpha = |2|^{p-1}$ and the result follows.

Theorem 6 *Let $f : X \to Y$ be a mapping satisfying (3) and $f(0) = 0$ for which there exists a function $\varphi : X \times X \to [0,\infty)$ such that there exists an $\alpha < 1$ with*

$$\varphi(x,y) \le \frac{\alpha}{|2|}\varphi\,(2x,2y) \tag{17}$$

for all $x, y \in X$ with $x \perp y$. If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive A-linear mapping $L : X\to Y$ satisfying (12).

Proof. Let (S, d) be the generalized metric space defined in the proof of Theorem 3.

We consider the linear mapping $J : S \to S$ such that

$$Jg(x) := 2g\left(\frac{x}{2}\right)$$

for all $x \in X$.

It follows from (5) that $d(f, Jf) \le 1$.

The rest of the proof is similar to the proofs of Theorems 3 and 5.

Corollary 4 *Let θ be a positive real number and p a real number with $0 < p < 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$ and (9). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ such that*

$$\|f(x)-L(x)\|_Y \le \frac{|2|^p\theta}{|2|^p-|2|}\|x\|^p \tag{18}$$

for all $x \in X$.

Proof. The proof follows from Theorem 6 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$ with $x \perp y$. Then we can choose $\alpha = |2|^{1-p}$ and the result follows.

4 Stability of the Orthogonally Additive Functional Equation (1) in Banach Modules Over a C^*-Algebra: Direct Method

Throughout this section, assume that A is a unital C^*-algebra with unit e and unitary group $U(A)$, $(X, \perp)$ is an orthogonality normed module over A and $(Y, \|.\|_Y)$ is a Banach module over A.

Theorem 7 *Let $\varphi : X \times X \to [0, \infty)$ be a function satisfying*

$$\Phi(x, y) := \sum_{j=1}^{\infty} 2^j \varphi\left(\frac{x}{2^j}, \frac{y}{2^j}\right) < \infty \tag{19}$$

for all $x, y \in X$ with $x \perp y$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$ and (3). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ such that

$$\|f(x) - L(x)\|_Y \leq \frac{1}{2}\Phi\,(0, x) \tag{20}$$

for all $x \in X$.

Proof. Putting $x = 0$ and $u = e$ in (3), we get

$$\|2f\,(y) - f(2y)\|_Y \leq \varphi(0, y) \tag{21}$$

for all $y \in X$, since $y \perp 0$. So

$$\left\| f(x) - 2f\left(\frac{x}{2}\right)\right\|_Y \leq \varphi\left(0, \frac{x}{2}\right) \tag{22}$$

for all $x \in X$.

It follows from (22) that

$$\left\|2^l f(\frac{x}{2^l}) - 2^m f(\frac{x}{2^m})\right\|_Y \leq \sum_{j=l+1}^{m} 2^j \varphi\left(0, \frac{x}{2^j}\right) \tag{23}$$

for all nonnegative integers m and l with $m > l$ and all $x \in X$. It follows from (19) and (23) that the sequence $\{2^k f(\frac{x}{2^k})\}$ is Cauchy for all $x \in X$. Since Y is complete, the sequence $\{2^k f(\frac{x}{2^k})\}$ converges. So one can define the mapping $L : X \to Y$ by

$$L(x) := \lim_{k\to\infty} 2^k f\left(\frac{x}{2^k}\right)$$

for all $x \in X$.

By (19) and (3),

$$\begin{aligned}\left\|2L\left(\frac{x}{2}+y\right)-L(x)-L(2y)\right\|_Y &= \lim_{k\to\infty} 2^k \left\|2f\left(\frac{x}{2^{k+1}}+\frac{y}{2^k}\right)\right. \\ &\quad \left.-f\left(\frac{x}{2^k}\right)-f\left(\frac{y}{2^{k-1}}\right)\right\|_Y \\ &\leq \lim_{k\to\infty} 2^k \varphi\left(\frac{x}{2^k},\frac{y}{2^k}\right)=0\end{aligned}$$

for all $x, y \in X$ with $x \perp y$. So $2L\left(\frac{x}{2}+y\right)-L(x)-L(2y)=0$. Thus the mapping $L : X \to Y$ is orthogonally additive. Moreover, letting $l = 0$ and passing the limit $m \to \infty$ in (23), we get (20). Therefore, there exists an orthogonally additive mapping $L : X \to Y$ satisfying (20).

Now, let $T : X \to Y$ be another orthogonally additive mapping satisfying (1) and (20). Then we have

$$\begin{aligned}\|L(x)-T(x)\|_Y &= 2^q\left\|L\left(\frac{x}{2^q}\right)-T\left(\frac{x}{2^q}\right)\right\|_Y \\ &\leq 2^q\left\|L\left(\frac{x}{2^q}\right)-f\left(\frac{x}{2^q}\right)\right\|_Y+2^q\left\|T\left(\frac{x}{2^q}\right)-f\left(\frac{x}{2^q}\right)\right\|_Y \\ &\leq 2\cdot 2^q \Phi\left(0,\frac{x}{2^q}\right),\end{aligned}$$

which tends to zero as $q \to \infty$ for all $x \in X$. So we can conclude that $L(x) = T(x)$ for all $x \in X$. This proves the uniqueness of L.

The rest of the proof is similar to the proof of Theorem 3.

Corollary 5 *Let θ be a positive real number and p a real number with $p > 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$ and (9). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ such that*

$$\|f(x)-L(x)\|_Y \leq \frac{\theta}{2^p-2}\|x\|^p \tag{24}$$

for all $x \in X$.

Proof. The proof follows from Theorem 7 by taking $\varphi(x,y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$ with $x \perp y$.

Theorem 8 *Let $\varphi : X \times X \to [0,\infty)$ be a function satisfying*

$$\Phi(x,y) := \sum_{j=0}^{\infty} \frac{1}{2^j} \varphi\left(2^j x, 2^j y\right) < \infty \tag{25}$$

for all $x, y \in X$ with $x \perp y$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$ and (3). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ such that

$$\|f(x) - L(x)\|_Y \le \frac{1}{2}\Phi\,(0,x) \tag{26}$$

for all $x \in X$.

Proof. It follows from (21) that

$$\left\| f(x) - \frac{1}{2} f(2x) \right\|_Y \le \frac{1}{2}\varphi(0,x)$$

for all $x \in X$. So

$$\left\| \frac{1}{2^l} f(2^l x) - \frac{1}{2^m} f(2^m x) \right\|_Y \le \sum_{j=l}^{m-1} \frac{1}{2 \cdot 2^j} \varphi(0, 2^j x) \tag{27}$$

for all nonnegative integers m and l with $m > l$ and all $x \in X$. It follows from (25) and (27) that the sequence $\{\frac{1}{2^k} f(2^k x)\}$ is Cauchy for all $x \in X$. Since the space Y is complete, the sequence $\{\frac{1}{2^k} f(2^k x)\}$ converges. So one can define the mapping $L : X \to Y$ by

$$L(x) := \lim_{k\to\infty} \frac{1}{2^k} f(2^k x)$$

for all $x \in X$.

The rest of the proof is similar to the proofs of Theorems 3 and 7.

Corollary 6 *Let θ be a positive real number and p a real number with $0 < p < 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$ and (9). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ such that*

$$\|f(x) - L(x)\|_Y \le \frac{\theta}{2 - 2^p} \|x\|^p \tag{28}$$

for all $x \in X$.

Proof. The proof follows from Theorem 8 by taking $\varphi(x,y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$ with $x \perp y$.

5 Stability of C^*-Algebra Homomorphisms Associated with the Orthogonally Additive Functional Equation (1): Fixed Point Method

Let X be a C^*-algebra and $x, y \in X$. Define $\langle x, y\rangle = xy^*$. Then both (O_1) and (O_3) hold, i.e., if $x, y \in X, x \perp y$, then $\alpha x \perp \beta y$ for all $\alpha, \beta \in \mathbb{C}$. We say that x and y are called *orthogonal* if $xy = 0 = yx$.

Throughout this section, assume that X is a unital C^*-algebra with unit e and unitary group $U(X)$ and $(Y, \|.\|_Y)$ is a unital C^*-algebra.

Definition 8. An orthogonally additive $\mathbb{C}$-linear mapping $f : X \to Y$ is called an *orthogonally additive C^*-algebra homomorphism* if $f(xz) = f(x)f(z)$ and $f(x^*) = f(x)^*$ for all $x, z \in X$.

Theorem 9 *Let $\varphi : X \times X \to [0, \infty)$ be a function such that there exists an $\alpha < 1$ with*

$$\varphi(x, y) \leq 2\alpha\varphi\left(\frac{x}{2}, \frac{y}{2}\right) \tag{29}$$

for all $x, y \in X$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$ and

$$\left\|2\mu f\left(\frac{x}{2} + y\right) - f(\mu x) - f(2\mu y)\right\|_Y \leq \varphi(x, y), \tag{30}$$

$$\|f(xz) - f(x)f(z)\|_Y \leq \varphi(x, z), \tag{31}$$

$$\left\|f\left(x^*\right) - f(x)^*\right\|_Y \leq \varphi(x, x) \tag{32}$$

for all $\mu \in \mathbb{T} := \{\lambda \in \mathbb{C} \mid |\lambda| = 1\}$ and all $x, y, z \in X$ with $x \perp y$. If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive C^*-algebra homomorphism $H : X \to Y$ such that

$$\|f(x) - H(x)\|_Y \leq \frac{\alpha}{1-\alpha}\varphi(x, 0) \tag{33}$$

for all $x \in X$.

Proof. Putting $y = 0$ and $\mu = 1$ in (30), we get

$$\left\|2f\left(\frac{x}{2}\right) - f(x)\right\|_Y \leq \varphi(x, 0) \tag{34}$$

for all $x \in X$, since $x \perp 0$. So

$$\left\|f(x) - \frac{1}{2}f(2x)\right\|_Y \leq \frac{1}{2}\varphi(2x, 0) \leq \alpha \cdot \varphi(x, 0) \tag{35}$$

for all $x \in X$.

Consider the set

$$S := \{h : X \to Y\}$$

and introduce the generalized metric on S:

$$d(g, h) = \inf\{\mu \in \mathbb{R}_+ : \|g(x) - h(x)\|_Y \leq \mu\varphi(x, 0), \ \ \forall x \in X\},$$

where, as usual, $\inf \phi = +\infty$. It is easy to show that (S, d) is complete (see [48, Lemma 2.1]).

We consider the linear mapping $J : S \to S$ such that

$$Jg(x) := \frac{1}{2} g(2x)$$

for all $x \in X$.

It follows from (35) that $d(f, Jf) \leq \alpha$.

By Theorem 1, there exists a mapping $H : X \to Y$ satisfying the following:

1. H is a fixed point of J, i.e.,

$$H(2x) = 2H(x) \tag{36}$$

for all $x \in X$. The mapping H is a unique fixed point of J in the set

$$M = \{g \in S : d(h, g) < \infty\}.$$

This implies that H is a unique mapping satisfying (36) such that there exists a $\mu \in (0, \infty)$ satisfying

$$\|f(x) - H(x)\|_Y \leq \mu\varphi(x, 0)$$

for all $x \in X$;

2. $d(J^n f, H) \to 0$ as $n \to \infty$. This implies the equality

$$\lim_{n\to\infty} \frac{1}{2^n} f(2^n x) = H(x)$$

for all $x \in X$;

3. $d(f, H) \leq \frac{1}{1-\alpha} d(f, Jf)$, which yields the inequality

$$d(f, H) \leq \frac{\alpha}{1 - \alpha}.$$

Thus (33) holds true.

By the same reasoning as in the proof of Theorem 6, one can show that the mapping $H : X \to Y$ is an orthogonally additive and $\mathbb{C}$-linear mapping satisfying (33).

It follows from (29) and (31) that

$$\begin{aligned}\|H\,(xz) - H(x)H(z)\|_Y &= \lim_{n\to\infty} \frac{1}{4^n}\|f(2^n x\cdot 2^n z) - f(2^n x)f(2^n z)\|_Y \\ &\le \lim_{n\to\infty}\frac{1}{4^n}\varphi(2^n x, 2^n z) \le \lim_{n\to\infty}\frac{2^n\alpha^n}{4^n}\varphi(x,z) = 0\end{aligned}$$

for all $x, z \in X$. So

$$H(xz) - H(x)H(y) = 0$$

for all $x, z \in X$. Hence $H : X \to Y$ is multiplicative.

It follows from (29) and (32) that

$$\begin{aligned}\left\|H\left(x^*\right) - H(x)^*\right\|_Y &= \lim_{n\to\infty}\frac{1}{2^n}\|f(2^n x^*) - f(2^n x)^*\|_Y \\ &\le \lim_{n\to\infty}\frac{1}{2^n}\varphi(2^n x, 2^n x) \le \lim_{n\to\infty}\frac{2^n\alpha^n}{2^n}\varphi(x,x) = 0\end{aligned}$$

for all $x \in X$. Therefore,

$$H\left(x^*\right) - H(x)^* = 0$$

for all $x \in X$. Hence

$$H(x^*) = H(x)^*$$

for all $x \in X$. Thus $H : X \to Y$ is a unique orthogonally additive C^*-algebra homomorphism satisfying (33).

Corollary 7 *Let θ be a positive real number and p a real number with $0 < p < 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$ and*

$$\left\|2\mu f\left(\frac{x}{2} + y\right) - f(\mu x) - f(2\mu y)\right\|_Y \le \theta(\|x\|^p + \|y\|^p), \tag{37}$$

$$\|f\,(xz) - f(x)f(z)\|_Y \le \theta(\|x\|^p + \|z\|^p), \tag{38}$$

$$\left\|f\left(x^*\right) - f(x)^*\right\|_Y \le 2\theta\|x\|^p \tag{39}$$

for all $\mu \in \mathbb{T}$ and all $x, y, z \in X$ with $x \perp y$. If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive C^*-algebra homomorphism $H : X \to Y$ such that

$$\|f(x) - H(x)\|_Y \leq \frac{2^p\theta}{2 - 2^p}\|x\|^p$$

for all $x \in X$.

Proof. The proof follows from Theorem 9 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$. Then we can choose $\alpha = 2^{p-1}$ and the result follows.

Theorem 10 *Let $f : X \to Y$ be a mapping satisfying (30)–(32) and $f(0) = 0$ for which there exists a function $\varphi : X \times X \to [0, \infty)$ such that there exists an $\alpha < 1$ with*

$$\varphi(x, y) \leq \frac{\alpha}{2}\varphi\,(2x, 2y)$$

for all $x, y \in X$. If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive C^*-algebra homomorphism $H : X \to Y$ such that

$$\|f(x) - H(x)\|_Y \leq \frac{1}{1-\alpha}\varphi\,(x, 0)$$

for all $x \in X$.

Proof. Let (S, d) be the generalized metric space defined in the proof of Theorem 9.

We consider the linear mapping $J : S \to S$ such that

$$Jg(x) := 2g\left(\frac{x}{2}\right)$$

for all $x \in X$.

It follows from (34) that $d(f, Jf) \leq 1$.

The rest of the proof is similar to the proof of Theorem 9.

Corollary 8 *Let θ be a positive real number and p a real number with $p > 2$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (37)–(39). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive C^*-algebra homomorphism $H : X \to Y$ such that*

$$\|f(x) - H(x)\|_Y \leq \frac{2^p\theta}{2^p - 2}\|x\|^p$$

for all $x \in X$.

Proof. The proof follows from Theorem 10 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$. Then we can choose $\alpha = 2^{1-p}$ and the result follows.

6 Stability of C^*-Algebra Homomorphisms Associated with the Orthogonally Additive Functional Equation (1): Direct Method

Throughout this section, assume that X is a unital C^*-algebra with unit e and unitary group $U(X)$ and $(Y, \|.\|_Y)$ is a unital C^*-algebra.

Theorem 11 *Let $\varphi : X \times X \to [0, \infty)$ be a function satisfying*

$$\Phi(x, y) := \sum_{j=1}^{\infty} 2^j \varphi\left(\frac{x}{2^j}, \frac{y}{2^j}\right) < \infty \tag{40}$$

for all $x, y \in X$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (30)–(32). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive C^*-algebra homomorphism $H : X \to Y$ such that

$$\|f(x) - H(x)\|_Y \le \frac{1}{2}\Phi\,(0, x) \tag{41}$$

for all $x \in X$.

Proof. Putting $x = 0$ and $\mu = 1$ in (30), we get

$$\|2f\ (y) - f(2y)\|_Y \le \varphi(0, y) \tag{42}$$

for all $y \in X$, since $y \perp 0$. So

$$\left\| f(x) - 2f\left(\frac{x}{2}\right)\right\|_Y \le \varphi\left(0, \frac{x}{2}\right) \tag{43}$$

for all $x \in X$.

It follows from (43) that

$$\|2^l f(\frac{x}{2^l}) - 2^m f(\frac{x}{2^m})\|_Y \le \sum_{j=l+1}^{m} 2^j \varphi\left(0, \frac{x}{2^j}\right) \tag{44}$$

for all nonnegative integers m and l with $m > l$ and all $x \in X$. It follows from (40) and (44) that the sequence $\{2^k f(\frac{x}{2^k})\}$ is Cauchy for all $x \in X$. Since Y is complete, the sequence $\{2^k f(\frac{x}{2^k})\}$ converges. So one can define the mapping $L : X \to Y$ by

$$L(x) := \lim_{k\to\infty} 2^k f\left(\frac{x}{2^k}\right)$$

for all $x \in X$.

By (40) and (30),

$$\begin{aligned}\left\|2L\left(\frac{x}{2}+y\right)-L(x)-L(2y)\right\|_Y &= \lim_{k\to\infty} 2^k \left\|2f\left(\frac{x}{2^{k+1}}+\frac{y}{2^k}\right)\right. \\ &\qquad \left. -f\left(\frac{x}{2^k}\right)-f\left(\frac{y}{2^{k-1}}\right)\right\|_Y \\ &\le \lim_{k\to\infty} 2^k\varphi\left(\frac{x}{2^k},\frac{y}{2^k}\right)=0\end{aligned}$$

for all $x, y \in X$ with $x \perp y$. So $2L\left(\frac{x}{2}+y\right) - L(x) - L(2y) = 0$. Thus the mapping $L : X \to Y$ is orthogonally additive. Moreover, letting $l = 0$ and passing the limit $m \to \infty$ in (44), we get (41). Therefore, there exists an orthogonally additive mapping $L : X \to Y$ satisfying (41).

Now, let $T : X \to Y$ be another orthogonally additive mapping satisfying (1) and (41). Then we have

$$\begin{aligned}\|L(x)-T(x)\|_Y &= 2^q\left\|L\left(\frac{x}{2^q}\right)-T\left(\frac{x}{2^q}\right)\right\|_Y \\ &\le 2^q\left\|L\left(\frac{x}{2^q}\right)-f\left(\frac{x}{2^q}\right)\right\|_Y + 2^q\left\|T\left(\frac{x}{2^q}\right)-f\left(\frac{x}{2^q}\right)\right\|_Y \\ &\le 2\cdot 2^q\Phi\left(0,\frac{x}{2^q}\right),\end{aligned}$$

which tends to zero as $q \to \infty$ for all $x \in X$. So we can conclude that $L(x) = T(x)$ for all $x \in X$. This proves the uniqueness of L.

The rest of the proof is similar to the proofs of Theorem 3 and 7.

Corollary 9 *Let θ be a positive real number and p a real number with $p > 2$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (37)–(39). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive C^*-algebra homomorphism $H : X \to Y$ such that*

$$\|f(x)-H(x)\|_Y \le \frac{\theta}{2^p-2}\|x\|^p$$

for all $x \in X$.

Proof. The proof follows from Theorem 11 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$.

Theorem 12 *Let $\varphi : X \times X \to [0,\infty)$ be a function satisfying*

$$\Phi(x,y) := \sum_{j=0}^{\infty}\frac{1}{2^j}\varphi\left(2^j x, 2^j y\right) < \infty \tag{45}$$

for all $x, y \in X$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (30)–(32). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive C^*-algebra homomorphism $H : X \to Y$ such that

$$\|f(x) - H(x)\|_Y \le \frac{1}{2}\Phi(0, x) \tag{46}$$

for all $x \in X$.

Proof. It follows from (42) that

$$\left\| f(x) - \frac{1}{2}f(2x) \right\|_Y \le \frac{1}{2}\varphi(0, x)$$

for all $x \in X$. So

$$\left\| \frac{1}{2^l}f(2^l x) - \frac{1}{2^m}f(2^m x) \right\|_Y \le \sum_{j=l}^{m-1} \frac{1}{2 \cdot 2^j}\varphi(0, 2^j x) \tag{47}$$

for all nonnegative integers m and l with $m > l$ and all $x \in X$. It follows from (45) and (47) that the sequence $\{\frac{1}{2^k}f(2^k x)\}$ is Cauchy for all $x \in X$. Since Y is a complete space, the sequence $\{\frac{1}{2^k}f(2^k x)\}$ converges. So one can define the mapping $L : X \to Y$ by

$$L(x) := \lim_{k\to\infty} \frac{1}{2^k}f(2^k x)$$

for all $x \in X$.

The rest of the proof is similar to the proofs of Theorems 3 and 11.

Corollary 10 *Let θ be a positive real number and p a real number with $0 < p < 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (37)–(39). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive C^*-algebra homomorphism $H : X \to Y$ such that*

$$\|f(x) - H(x)\|_Y \le \frac{\theta}{2 - 2^p}\|x\|^p$$

for all $x \in X$.

Proof. The proof follows from Theorem 12 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$.

7 Stability of Isometric Linear Mappings in Banach Modules Over a C^*-Algebra Associated with the Orthogonally Additive Functional Equation (1): Fixed Point Method

Throughout this section, assume that A is a unital C^*-algebra with unit e and unitary group $U(A)$, $(X, \perp)$ is an orthogonality normed module over A and $(Y, \|.\|_Y)$ is a Banach module over A.

Definition 9. An orthogonally additive A-linear mapping $f : X \to Y$ is called an *orthogonally additive isometric A-linear mapping* if $\|f(x)\|_Y = ||x||$ for all $x \in X$.

Theorem 13 *Let $\varphi : X \times X \to [0, \infty)$ be a function satisfying (2). Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (3) and*

$$|\ \|f(x)\|_Y - ||x||\ | \leq \varphi(x, 0) \tag{48}$$

for all $x \in X$. If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive isometric A-linear mapping $L : X \to Y$ satisfying (4).

Proof. By Theorem 3, there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ satisfying (4).

It follows from (2) and (48) that

$$\begin{aligned} |\ \|L(x)\|_Y - ||x||\ | &= \lim_{n\to\infty} \frac{1}{2^n} |\ \|f(2^n x)\|_Y - ||2^n x||\ | \\ &\leq \lim_{n\to\infty} \frac{1}{2^n}\varphi(2^n x, 0) \leq \lim_{n\to\infty} \frac{2^n\alpha^n}{2^n}\varphi(x, 0) = 0 \end{aligned}$$

for all $x \in X$. So $\|L(x)\|_Y - ||x|| = 0$ for all $x \in X$. Hence

$$\|L(x)\|_Y = ||x||$$

for all $x \in X$. Thus $L : X \to Y$ is a unique orthogonally isometric A-linear mapping satisfying (4).

Corollary 11 *Let θ be a positive real number and p a real number with $0 < p < 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (9) and*

$$|\ \|f(x)\|_Y - ||x||\ | \leq \theta\|x\|^p \tag{49}$$

for all $x \in X$. If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive isometric A-linear mapping $L : X \to Y$ satisfying (10).

Proof. The proof follows from Theorem 13 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$ with $x \perp y$. Then we can choose $\alpha = 2^{p-1}$ and the result follows.

Theorem 14 *Let $f : X \to Y$ be a mapping satisfying (3), (48) and $f(0) = 0$ for which there exists a function $\varphi : X \times X \to [0, \infty)$ satisfying (11). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive isometric A-linear mapping $L : X \to Y$ satisfying (12).*

Proof. The proof is similar to the proofs of Theorems 3 and 13.

Corollary 12 *Let θ be a positive real number and p a real number with $p > 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (9) and (49). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive isometric A-linear mapping $L : X \to Y$ satisfying (13).*

Proof. The proof follows from Theorem 14 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$ with $x \perp y$. Then we can choose $\alpha = 2^{1-p}$ and the result follows.

8 Stability of Isometric Linear Mappings in Non-Archimedean Banach Modules Over a C^*-Algebra: Fixed Point Method

Throughout this section, assume that A is a unital C^*-algebra with unit e and unitary group $U(A)$, $(X, \perp)$ is an orthogonality non-Archimedean normed module over A and $(Y, \|.\|_Y)$ is a non-Archimedean Banach module over A. Assume that $|2| \neq 1$.

Theorem 15 *Let $\varphi : X \times X \to [0, \infty)$ be a function satisfying (14). Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (3) and (48). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive isometric A-linear mapping $L : X \to Y$ satisfying (4).*

Proof. The proof is similar to the proofs of Theorems 3, 5 and 13.

Corollary 13 *Let θ be a positive real number and p a real number with $p > 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (9) and (49). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive isometric A-linear mapping $L : X \to Y$ satisfying (16).*

Proof. The proof follows from Theorem 15 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$ with $x \perp y$. Then we can choose $\alpha = |2|^{p-1}$ and the result follows.

Theorem 16 *Let $f : X \to Y$ be a mapping satisfying (3), (48) and $f(0) = 0$ for which there exists a function $\varphi : X \times X \to [0, \infty)$ satisfying (17). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive isometric A-linear mapping $L : X \to Y$ satisfying (12).*

Proof. The proof is similar to the proofs of Theorems 3, 5 and 13.

Corollary 14 *Let θ be a positive real number and p a real number with $0 < p < 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (9) and (49). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive isometric A-linear mapping $L : X \to Y$ satisfying (18).*

Proof. The proof follows from Theorem 16 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$ with $x \perp y$. Then we can choose $\alpha = |2|^{1-p}$ and the result follows.

9 Stability of Isometric Linear Mappings in Banach Modules Over a C^*-Algebra Associated with the Orthogonally Additive Functional Equation (1): Direct Method

Throughout this section, assume that A is a unital C^*-algebra with unit e and unitary group $U(A)$, $(X, \perp)$ is an orthogonality normed module over A and $(Y, \|.\|_Y)$ is a Banach module over A.

Theorem 17 *Let $\varphi : X \times X \to [0, \infty)$ be a function satisfying (19). Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (3) and (48). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive isometric A-linear mapping $L : X \to Y$ satisfying (20).*

Proof. By Theorem 7, there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ satisfying (20).

The rest of the proof is similar to the proof of Theorem 13.

Corollary 15 *Let θ be a positive real number and p a real number with $p > 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (9) and (49). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive isometric A-linear mapping $L : X \to Y$ satisfying (24).*

Proof. The proof follows from Theorem 17 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$ with $x \perp y$.

Theorem 18 *Let $\varphi : X \times X \to [0, \infty)$ be a function satisfying (25). Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (3) and (48). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive isometric A-linear mapping $L : X \to Y$ satisfying (26).*

Proof. The proof is similar to the proofs of Theorems 3, 8 and 13.

Corollary 16 *Let θ be a positive real number and p a real number with $0 < p < 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (9) and (49). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive isometric A-linear mapping $L : X \to Y$ satisfying (28).*

Proof. The proof follows from Theorem 18 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$ with $x \perp y$.

10 Stability of 2-Isometric Linear Mappings in Banach Modules Over a C^*-Algebra Associated with the Orthogonally Additive Functional Equation (1): Fixed Point Method

Throughout this section, assume that A is a unital C^*-algebra with unit e and unitary group $U(A)$, $(X, \perp)$ is an orthogonality normed module over A and $(Y, \|\cdot\|_Y)$ is a Banach module over A.

Definition 10. An orthogonally additive A-linear mapping $f : X \to Y$ is called an *orthogonally additive 2-isometric linear mapping* if $\|f(x), f(z)\|_Y = ||x, z||$ for all $x, z \in X$.

Theorem 19 *Let $\varphi : X \times X \to [0, \infty)$ be a function such that there exists an $\alpha < 1$ with*

$$\varphi(x, y) \le 2\alpha\varphi\left(\frac{x}{2}, \frac{y}{2}\right) \tag{50}$$

for all $x, y \in X$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (3) and

$$|\ \|f(x), f(z)\|_Y - ||x, z||\ | \le \varphi(x, z) \tag{51}$$

for all $x, z \in X$. If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive 2-isometric A-linear mapping $L : X \to Y$ satisfying (4).

Proof. By Theorem 3, there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ satisfying (4).

It follows from (50) and (51) that

$$\begin{aligned} |\ \|L(x), L(z)\|_Y - ||x, z||\ | &= \lim_{n\to\infty} \frac{1}{2^n} |\ \|f(2^n x), f(2^n z)\|_Y - ||2^n x, 2^n z||\ | \\ &\le \lim_{n\to\infty} \frac{1}{2^n}\varphi(2^n x, 2^n z) \le \lim_{n\to\infty} \frac{2^n\alpha^n}{2^n}\varphi(x, z) = 0 \end{aligned}$$

for all $x, z \in X$. So $\|L(x), L(z)\|_Y - ||x, z|| = 0$ for all $x, z \in X$. Hence

$$\|L(x), L(z)\|_Y = ||x, z||$$

for all $x, z \in X$. Thus $L : X \to Y$ is a unique orthogonally 2-isometric A-linear mapping satisfying (4).

Corollary 17 *Let θ be a positive real number and p a real number with $0 < p < 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (9) and*

$$|\ \|f(x), f(z)\|_Y - ||x, z||\ | \le \theta\|x\|^p \tag{52}$$

for all $x, z \in X$. If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive 2-isometric A-linear mapping $L : X \to Y$ satisfying (10).

Proof. The proof follows from Theorem 19 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$. Then we can choose $\alpha = 2^{p-1}$ and the result follows.

Theorem 20 *Let $f : X \to Y$ be a mapping satisfying (3), (51) and $f(0) = 0$ for which there exists a function $\varphi : X \times X \to [0, \infty)$ such that there exists an $\alpha < 1$ with*

$$\varphi(x, y) \le \frac{\alpha}{2}\varphi(2x, 2y)$$

for all $x, y \in X$. If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive 2-isometric A-linear mapping $L : X \to Y$ satisfying (12).

Proof. The proof is similar to the proofs of Theorems 3 and 19.

Corollary 18 *Let θ be a positive real number and p a real number with $p > 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (9) and (52). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive 2-isometric A-linear mapping $L : X \to Y$ satisfying (13).*

Proof. The proof follows from Theorem 20 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$. Then we can choose $\alpha = 2^{1-p}$ and the result follows.

11 Stability of 2-Isometric Linear Mappings in Non-Archimedean Banach Modules Over a C^*-Algebra Associated with the Orthogonally Additive Functional Equation (1): Fixed Point Method

Throughout this section, assume that A is a unital C^*-algebra with unit e and unitary group $U(A)$, $(X, \perp)$ is an orthogonality non-Archimedean normed module over A and $(Y, \|.\|_Y)$ is a non-Archimedean Banach module over A. Assume that $|2| \neq 1$.

Theorem 21 *Let $\varphi : X \times X \to [0, \infty)$ be a function such that there exists an $\alpha < 1$ with*

$$\varphi(x, y) \le |2|\alpha\varphi\left(\frac{x}{2}, \frac{y}{2}\right)$$

for all $x, y \in X$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (3) and (51). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive 2-isometric A-linear mapping $L : X \to Y$ satisfying (4).

Proof. The proof is similar to the proofs of Theorems 3, 5 and 19.

Corollary 19 *Let θ be a positive real number and p a real number with $p > 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (9) and (52). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive 2-isometric A-linear mapping $L : X \to Y$ satisfying (16).*

Proof. The proof follows from Theorem 21 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$. Then we can choose $\alpha = |2|^{p-1}$ and the result follows.

Theorem 22 *Let $f : X \to Y$ be a mapping satisfying (3), (51) and $f(0) = 0$ for which there exists a function $\varphi : X \times X \to [0, \infty)$ such that there exists an $\alpha < 1$ with*

$$\varphi(x, y) \le \frac{\alpha}{|2|}\varphi(2x, 2y)$$

for all $x, y \in X$. If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive 2-isometric A-linear mapping $L : X \to Y$ satisfying (12).

Proof. The proof is similar to the proofs of Theorems 3, 5 and 19.

Corollary 20 *Let θ be a positive real number and p a real number with $0 < p < 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (9) and (52). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive 2-isometric A-linear mapping $L : X \to Y$ satisfying (18).*

Proof. The proof follows from Theorem 22 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$. Then we can choose $\alpha = |2|^{1-p}$ and the result follows.

12 Stability of 2-Isometric Linear Mappings in Banach Modules Over a C^*-Algebra Associated with the Orthogonally Additive Functional Equation (1): Direct Method

Throughout this section, assume that A is a unital C^*-algebra with unit e and unitary group $U(A)$, $(X, \perp)$ is an orthogonality normed module over A and $(Y, \|.\|_Y)$ is a Banach module over A.

Theorem 23 *Let $\varphi : X \times X \to [0, \infty)$ be a function satisfying*

$$\Phi(x, y) := \sum_{j=1}^{\infty} 2^j \varphi\left(\frac{x}{2^j}, \frac{y}{2^j}\right) < \infty$$

for all $x, y \in X$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (3) and (51). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive 2-isometric A-linear mapping $L : X \to Y$ satisfying (20).

Proof. By Theorem 7, there exists a unique orthogonally additive A-linear mapping $L : X \to Y$ satisfying (20).

The rest of the proof is similar to the proof of Theorem 19.

Corollary 21 *Let θ be a positive real number and p a real number with $p > 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (9) and (52). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive 2-isometric A-linear mapping $L : X \to Y$ satisfying (24).*

Proof. The proof follows from Theorem 23 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$.

Theorem 24 *Let $\varphi : X \times X \to [0, \infty)$ be a function satisfying*

$$\Phi(x, y) := \sum_{j=0}^{\infty} \frac{1}{2^j} \varphi\left(2^j x, 2^j y\right) < \infty$$

for all $x, y \in X$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (3) and (51). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive 2-isometric A-linear mapping $L : X \to Y$ satisfying (26).

Proof. The proof is similar to the proofs of Theorems 3, 8 and 19.

Corollary 22 *Let θ be a positive real number and p a real number with $0 < p < 1$. Let $f : X \to Y$ be a mapping satisfying $f(0) = 0$, (9) and (52). If for each $x \in X$ the mapping $f(tx)$ is continuous in $t \in \mathbb{R}$, then there exists a unique orthogonally additive 2-isometric A-linear mapping $L : X \to Y$ satisfying (28).*

Proof. The proof follows from Theorem 3 by taking $\varphi(x, y) = \theta(\|x\|^p + \|y\|^p)$ for all $x, y \in X$.

References

1. A.D. Aleksandrov, Mappings of families of sets. Sov. Math. Dokl. **11**, 116–120 (1970)
2. J. Alonso, C. Benítez, Orthogonality in normed linear spaces: a survey I. Main properties. Extracta Math. **3**, 1–15 (1988)
3. J. Alonso, C. Benítez, Orthogonality in normed linear spaces: a survey II. Relations between main orthogonalities. Extracta Math. **4**, 121–131 (1989)
4. J. Baker, Isometries in normed spaces. Am. Math. Mon. **78**, 655–658 (1971)
5. G. Birkhoff, Orthogonality in linear metric spaces. Duke Math. J. **1**, 169–172 (1935)
6. J. Bourgain, Real isomorphic complex Banach spaces need not be complex isomorphic. Proc. Am. Math. Soc. **96**, 221–226 (1986)

7. L. Cădariu, V. Radu, Fixed points and the stability of Jensen's functional equation. J. Inequal. Pure Appl. Math. **4**(1), Art. ID 4 (2003)
8. L. Cădariu, V. Radu, On the stability of the Cauchy functional equation: a fixed point approach. Grazer Math. Ber. **346**, 43–52 (2004)
9. L. Cădariu, V. Radu, Fixed point methods for the generalized stability of functional equations in a single variable. Fixed Point Theory Appl. **2008**, Art. ID 749392 (2008)
10. S.O. Carlsson, Orthogonality in normed linear spaces. Ark. Mat. **4**, 297–318 (1962)
11. Y. Cho, P.C.S. Lin, S. Kim, A. Misiak, *Theory of* 2*-Inner Product Spaces* (Nova Science Publ., New York, 2001)
12. P.W. Cholewa, Remarks on the stability of functional equations. Aequ. Math. **27**, 76–86 (1984)
13. H. Chu, K. Lee, C. Park, On the Aleksandrov problem in linear n-normed spaces. Nonlinear Anal. – TMA **59**, 1001–1011 (2004)
14. H. Chu, C. Park, W. Park, The Aleksandrov problem in linear 2-normed spaces. J. Math. Anal. Appl. **289**, 666–672 (2004)
15. S. Czerwik, On the stability of the quadratic mapping in normed spaces. Abh. Math. Semin. Univ. Hambg. **62**, 59–64 (1992)
16. S. Czerwik, *Functional Equations and Inequalities in Several Variables* (World Scientific Publishing Company, New Jersey/London/Singapore/Hong Kong, 2002)
17. S. Czerwik, *Stability of Functional Equations of Ulam-Hyers-Rassias Type* (Hadronic Press, Palm Harbor/Florida, 2003)
18. D. Deses, On the representation of non-Archimedean objects. Topol. Appl. **153**, 774–785 (2005)
19. J. Diaz, B. Margolis, A fixed point theorem of the alternative for contractions on a generalized complete metric space. Bull. Am. Math. Soc. **74**, 305–309 (1968)
20. C.R. Diminnie, A new orthogonality relation for normed linear spaces. Math. Nachr. **114**, 197–203 (1983)
21. G. Dolinar, Generalized stability of isometries. J. Math. Anal. Appl. **242**, 39–56 (2000)
22. F. Drljević, On a functional which is quadratic on A-orthogonal vectors. Publ. Inst. Math. (Beograd) **54**, 63–71 (1986)
23. M. Fochi, Functional equations in A-orthogonal vectors. Aequ. Math. **38**, 28–40 (1989
24. R. Ger, J. Sikorska, Stability of the orthogonal additivity. Bull. Pol. Acad. Sci. Math. **43**, 143–151 (1995)
25. J. Gevirtz, Stability of isometries on Banach spaces. Proc. Am. Math. Soc. **89**, 633–636 (1983)
26. P. Gruber, Stability of isometries. Trans. Am. Math. Soc. **245**, 263–277 (1978)
27. S. Gudder, D. Strawther, Orthogonally additive and orthogonally increasing functions on vector spaces. Pac. J. Math. **58**, 427–436 (1975)
28. K. Hensel, Ubereine news Begrundung der Theorie der algebraischen Zahlen. Jahresber. Dtsch. Math. Ver. **6**, 83–88 (1897)
29. D.H. Hyers, On the stability of the linear functional equation. Proc. Natl. Acad. Sci. U.S.A. **27**, 222–224 (1941)
30. D.H. Hyers, G. Isac, Th.M. Rassias, *Stability of Functional Equations in Several Variables* (Birkh$\ddot{a}$user, Basel, 1998)
31. G. Isac, Th.M. Rassias, Stability of ψ-additive mappings: appications to nonlinear analysis. Int. J. Math. Math. Sci. **19**, 219–228 (1996)
32. R.C. James, Orthogonality in normed linear spaces. Duke Math. J. **12**, 291–302 (1945)
33. R.C. James, Orthogonality and linear functionals in normed linear spaces. Trans. Am. Math. Soc. **61**, 265–292 (1947)
34. S.-M. Jung, On the Hyers-Ulam stability of the functional equations that have the quadratic property. J. Math. Anal. Appl. **222**, 126–137 (1998)
35. S.-M. Jung, *Hyers-Ulam-Rassias Stability of Functional Equations in Mathematical Analysis* (Hadronic Press, Palm Harbor/Florida, 2001)
36. S.-M. Jung, Inequalities for distances between points and distance preserving mappings. Nonlinear Anal. **62**, 675–681 (2005)

37. S.-M. Jung, A fixed point approach to the stability of isometries. J. Math. Anal. Appl. **329**, 879–890 (2007)
38. S.-M. Jung, T.-S. Kim, A fixed point approach to the stability of the cubic functional equation. Bol. Soc. Mat. Mex. **12**(3), 51–57 (2006)
39. S.-M. Jung, T.-S. Kim, K.-S. Lee, A fixed point approach to the stability of quadratic functional equation. Bull. Korean Math. Soc. **43**, 531–541 (2006)
40. S.-M. Jung, K.-S. Lee, An inequality for distances between $2n$ points and the Aleksandrov-Rassias problem. J. Math. Anal. Appl. **324**, 1363–1369 (2006)
41. S.-M. Jung, Th.M. Rassias, On distance-preserving mappings. J. Korean Math. Soc. **41**, 667–680 (2004)
42. R.V. Kadison, J.R. Ringrose, *Fundamentals of the Theory of Operator Algebras* (Academic Press, New York, 1983)
43. A.K. Katsaras, A. Beoyiannis, Tensor products of non-Archimedean weighted spaces of continuous functions. Georgian Math. J. **6**, 33–44 (1999)
44. A. Khrennikov, *Non-Archimedean Analysis: Quantum Paradoxes, Dynamical Systems and Biological Models*. Mathematics and Its Applications, vol. 427 (Kluwer, Dordrecht, 1997)
45. Y. Ma, The Aleksandrov problem for unit distance preserving mapping. Acta Math. Sci. **20**, 359–364 (2000)
46. S. Mazur, S. Ulam, Sur les transformation d'espaces vectoriels normés. C.R. Acad. Sci. Paris **194**, 946–948 (1932)
47. B. Mielnik, Th.M. Rassias, On the Aleksandrov problem of conservative distances. Proc. Am. Math. Soc. **116**, 1115–1118 (1992)
48. D. Miheţ, V. Radu, On the stability of the additive Cauchy functional equation in random normed spaces. J. Math. Anal. Appl. **343**, 567–572 (2008)
49. M. Mirzavaziri, M.S. Moslehian, A fixed point approach to stability of a quadratic equation. Bull. Braz. Math. Soc. **37**, 361–376 (2006)
50. M.S. Moslehian, On the orthogonal stability of the Pexiderized quadratic equation. J. Differ. Equ. Appl. **11**, 999–1004 (2005)
51. M.S. Moslehian, On the stability of the orthogonal Pexiderized Cauchy equation. J. Math. Anal. Appl. **318**, 211–223 (2006)
52. M.S. Moslehian, Th.M. Rassias, Orthogonal stability of additive type equations. Aequ. Math. **73**, 249–259 (2007)
53. M.S. Moslehian, Gh. Sadeghi, A Mazur-Ulam theorem in non-archimedean normed spaces. Nonlinear Anal. – TMA **69**, 3405–3408 (2008)
54. P.J. Nyikos, On some non-Archimedean spaces of Alexandrof and Urysohn. Topol. Appl. **91**, 1–23 (1999)
55. L. Paganoni, J. Rätz, Conditional function equations and orthogonal additivity. Aequ. Math. **50**, 135–142 (1995)
56. C. Park, Fixed points and Hyers-Ulam-Rassias stability of Cauchy-Jensen functional equations in Banach algebras. Fixed Point Theory Appl. **2007**, Art. ID 50175 (2007)
57. C. Park, Generalized Hyers-Ulam-Rassias stability of quadratic functional equations: a fixed point approach. Fixed Point Theory Appl. **2008**, Art. ID 493751 (2008)
58. C. Park, J. Park, Generalized Hyers-Ulam stability of an Euler-Lagrange type additive mapping. J. Differ. Equ. Appl. **12**, 1277–1288 (2006)
59. C. Park, Th.M. Rassias, Isometries on linear n-normed spaces. J. Inequal. Pure Appl. Math. **7**, Art. ID 168 (2006)
60. C. Park, Th.M. Rassias, Additive isometries on Banach spaces. Nonlinear Funct. Anal. Appl. **11**, 793–803 (2006)
61. C. Park, Th.M. Rassias, Inequalities in additive N-isometries on linear N-normed Banach spaces. J. Inequal. Appl. **2007**, Art. ID 70597 (2006)
62. C. Park, Th.M. Rassias, Isometric additive mappings in quasi-Banach spaces. Nonlinear Funct. Anal. Appl. **12**, 377–385 (2007)
63. C. Park, Th.M. Rassias, Isometric additive mappings in generalized quasi-Banach spaces. Banach J. Math. Anal. **2**, 59–66 (2008)

64. C. Park, Th.M. Rassias, d-Isometric linear mappings in linear d-normed Banach modules. J. Korean Math. Soc. **45**, 249–271 (2008)
65. A.G. Pinsker, Sur une fonctionnelle dans l'espace de Hilbert. C. R. (Dokl.) Acad. Sci. URSS, n. Ser. **20**, 411–414 (1938)
66. V. Radu, The fixed point alternative and the stability of functional equations. Fixed Point Theory **4**, 91–96 (2003)
67. Th.M. Rassias, On the stability of the linear mapping in Banach spaces. Proc. Am. Math. Soc. **72**, 297–300 (1978)
68. Th.M. Rassias, Properties of isometic mappings. J. Math. Anal. Appl. **235**, 108–121 (1997)
69. Th.M. Rassias, On the stability of the quadratic functional equation and its applications. Studia Univ. Babeş-Bolyai Math. **43**, 89–124 (1998)
70. Th.M. Rassias, The problem of S.M. Ulam for approximately multiplicative mappings. J. Math. Anal. Appl. **246**, 352–378 (2000)
71. Th.M. Rassias, On the stability of functional equations in Banach spaces. J. Math. Anal. Appl. **251**, 264–284 (2000)
72. Th.M. Rassias, On the A.D. Aleksandrov problem of conservative distances and the Mazur-Ulam theorem. Nonlinear Anal. – TMA **47**, 2597–2608 (2001)
73. Th.M. Rassias (ed.), *Functional Equations, Inequalities and Applications* (Kluwer, Dordrecht/Boston/London, 2003)
74. Th.M. Rassias, P. Šemrl, On the Mazur-Ulam theorem and the Aleksandrov problem for unit distance preserving mapping. Proc. Am. Math. Soc. **118**, 919–925 (1993)
75. J. Rätz, On orthogonally additive mappings. Aequ. Math. **28**, 35–49 (1985)
76. J. Rätz, Gy. Szabó, On orthogonally additive mappings IV. Aequ. Math. **38**, 73–85 (1989)
77. F. Skof, Proprietà locali e approssimazione di operatori. Rend. Semin. Mat. Fis. Milano **53**, 113–129 (1983)
78. K. Sundaresan, Orthogonality and nonlinear functionals on Banach spaces. Proc. Am. Math. Soc. **34**, 187–190 (1972)
79. Gy. Szabó, Sesquilinear-orthogonally quadratic mappings. Aequ. Math. **40**, 190–200 (1990)
80. S.M. Ulam, *Problems in Modern Mathematics* (Wiley, New York, 1960)
81. F. Vajzović, Über das Funktional H mit der Eigenschaft: $(x, y) = 0 \Rightarrow H(x + y) + H(x - y) = 2H(x) + 2H(y)$. Glas. Mat. Ser. III **2**(22), 73–81 (1967)
82. S. Xiang, Mappings of conservative distances and the Mazur-Ulam theorem. J. Math. Anal. Appl. **254**, 262–274 (2001)

Exotic Heat PDE's.II*

Agostino Prástaro

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract Exotic heat equations that allow to prove the Poincaré conjecture and its generalizations to any dimension are considered. The methodology used is the PDE's algebraic topology, introduced by A. Prástaro in the geometry of PDE's, in order to characterize global solutions. In particular it is shown that this theory allows us to identify n-dimensional *exotic spheres*, i.e., homotopy spheres that are homeomorphic, but not diffeomorphic to the standard S^n.

1 Introduction

How exotic are exotic spheres?

The term "exotic sphere" was used by J. Milnor to characterize smooth manifolds that are homotopy equivalent and homeomorphic to S^n, but not diffeomorphic to S^n.[1] This strange mathematical phenomenon, never foreseen before the introduction

*Panos Pardalos and Themistocles M. Rassias (eds.), Essays in Mathematics and its Applications. Dedicated to Stephen Smale, Springer, New York, 2011

[1] In this paper we will use the following notation: $\approx$ homeomorphism; $\cong$ diffeomorphism; $\approxeq$ homotopy equivalence; $\simeq$ homotopy.

A. Prástaro (✉)
Department SBAI, Mathematics (ex MEMOMAT), University of Roma "La Sapienza", Via A. Scarpa, 16, 00161 Roma, Italy
e-mail: agostino.prastaro@uniroma1.it

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_15, © Springer-Verlag Berlin Heidelberg 2012

just by J. Milnor of the famous 7-dimensional exotic sphere [20], has stimulated a lot of mathematical research in algebraic topology. The starting points, were, other than the cited paper by J. W. Milnor, also a joint paper with Kervaire [18] and some papers by Smale [45], Freedman [7] and Cerf [4] on generalizations of the Poincaré conjecture in dimension $n \geq 4$. There the principal mathematical tools utilized were Morse theory (Milnor), h-cobordism theory (Smale), surgery techniques and Hirzebruch signature formula. Surprising, from this beautiful mathematical architecture was remained excluded just the famous Poincaré conjecture for 3-dimensional manifolds. In fact, the surgery techniques do not give enough tools in low dimension ($n < 5$), where surgery obstructions disappear. Really, it was necessary to recast the Poincaré problem as a problem to find solutions in a suitable PDE equation (*Ricci flow equation*), to be able to obtain more informations just on dimension three. (See works by Hamilton [11–15], Perelman [24, 25] and Prástaro [1, 40].) The idea by R.S. Hamilton to recast the problem in the study of the Ricci flow equation has been the real angular stone that has allowed to look to the solution of the Poincaré conjecture from a completely new point of view. In fact, with this new prospective it was possible to G. Perelman to obtain his results and to A. Prástaro to give a new proof of this conjecture, by using his PDE's algebraic topologic theory. To this respect, let us emphasize that the usual geometric methods for PDE's (Spencer, Cartan), were able to formulate for nonlinear PDE's, local existence theorems only, until the introduction, by A. Prástaro, of the algebraic topologic methods in the PDE's geometric theory. These give suitable tools to calculate integral bordism groups in PDE's, and to characterize global solutions. Then, on the ground of integral bordism groups, a new geometric theory of stability for PDE's and solutions of PDE's has been built. These general methodologies allowed to A. Prástaro to solve fundamental mathematical problems too, other than the Poincaré conjecture and some of its generalizations, like characterization of global smooth solutions for the Navier-Stokes equation and global smooth solutions with mass-gap for the quantum Yang-Mills superequation. (See [29–41].[2])

The main purpose of this paper is to show how, by using the PDE's algebraic topology, introduced by A. Prástaro, one can prove the Poincaré conjecture in any dimension for the category of smooth manifolds, but also to identify exotic spheres. In the part I [42] we have just emphasized as in dimension 3, the method followed by A. Prástaro allows us to prove the Poincaré conjecture and to state also that 3-dimensional homotopy spheres are diffeomorphic to S^3. (Related problems are considered there too.) In the framework of the PDE's algebraic topology, the identification of exotic spheres is possible thanks to an interaction between integral bordism groups of PDE's, conservation laws, surgery and geometric topology of

[2] See also [1, 2, 44], where interesting related applications of the PDE's Algebraic Topology are given.

manifolds. With this respect we shall enter in some details on these subjects, in order to well understand and explain the meaning of such interactions. So the paper splits in three sections other this Introduction. 2. Integral bordism groups in Ricci flow PDE's. 3. Morse theory in Ricci flow PDE's. 4. h-Cobordism in Ricci flow PDE's. The main result is contained just in this last section and it is Theorem 30.[3]

2 Integral Bordism Groups in Ricci Flow PDE's

In this section we shall characterize the local and global solutions of the Ricci flow equation, following the geometric approach of some our previous works on this equation [1, 30, 38, 40]. Let M be a n-dimensional smooth manifold and let us consider the following fiber bundle $\bar{\pi}: E \equiv \mathbb{R} \times \widetilde{S_2^0 M} \to \mathbb{R} \times M$, $(t, x^i, y_{ij})_{1 \le i,j \le n} \mapsto (t, x^i) \equiv (x^\alpha)_{0 \le \alpha \le n}$, where $\widetilde{S_2^0 M} \subset S_2^0 M$ is the open subbundle of non-degenerate Riemannian metrics on M. Then the Ricci flow equation is the closed second order partial differential relation, (in the sense of Gromov [10]), on the fiber bundle $\bar{\pi} : E \to \mathbb{R} \times M$, $(RF) \subset JD^2(E)$, defined by the differential polynomials on $JD^2(E)$ given in (1):

$$\left\{ \begin{aligned} F_{jl} &\equiv |y|[y_{ik}](y_{il,jk} + y_{jk,il} - y_{jl,ik} - y_{ik,jl}) \\ &\quad + [y_{ik}][y_{rs}]([jk, r][il, s] - [jl, r][ik, s]) + \frac{|y|^2}{2} y_{jl,t} \\ &\quad \equiv S_{jl}(y_{rs}, y_{rs,\alpha}, y_{rs,pq}) + \frac{|y|^2}{2} y_{jl,t} = 0, \end{aligned} \right. \tag{1}$$

where $[ij, r]$ are the usual Christoffels symbols, given by means of the coordinates $y_{rs,i}$, $|y| = \det(y_{ik})$, and $[y_{ik}]$ is the algebraic complement of y_{ik}. The ideal $\mathfrak{p} \equiv < F_{jl} >$ is not prime in $\mathbb{R}[y_{rs}, y_{rs,\alpha}, y_{rs,ij}]$. However, an irreducible component is described by the system in solved form: $y_{rs,t} = -\frac{2}{|y|^2} S_{jl}$. This is formally integrable and also completely integrable.[4] In fact,

[3] In order to allow a more easy understanding, this paper has been written in a large expository style. (A first version of this paper has been put in arxiv: [43].)

[4] We shall denote with the same symbol (RF) the corresponding algebraic manifold. For a geometric algebraic theory of PDE's see the monograph [29], and references quoted there.

$$\begin{cases}\dim JD^{2+s}(E)=n+1+\sum_{0\le r\le 2+s}\frac{(n+1)n}{2}\frac{(n+r)!}{r!n!}\\ \dim(RF)_{+s}=n+1+\frac{(n+1)n}{2}\left[\sum_{0\le r\le 2+s}\frac{(n+r)!}{r!n!}-\sum_{0\le r'\le s}\frac{(n+r')!}{r'!n!}\right]\\ \dim g_{2+s}=\frac{(n+1)n}{2}\frac{(n+2+s)!}{(2+s)!n!}-\frac{(n+1)n}{2}\frac{(n+s)!}{s!n!}.\\ \end{cases} \tag{2}$$

Therefore, one has: $\dim(RF)_{+s}=\dim(RF)_{+(s-1)}+\dim g_{2+s}$. This assures that one has the exact sequences in (3).

$$(RF)_{+s}\longrightarrow(RF)_{+(s-1)}\longrightarrow 0,\quad s\ge 1. \tag{3}$$

One can also see that the symbol g_2 is not involutive. By the way a general theorem of the formal geometric theory of PDE's assures that after a finite number of prolongations, say s, the corresponding symbol g_{2+s} becomes involutive. (See [9, 29].) Then, taking into account the surjectivity of the mappings (3), we get that (RF) is formally integrable. Furthermore, from the algebraic character of this equation, we get also that is completely integrable. Therefore, in the neighborhood of any of its points $q\in(RF)$ we can find solutions. (These can be analytic ones, but also smooth if we consider to work on the infinity prolongation $(RF)_{+\infty}$, where the Cartan distribution is "involutive" and of dimension $(n+1)$.) Finally, taking into account that $\dim(RF)>2(n+1)+1=2n+3$, we can use Theorem 2.15 in [30] to calculate the n-dimensional singular integral bordism group, $\Omega_{n,s}^{(RF)}$, for n-dimensional closed smooth admissible integral manifolds bording by means of (singular) solutions. (Note that the symbols of (RF) and its prolongations are non-zero.) This group classifies the structure of the global singular solutions of the Ricci-flow equation. One has:

$$\Omega_{n,s}^{(RF)}\cong\bigoplus_{r+s=n}H_r(M;\mathbb{Z}_2)\otimes_{\mathbb{Z}_2}\Omega_s, \tag{4}$$

where Ω_s is the bordism group for s-dimensional closed smooth manifolds.[5]

[5] We used the fact that the fiber of $E\to M$ is contractible.

$$\begin{tikzcd}
0 \arrow[dr] & & & 0 \arrow[d] & \\
& \overline{K}_n^{(RF)} \arrow[rr] \arrow[dr] & & K_{n,s;n}^{(RF)} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & K_{n,s}^{(RF)} \arrow[r] \arrow[u] & \Omega_n^{(RF)} \arrow[r, "a"] \arrow[dr, "c"] & \Omega_{n,s}^{(RF)} \arrow[r] \arrow[d, "b"] & 0 \\
& 0 \arrow[u] \arrow[rr] & & \Omega_n \arrow[d] & \\
& & & 0 &
\end{tikzcd} \tag{5}$$

It is important to underline that with the term "n-dimensional closed smooth admissible integral manifolds" we mean smooth integral manifolds, $N \subset (RF) \subset JD^2(E)$, that diffeomorphically project on their image on E, via the canonical projection $\pi_{2,0}{:}JD^2(E) \to E$. In [40] we have proved, that any smooth section $g{:}M \to \widetilde{S^0_2M}$, identifies a space-like n-dimensional smooth integral manifold $N \subset (RF)$, and that for such a Cauchy manifold pass local smooth solutions, contained in a tubular neigbourhood $N \times [0, \epsilon) \subset (RF)$, for suitable $\epsilon > 0$. Therefore, we can represent any n-dimensional smooth compact Riemannian manifold (M, γ) as a space-like Cauchy manifold $N_0 \subset (RF)_{t_0}$, for some initial time t_0, and ask if there are solutions that bord N_0 with (S^n, γ'), where γ' is the canonical metric of S^n, identified with another space-like Cauchy manifold $N_1 \subset (RF)_{t_1}$, with $t_0 < t_1$. The answer depends on the class of solution that we are interested to have. For weak-singular solutions the corresponding integral bordism group $\Omega_{n,s}^{(RF)}$ is given in (4). The relation with the integral bordism group $\Omega_n^{(RF)}$, for smooth solutions of (RF) is given by the exact commutative diagram (5) where is reported the relation with the bordism group Ω_n for smooth manifolds.

Theorem 1. *Let M in the Ricci flow equation $(RF) \subset JD^2(E) \subset J^2_n(W)$ be a smooth compact n-dimensional manifold homotopy equivalent to S^n. Then the n-dimensional singular integral bordism group of (RF) is given in (6).*

$$\Omega_{n,s}^{(RF)} = \Omega_n \bigoplus \mathbb{Z}_2. \tag{6}$$

Then one has the exact commutative diagram given in (7)

$$\begin{tikzcd}
0 \arrow[dr] & & & 0 \arrow[d] & \\
& \overline{K}_n^{(RF)} \arrow[rr] \arrow[dr] & & K_{n,s;n}^{(RF)} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & K_{n,s}^{(RF)} \arrow[r] \arrow[u] & \Omega_n^{(RF)} \arrow[r, "a"] \arrow[dr, "c"] & \Omega_n \bigoplus \mathbb{Z}_2 \arrow[r] \arrow[d, "b"] & 0 \\
& 0 \arrow[u] \arrow[rr] & & \Omega_n \arrow[d] & \\
& & & 0 &
\end{tikzcd} \tag{7}$$

One has the isomorphisms reported in (8).

$$\left\{\begin{array}{l} \text{(a)} : \Omega_n^{(RF)}/K_{n,s}^{(RF)} \cong \Omega_{n,s}^{(RF)} \\ \\ \text{(b)} : \Omega_n^{(RF)}/\overline{K}_n^{(RF)} \cong \Omega_n \end{array}\right. \tag{8}$$

Proof. Since we have assumed that M is homotopy equivalent to S^n, ($M \approxeq S^n$), we can state that M has the same homology groups of S^n. Therefore we get the isomorphisms reported in (9).

$$H_p(M;\mathbb{Z}_2) \cong H_p(S^n;\mathbb{Z}_2) \cong \left\{\begin{array}{l} \mathbb{Z}_2\,,\ p = 0, n \\ 0\,,\ \text{otherwise} \end{array}\right. \tag{9}$$

Therefore, taking into account (4) we get the isomorphism (6).

Example 1. In Table 1 we report some explicitly calculated cases of integral singular bordism groups for $1 \leq n \leq 7$.

3 Morse Theory in Ricci Flow PDE's

Let us now give the fundamental theorem that describe quantum tunnel effects in solutions of PDEs, i.e., the change of sectional topology in the (singular) solutions of (RF).

Theorem 2 (Topology transitions as quantum tunnel effects in Ricci flow equation). *Let $N_0, N_1 \subset (RF) \subset JD^2(E)$ be space-like Cauchy manifolds of (RF), at two different times $t_0 \neq t_1$. Let $V \subset (RF)$ be a (singular) solution such that*

Table 1 Examples of singular integral bordism groups for n-dimensional homotopy spheres

n	$\Omega_{n,s}^{(RF)}$
1	$\mathbb{Z}_2$
2	$\mathbb{Z}_2 \oplus \mathbb{Z}_2$
3	$\mathbb{Z}_2$
4	$\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$
5	$\mathbb{Z}_2 \oplus \mathbb{Z}_2$
6	$\mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$
7	$\mathbb{Z}_2 \oplus \mathbb{Z}_2$

$\partial V = N_0 \sqcup N_1$. *Then there exists an admissible Morse function* $f : V \to [a, b] \subset \mathbb{R}$ *such that:*

(A) (Simple quantum tunnel effect). *If* f *has a critical point* q *of index* k *then there exists a* k*-cell* $e^k \subset V - N_1$ *and an* $(n-k+1)$*-cell* ${e_*}^{n-k+1} \subset V - N_0$ *such that:*

- (i) $e^k \cap N_0 = \partial e^k$;
- (ii) ${e_*}^{n-k+1} \cap N_1 = \partial {e_*}^{n-k+1}$;
- (iii) *there is a deformation retraction of* V *onto* $N_0 \cup e^k$;
- (iv) *there is a deformation retraction of* V *onto* $N_1 \cup {e_*}^{n-k+1}$;
- (v) ${e_*}^{n-k+1} \cap e^k = q$; ${e_*}^{n-k+1} \pitchfork e^k$.

(B) (Multi quantum tunnel effect). *If* f *is of type* $(\nu_0, \cdots, \nu_{n+1})$ *where* ν_k *denotes the number of critical points with index* k *such that* f *has only one critical values* c, $a < c < b$, *then there are disjoint* k*-cells* $e_i^k \subset V \setminus N_1$ *and disjoint* $(n{-}k{+}1)$*-cells* $(e_*)_i^{n-k+1} \subset V \setminus N_0$, $1 \leq i \leq \nu_k$, $k{=}0, \cdots, n{+}1$, *such that:*

- (i) $e_i^k \cap N_0 = \partial e_i^k$;
- (ii) ${e_*}_i^{n-k+1} \cap N_1 = \partial {e_*}_i^{n-k+1}$;
- (iii) *there is a deformation retraction of* V *onto* $N_0 \bigcup \{\cup_{i,k}(e_*)_i^k\}$;
- (iv) *there is a deformation retraction of* V *onto* $N_1 \bigcup \{\cup_{i,k}(e_*)_i^{n-k}\}$;
- (v) $(e_*)_i^{n-k} \cap e_i^k = q_i$; $(e_*)_i^{n-k+1} \pitchfork e_i^k$.

(C) (No topology transition). *If* f *has no critical point then* $V \cong N_0 \times I$ *where* $I \equiv [0, 1]$.

Proof. The proof can be conducted by adapting to the Ricci flow equation (RF) Theorem 23 in [26]. Let us emphasize here some important lemmas only.

Lemma 1 (Morse-Smale functions).

(1) *On a closed connected compact smooth manifold* M, *there exists a Morse function* $f : M \to \mathbb{R}$ *such that the critical values are ordened with respect to the indexes, i.e.,* $f(x_\lambda) = f(x_\mu)$, *if* $\lambda = \mu$, *and* $f(x_\lambda) > f(x_\mu)$, *if* $\lambda > \mu$, *where* x_λ, *(resp.* x_μ*), is the critical point of* f *with index* λ *(resp.* μ*). Such functions are called* regular functions, *or* Morse-Smale functions, *and are not dense in* $C^\infty(M, \mathbb{R})$, *as Morse functions instead are. Furthermore, such functions can be chosen in such a way that they have an unique maximum point (with index* $\lambda{=}n{=}\dim M$*), and an unique minimum point (with index* $\lambda = 0$*).*

(2) *To such functions are associated vector fields* $\zeta = \mathrm{grad}\, f : M \to TM$ *such that* $\zeta(x_\lambda) = 0$ *iff* x_λ *is a critical point. Then in a neighborhood of a* x_λ*, the integral curves of* ζ *are of two types: ingoing in* x_λ*, and outgoing from* x_λ*. These fit in two different disks* D^λ *and* $D^{n-\lambda}$ *contained in* M *called* separatrix diagram. *(See Fig. 1.)*

Lemma 2 (Morse functions and CW complexes).

(1) *Let* M *be a compact* n*-dimension manifold and* $f : M \to [a,b]$ *an admissible Morse function of type* $(\nu_0, \cdots, \nu_n)$ *such that* $\partial M = f^{-1}(b)$*. Then* M *has the homotopy type of a finite CW complex having* ν_k *cells at each dimension* $k = 0, \cdots, n$*. Furthermore* $(M, \partial M)$ *has the homotopy type of a CW-pair of dimension* n*.*
Furthermore $(M, f^{-1}(a))$ *has the homotopy of relative CW complex having* ν_k *cells of dimension* k*, for each* $k = 0, \cdots, n$*.*[6]

(2) *An* n*-dimension manifold* M *has the homotopy type of a CW complex of dimension* $\leq n$*.*

(3) *Let* M *be a compact manifold and* $N \subset M$ *a compact submanifold with* $\partial N = \partial M = \varnothing$*. Then* (M, N) *has the homotopy type of CW pair.*

(4) *The cell decomposition of a closed connected compact smooth manifold* M*, related to a Morse-Smale function, is obtained attaching step-by-step a cell of higher dimension to the previous ones.*

Lemma 3 (Homological Euler characteristic).

(1) *If the compact solution* V *of* (RF) *is characterized by an admissible Morse function* $f : V \to [a,b]$ *of type* $(\nu_0, \cdots, \nu_{n+1})$*, then its homological Euler characteristic* $\chi_{hom}(V)$ *is given by the formula (10).*

$$\chi_{hom}(V) = \sum_{0 \leq k \leq (n+1)} (-1)^k \beta_k, \quad \beta_k = \dim_F H_k(V, f^{-1}(a); F) \tag{10}$$

where F *is any field.*[7] *Furthermore, if* $f^{-1}(a) = \varnothing$*, then* β_k *in (10) is given by* $\beta_k = \dim_F H_k(V; F)$*, and* $\chi_{hom}(V) = \chi(V)$*.*

(2) *If* M *is a compact odd dimensional manifold with* $\partial M = \varnothing$*, then* $\chi_{hom}(M) = 0$*.*

(3) *Let* M *be a compact manifold such that its boundary can be divided in two components:* $\partial M = \partial_- M \bigcup \partial_+ M$*, then* $\chi_{hom}(M, \partial_+ M) = \chi_{hom}(M, \partial_- M)$*.*

(4) *Let* M *be a compact manifold such that* $\partial M = N_0 \sqcup N_1$*, with* N_i*,* $i = 0, 1$*, disjoint closed sets. Let* $f : M \to \mathbb{R}$ *be a* C^2 *map without critical points, such*

[6] A *relative CW complex* (Y, X) is a space Y and a closed subspace X such that $Y = \bigcup_{r=-1}^{\infty} Y_r$, such that $X = Y_{-1} \subset Y_0 \subset \cdots$, and Y_r is obtained from Y_{r-1} by attaching r-cells.

[7] Let $H_k(Y, X; F)$ denote the singular homology group of the pair (Y, X) with coefficients in the field F. $\beta_k = \dim_F H_k(Y, X; F)$ are called the *F-Betti numbers* of (Y, X). If these numbers are finite and only finitely are nonzero, then the *homological Euler characteristic* of (Y, X) is defined by the formula: $\chi_{hom}(Y, X) = \sum_{0 \leq k \leq \infty} (-1)^k \beta_k$. When Y is a compact manifold

that $f(N_0)=0$, $f(N_1)=1$. Then one has the diffeomorphisms: $M\cong N_0\times I$, $M\cong N_1\times I$.

(5) *Let M be a compact n-dimensional manifold with $\partial M=\varnothing$, such that has a Morse function $f:M\to\mathbb{R}$ with only two critical points. Then M is homeomorphic to S^n.*

Lemma 4. *Let X and Y be closed compact differentiable manifolds without boundaries, then there exists a compact manifold V, such that $\partial V=X\sqcup Y$ iff Y is obtained from X by a sequence of surgeries. (For details see below Theorem 13.)*

Theorem 3 (Smooth solutions and characteristic vector fields). *The characteristic vector field ξ, propagating a space-like n-dimensional smooth, compact, Cauchy manifold $N\subset V$, where V is a smooth solution of (RF), hence a time-like, $(n+1)$-dimensional smooth integral manifold of (RF), cannot have zero points.*

Proof. In fact, the characteristic vector field ξ coincides with the time-like $\zeta_0\equiv\partial x_0+\sum_{|\beta|\geq 0}y^j_{\alpha\beta}\partial y^\beta_j$, where $y^j_{\alpha\beta}$ are determined by the infinity prolongation $(RF)_{+\infty}$ of (RF). Therefore such a vector field cannot have zero points on a compact smooth solution V, of (RF), such that $\partial V=N_0\sqcup N_1$. On the other hand, if $f:V\to\mathbb{R}$ is the Morse function whose gradient gives just the vector field ξ, then f cannot have critical points.[8]

Corollary 4. *A $(n+1)$-dimensional smooth, compact, manifold $V\subset(RF)$, smooth solution of (RF), such that $\partial V=N_0\sqcup N_1$, where N_i, $i=0,1$, are smooth Cauchy manifolds, cannot produce a change of topology from N_0 to N_1, hence these manifolds must necessarily be homeomorphic.*

The following theorem emphasizes the difference between homeomorphic manifolds and diffeomorphic ones.

and X is a compact submanifold, then $\chi_{hom}(Y,X)$ is defined. The evaluation of homological Euler characteristic for topological spaces coincides with that of Euler characteristic for CW complexes X given by $\chi(X)=\sum_{i\geq 0}(-1)^i k_i$, where k_i is the number of cells of dimension i. For closed smooth manifolds M, $\chi(M)$ coincides with the *Euler number*, that is the Euler class of the tangent bundle TM, evalued on the fundamental class of M. For closed Riemannian manifolds, $\chi(M)$ is given as an integral on the curvature, by the *generalized Gauss-Bonnet theorem*: $\chi(V)=\frac{1}{(2\pi)^n}\int_V Pf(\Omega)$, where $\partial V=\varnothing$, $\dim V=2n$, Ω is the curvature of the Levi-Civita connection and $Pf(\Omega)=\frac{1}{2^n n!}\sum_{\sigma\in S_{2n}}\epsilon(\sigma)\prod_{i=1}^n\Omega_{\sigma(2i-1)\sigma(2i)}$, where (Ω_{rs}) is the skew-symmetric $(2n)\times(2n)$ matrix representing $\Omega:V\to\mathfrak{so}(2n)\bigotimes\Lambda^0_2(V)$, hence $Pf(\Omega):V\to\Lambda^0_{2n}(V)$. In Table 2 are reported some important properties of Euler characteristic, that are utilized in this paper.

[8] Let us recall that a compact connected manifold M with boundary $\partial M\neq\varnothing$, admits a nonvanishing vector field. Furthermore, a compact, oriented n-dimensional submanifold $M\subset\mathbb{R}^{2n}$ has a nonvanishing normal vector field. Therefore, above statements about smooth solutions of (RF) agree with well known results of differential topology. (See, e.g., [17]).

Table 2 Euler characteristic χ: properties and examples

Definition	χ	Remarks	Examples
$\partial V = \varnothing$, $\dim V = 2n+1, n \geq 0$	$\chi(V) = 0$	(from Poincaré duality)	
$M \approxeq N$ (homotopy equivalence)	$\chi(M) = \chi(N)$	from $H^{\bullet}(M) \cong H^{\bullet}(N)$	$\chi(pt) = 1$ $\chi(S^n) = 1 + (-1)^n = 0$ $(n = \text{odd}), 2$ $(n = \text{even})$ $\chi(D^3) = \chi(P^3) = \chi(\mathbb{R}^n) = 1$, P^3=convex polyhedron) $\chi(S^2) = \chi(\partial P^3) = 2$, ∂P^3=surface convex polyhedron)
$V = M \sqcup N$	$\chi(V) = \chi(M) + \chi(N)$	(from homology additivity)	$\chi(\underbrace{S^2 \sqcup \cdots \sqcup S^2}_{n}) = 2n$
excision couple (M, N)	$\chi(M \cup N) = \chi(M)$ $+\chi(N) - \chi(M \cap N)$		$\chi(S^2) = \chi(D^2) + \chi(D^2) - \chi(S^1) = 1 + 1 - 0 = 2$ $\chi(K_{lb}) = \chi(M_{ob}) + \chi(M_{ob}) - \chi(S^1) = 0 + 0 - 0 = 0$ $\chi(C_{rc} \bigcup_{S^1} D^2 = \mathbb{R}P^2)$=$\chi(M_{ob})$+$\chi(D^2) - \chi(S^1) = 0$+1−0=1
$V = M \times N$	$\chi(V) = \chi(M).\chi(N)$		$\chi(T^n) = \chi(\underbrace{S^1 \times \cdots \times S^1}_{n}) = 0$
$p : V \to M$ orientable fibration over field with fibre F M path-connected	$\chi(V) = \chi(F).\chi(M)$	(from Serre-spectral sequence) (also from transfer map) $(\tau : H_{\bullet}(M) \to H_{\bullet}(V))$ $(p_{\bullet} \circ \tau = \chi(F).1_{H_{\bullet}(M)})$	$S^n \to \mathbb{R}P^n$: $\chi(S^n) = \chi(\{1, -1\}).\chi(\mathbb{R}P^n) = 2.\chi(\mathbb{R}P^n)$ $\chi(\mathbb{R}P^n) = 0$ $(n = \text{odd}), 1$ $(n = \text{even})$.
$p : V \to M$ k-sheeted covering	$\chi(V) = k.\chi(M)$		$p : M_{ob} \to S^1$: M_{ob}=Möbius strip: $\chi(M_{ob}) = 2\chi(S^1) = 0$
$V = \partial M$, $\dim M = 2n, n \geq 0$	$\chi(V) = 2m, m \geq 0$	(from excision couple)	$\mathbb{R}P^{2n} \neq \partial M$, since $\chi(\mathbb{R}P^{2n}) = 1$

$\chi : \Omega^{O}_{2i} \to \mathbb{Z}$ is a surjective mapping for $i \geq 1$ and an isomorphism for $i = 1$
$\chi(\partial P^3) = V - E + F$, V=vertex-number, E=edge-number, F = face-number
Closed oriented surfaces: $\chi = 2 - 2g$, g = genus, (number of handles)
Closed nonorientable surfaces: $\chi = 2 - \kappa$, κ=nonorientable genus, (number of real projective planes in a connected decomposition)
Examples of nonorientable surfaces: K_{lb}= Klein bottle $(\partial K_{lb} = \varnothing)$; M_{ob} = Möbius strip $(\partial M_{ob} = S^1)$; $\mathbb{R}P^2$ = Projective plane $(\partial \mathbb{R}P^2 = \varnothing)$
Examples of nonorientable surfaces: $C_{rc} \approxeq M_{ob}$ = cross-cap: surface homotopy equivalent to Möbius strip

Theorem 5 (Exotic differentiable structures on compact smooth manifolds). *Let M and N be n-dimensional homeomorphic compact smooth manifolds. Then it does not necessitate that M is diffeomorphic to N.*

Proof. Since M is considered homeomorphic to N, there exist continuous mappings $f : M \to N$ and $g : N \to M$, such that $g \circ f = id_M$, and $f \circ g = id_N$. Let us consider, now, the following lemma.

Lemma 5. *Let M and N be C^s manifolds, $1 \leq s \leq \infty$, without boundary. Then $C^s(M, N)$ is dense in $C^r_S(M, N)$, (in the strong topology), $0 \leq r < s$.*

Proof. See, e.g., [17].

From Lemma 5 we can state that the above continuous mappings f and g can be approximated with differentiable mapping, but these do no necessitate to be diffeomorphisms. In fact we have the following lemma.

Lemma 6. *Let $G^k(M, N) \subset C^k(M, N)$, $k \geq 1$, denote any one of the following subsets: diffeomorphisms, embeddings, closed embeddings, immersions, submersions, proper maps. Let M and N be compact C^s manifolds, $1 \leq s \leq \infty$, without boundary. Then $G^s(M, N)$ is dense in $G^r(M, N)$ in the strong topology, $1 \leq r < s$. In particular, M and N are C^s diffeomorphic iff they are C^r diffeomorphic with $r \geq 1$.*

Proof. See, e.g., [17].

Above lemma can be generalized also to compact manifolds with boundary. In fact, we have the following lemma.

Lemma 7. *Let us consider compact manifolds with boundary. Then the following propositions hold.*

(i) *Every C^r manifold M, $1 \leq r < \infty$, is C^∞ diffeomorphic to a C^∞ manifold and the latter is unique up to C^∞ diffeomorphisms.*
(ii) *Let $(M, \partial M)$ and $(N, \partial N)$, be C^s manifold pairs, $1 \leq s \leq \infty$. Then, the inclusion $C^s(M, \partial M; N, \partial N) \hookrightarrow C^r(M, \partial M; N, \partial N)$, $0 \leq r < s$, is dense in the strong topology. If, $1 \leq r < s$ and $(M, \partial M)$ and $(N, \partial N)$ are C^r diffeomorphic, they are also C^s diffeomorphic.*

Proof. See, e.g., [17].

Therefore, it is not enough to assume that compact smooth manifolds should be homeomorphic in order to state that they are also diffeomorphic, hence the proof of Theorem 5 is complete. (To complement Theorem 5 see also Lemmas 12 and 13 below.)

From Theorem 5 we are justified to give the following definition.

Definition 1. Let M and N be two n-dimensional smooth manifolds that are homeomorphic but not diffeomorphic. Then we say that N is an *exotic substitute* of M.

Example 2. The sphere S^7 has 28 exotic substitutes, just called *exotic* 7*-dimensional spheres*. (See [18, 20].) These are particular 7-dimensional manifolds, built starting from oriented fiber bundle pairs over S^4. More precisely let us consider $(D^4, S^3) \to (W, V) \to S^4$. The 4-plane bundle $D^4 \to S^4$ is classified by the isomorphism

$$[S^4, BSO(4)] \cong \mathbb{Z} \bigoplus \mathbb{Z},$$

given by $\omega \mapsto (\frac{1}{4}(2\chi(\omega) + p_1(\omega)), \frac{1}{4}(2\chi(\omega) - p_1(\omega)))$, where $\chi(\omega), p_1(\omega) \in H^4(S^4) = \mathbb{Z}$ are respectively the Euler number and the Pontrjagin class of ω, related by the congruence $p_1(\omega) = 2\chi(\omega) \pmod 4$. Let us denote by $(W(\omega), V(\omega))$ the above fiber bundle pair identified by ω. The homology groups of $V(\omega) \to S^4$, are given in (11).[9]

$$H_p(V(\omega)) = \begin{cases} \mathbb{Z} \ \text{ if } p = 0, 7 \\ \text{coker}\,(\chi(\omega) : \mathbb{Z} \to \mathbb{Z}) \ \text{ if } p = 3 \\ \ker(\chi(\omega) : \mathbb{Z} \to \mathbb{Z}) \ \text{ if p = 4} \\ 0 \ \text{ otherwise.} \end{cases} \tag{11}$$

The Euler number $\chi(\omega)$ is the Hopf invariant of $J(\omega) \in \pi_7(S^4)$, i.e., $\chi(\omega) = \mathit{Hopf}(J(\omega)) \in \mathbb{Z}$. If $\chi(\omega) = 1 \in \mathbb{Z}$, then $V(\omega)$ is a homotopy 7-sphere which is boundary of an oriented 8-dimensional manifold $W(\omega)$. In fact ${}^{+}\Omega_7 = 0$ and for $\chi(\omega) = 1 \in \mathbb{Z}$ one has $H_p(V(\omega)) = H_p(S^7)$. Let k be an odd integer and let $\omega_k : S^4 \to BSO(4)$ be the classifying map for orientable 4-plane bundle over S^4 with $p_1(\omega_k) = 2k$, $\chi(\omega_k) = 1 \in \mathbb{Z}$. There exists a Morse function $V(\omega_k) \to \mathbb{R}$ with two critical points, such that $V(\omega_k) \setminus \{pt\} \cong \mathbb{R}^7$, hence $V(\omega_k)$ is homeomorphic to S^7. Let us investigate under which conditions $V(\omega_k)$ is diffeomorphic to S^7 too. So let us assume that such diffeomorphism $f : V(\omega_k) \cong S^7$ exists. Then let us consider the closed oriented 8-dimensional manifold $M = W(\omega_k) \bigcup_f D^8$. For such a manifold we report in (12) its intersection form and signature.

$$\begin{cases} (H^4(M), \lambda) = (\mathbb{Z}, 1) \\ \sigma(M) = \sigma(H^4(M), \lambda) = 1. \end{cases} \tag{12}$$

By the Hirzebruch signature theorem one has $\sigma(M) =< \mathcal{L}_2(p_1, p_2), [M] >= 1 \in \mathbb{Z}$, with $< \mathcal{L}_2(p_1, p_2), [M] >= \frac{1}{45}(7p_2(M) - p_1(M)^2) = 1 \in H^8(M) = \mathbb{Z}$, $p_1(M) = 2k$, $p_2(M) = \frac{1}{7}(45 + 4k^2) = \frac{4}{7}(k^2 - 1) + 7 \in H^4(M) = \mathbb{Z}$. Since $p_2(M)$ is an integer, it follows that must be $k^2 \equiv 1 \pmod 7$. This condition on k, comes from the assumption that $V(\omega_k)$ is diffeomorphic to S^7, therefore, it follows that under the condition $k^2 \not\equiv 1 \pmod 7$, $V(\omega_k)$ can be only homeomorphic to S^7,

[9] Recall that an odd dimensional oriented compact manifold M, with $\partial M = \varnothing$ has $\chi(M) = 0$. In particular $\chi(S^{2k+1}) = 0$, instead $\chi(S^{2k}) = 2$. Furthermore, if M and N are compact oriented manifolds with $\partial M = \partial N = \varnothing$, then $\chi(M \times N) = \chi(M)\chi(N)$.

but not diffeomorphic, hence it is an exotic sphere, and M is only a 8-dimensional topological manifold, to which the Hirzebruch signature theorem does not apply.

Example 3. The 4-dimensional affine space $\mathbb{R}^4$ has infinity exotic substitutes, just called *exotic* $\mathbb{R}^4$. (See [5,7].)

The surgery theory is a general algebraic topological framework to decide if a homotopy equivalence between n-dimensional manifolds is a diffeomorphism. (See, e.g., [49].) We shall resume here some definitions and results about this theory. In the following section we will enter in some complementary informations and we will continue to develop such approach in connection with other algebraic topological aspects.

Definition 2. An *n-dimensional geometric Poincaré complex* is a finite CW complex such that one has the isomorphism $H^p(X;\Lambda) \cong H_{n-p}(X;\Lambda)$, induced by the cap product, i.e., $[\omega] \mapsto [X] \cap [\omega]$, for every $\mathbb{Z}[\pi_1(X)]$-module Λ.

Theorem 6 (Geometric Poincaré complex properties).

(1) *An n-dimensional manifold is an n-dimensional geometric Poincaré complex.*

(2) *Let X be a geometric Poincaré complex and Y another CW complex homotopy related to X. Then also Y is a geometric Poincaré complex.*

(3) *Any CW complex homotopy equivalent to a manifold is a geometric Poincaré complex.*

(4) *Geometric Poincaré complexes that are not homotopy equivalent to a manifold may be obtained by gluing together n-dimensional manifolds with boundary, $(M, \partial M)$, $(N, \partial N)$, having an homotopic equivalence on the boundaries, $f : \partial M \approxeq \partial N$, which is not homotopic to a diffeomorphism.*

(5) (Transfer or Umkehr map). *Let $f : N \to M$ be a mapping between oriented, compact, closed manifolds of arbitrary dimensions. Then the Poincaré duality identifies an homomorphism $\tau : H^\bullet(N;\mathbb{Z}) \to H^{\bullet-d}(M;\mathbb{Z})$, where $d = \dim N - \dim M$. More precisely one has the commutative diagram (13) that defines τ.*

$$\begin{tikzcd}
H^\bullet(N;\mathbb{Z}) \arrow[r, "\tau"] \arrow[d, "D_N"'] & H^{\bullet-d}(M;\mathbb{Z}) \\
H_{\dim N-\bullet}(N;\mathbb{Z}) \arrow[r, "f_*"'] & H_{\dim M-\bullet}(M;\mathbb{Z}) \arrow[u, "D_M^{-1}"]
\end{tikzcd} \tag{13}$$

where $d = \dim N - \dim M$. D_N and D_M are the Poincaré isomorphisms on N and M respectively. One has $\tau(f^*(x) \cup y) = x \cup \tau(y)$, $\forall x \in H^\bullet(M;\mathbb{Z})$ and $y \in H^\bullet(N;\mathbb{Z})$.[10] *In particular when $f : \widetilde{M} \to M$ is a covering map, then one can write $\tau(x)(\sigma) = x(\sum_{f(\widetilde{\sigma})=\sigma})\widetilde{\sigma}$, $\forall x \in C^\bullet(\widetilde{M})$ and $\sigma \in C_\bullet(M)$.*

[10] If $f : N \to M$ is an orientable fiber bundle with compact, orientable fiber F, integration over the fiber provides another definition of the transfer map: $\tau : H^\bullet_{de-Rham}(N) \to H_{de-Rham}(M)^{\bullet-r}$, where $r = \dim F$.

Definition 3. Let X be a closed n-dimensional geometric Poincaré complex. A *manifold structure* (M, f) on X is a closed n-dimensional manifold M together with a homotopy equivalence $f : M \approxeq X$. We say hat such two manifold structures (M, f), (N, g) on X are equivalent if there exists a bordism $(F; f, g) : (V; M, N) \to X \times (I; \{0\}, \{1\})$, with F a homotopy equivalence. (This means that $(V; M, N)$ is an h-cobordism, (see the next section).) Let $\mathfrak{S}(X)$ denote the set of such equivalence classes. We call $\mathfrak{S}(X)$ the *manifold structure set* of X. $\mathfrak{S}(X) = \varnothing$ means that X is without manifold structures.

Theorem 7 (Manifold structure set properties).

(1) *$\mathfrak{S}(X)$ is homotopy invariant of X, i.e., a homotopy equivalence $f : X \approxeq Y$ induces a bijection $\mathfrak{S}(X) \to \mathfrak{S}(Y)$.*
(2) *A homotopy equivalence $f : M \approxeq N$ of n-dimensional manifolds determines an element $(M, f) \in \mathfrak{S}(N)$, such that f is h-cobordant to* $1 : N \to N$ *iff* $(M, f) \in [(N, 1)] \in \mathfrak{S}(N)$.
(3) *Let M be a n-dimensional closed differentiable manifold. If $\mathfrak{S}(X) = \{pt\}$ then M does not admit exotic substitutes.*
(4) (Differential structures by gluing manifolds together). *Let M and N be n-dimensional manifolds such that their boundary are diffeomorphic: $\partial M \cong \partial N$. Let α and β be two differential structures on $M \bigcup_f N$ that agree with the differential structures on M and N respectively. Then there exists a diffeomorphism $h : W_\alpha \cong W_\beta$ such that $h|_M = 1_M$.*

Definition 4. A *degree* 1 *normal map* from an n-dimensional manifold M to an n-dimensional geometric Poincaré complex X is given by a couple (f, b), where $f : M \to X$ is a mapping such that $f_*[M] = [X] \in H_n(X)$, and $b : \nu_M \to \eta$ is a stable bundle map over f, from the stable normal bundle $\nu_M : M \to BO$ to the stable bundle $\eta : X \to BO$. We write also $(f, b) : M \to X$.

Theorem 8 (Obstructions for manifold structures on geometric Poincaré complex).

(1) *Let X be an n-dimensional geometric Poincaré complex. Then the criterion to decide if X is homotopy equivalent to an n-dimensional manifold M, is to verify that are satisfied the following two conditions.*

(i) *X admits a degreee* 1 *normal map $(f, b) : M \to X$. This is the case when the map $t(\nu_X) : X \to B(G/O)$, given in (14), is null-homotopic.*

$$\begin{tikzcd} X \arrow[r] \arrow[rr, bend right, "t(\nu_X)"'] & BG \arrow[r] & B(G/O) \end{tikzcd} \tag{14}$$

Then there exists a null-homotopy $t(\nu_X) \simeq \{*\}$ iff the Spivak normal fibration $\nu_X : X \to BG = \lim\limits_{\overrightarrow{k}} BG(k)$ admits a vector bundle reduction $\widetilde{\nu_X} : X \to BO$.

Table 3 Simply connected surgery obstruction groups

$n \pmod 4$	0	1	2	3
$L_n(\mathbb{Z})$	$\mathbb{Z}$	0	$\mathbb{Z}_2$	0

(ii) $(f, b) : M \to X$ *is bordant to a homotopy equivalence* $(g, h) : N \approxeq X$.

(2) (J. H. C. Whitehad's theorem). $f : M \to X$ *is a homotopy equivalence iff* $\pi_*(f) = 0$. *Let* $n = 2k$, *or* $n = 2k + 1$. *It is always possible to kill* $\pi_i(f)$, *for* $i \leq k$, *i.e., there is a bordant degree* 1 *normal map* $(h, b) : N \to X$, *with* $\pi_i(h) = 0$ *for* $i \leq k$. *There exists a normal bordism of* (f, b) *to a homotopy equivalence iff it is also possible kill* $\pi_{k+1}(h)$. *In general there exists an obstruction to killing* $\pi_{k+1}(h)$, *which for* $n \geq 5$ *is of algebraic nature.*

(3) (C. T. C. Wall's surgery obstruction theorem).[49] *For any group* π *there are defined algebraic L-groups* $L_n(\mathbb{Z}[\pi])$ *depending only on* n(mod 4) *as group of stable isomorphism classes of* $(-1)^k$*-quadratic forms over* $\mathbb{Z}[\pi]$ *for* $n = 2k$, *or as group of stable automorphisms of such forms for* $n = 2k + 1$. *An* n*-dimensional degree* 1 *normal map* $(f, b) : N \to X$ *has a* surgery obstruction $\sigma_*(f, b) \in L_n(\mathbb{Z}[\pi_1(X)])$, *such that* $\sigma_*(f, b) = 0$ *if (and for* $n \geq 5$ *only if)* (f, b) *is bordant to a homotopy equivalence.*

Example 4. The simply-connected surgery obstruction groups are given in Table 3. In particular, we have the following.

- The surgery obstruction of a $4k$-dimensional normal map $(f, b) : M \to X$ with $\pi_1(X) = \{1\}$ is $\sigma_*(f, b) = \frac{1}{8}\sigma(K_{2k}(M), \lambda) \in L_{4k}(\mathbb{Z}) = \mathbb{Z}$, with λ the nonsingular symmetric form on the middle-dimensional homology kernel $\mathbb{Z}$-module
$$K_{2k}(M) = \ker(f_* : H_{2k}(M) \to H_{2k}(X)).$$
- The surgery obstruction of a $(4k + 2)$-dimensional normal map $(f, b) : M \to X$ with $\pi_1(X) = \{1\}$ is $\sigma_*(f, b) = Arf(K_{2k+1}(M; \mathbb{Z}_2), \lambda, \mu) \in L_{4k+2}(\mathbb{Z}) = \mathbb{Z}_2$, with λ, μ the nonsingular quadratic form on the middle-dimensional homology $\mathbb{Z}_2$-coefficient homology kernel $\mathbb{Z}_2$-module
$$K_{2k+1}(M; \mathbb{Z}_2) = \ker(f_* : H_{2k+1}(M; \mathbb{Z}_2) \to H_{2k+1}(X; \mathbb{Z}_2)).$$

Theorem 9 (Browder-Novikov-Sullivan-Wall's surgery exact sequence). *One has the following propositions.*

(i) *Let* X *be an* n*-dimensional geometric Poincaré complex with* $n \geq 5$. *The manifold structure set* $\mathfrak{S}(X) \neq \varnothing$ *iff there exists a normal map* $(f, b) : M \to X$ *with surgery obstruction* $\sigma_*(f, b) = 0 \in L_n(\mathbb{Z}[\pi_1(X)])$.

(ii) *Let* M *be an* n*-dimensional manifold. Then* $\mathfrak{S}(M)$ *fits into the surgery exact sequence of pointed sets reported in (15).*

$$\cdots L_{n+1}(\mathbb{Z}[\pi_1(M)]) \longrightarrow \mathfrak{S}(M) \longrightarrow [M, G/O] \to L_n(\mathbb{Z}[\pi_1(M)]) . \tag{15}$$

4 h-Cobordism in Ricci Flow PDE's

In this section we shall relate the h-cobordism with the geometric properties of the Ricci flow equation considered in the previous two sections. With this respect let us recall first some definitions and properties about surgery on manifolds.

Definition 5. (1) The *n-dimensional handle*, of *index p*, is $h^p \equiv D^p \times D^{n-p}$. Its *core* is $D^p \times \{0\}$. The *boundary of the core* is $S^{p-1} \times \{0\}$. Its *cocore* is $\{0\} \times D^{n-p}$ and its *transverse sphere* is $\{0\} \times S^{n-p-1}$.

(2) Given a topological space Y, the images of continuous maps $D^n \to Y$ are called the *n-cells* of Y.

(3) Given a topological space X and a continuous map $\alpha : S^{n-1} \to X$, we call $Y \equiv X \bigcup_\alpha D^n$ obtained from X by *attaching a n-dimensional cell* to X.

(4) We call *CW-complex* a topological space X obtained from $\varnothing$ by successively attaching cells of non-decreasing dimension:

$$X \equiv (\cup D^0) \cup D^1 \cup D^2) \cup \cdots \tag{16}$$

We call $X^n \equiv \bigcup_{1 \leq i \leq n} D^i$, $n \geq 0$, the *(n)-skeleta*.

Definition 6. *(Homotopy groups.)*

(1) We define *homotopy groups* of manifold, (resp. CW-complex), M, the groups

$$\pi_p(M) = [S^p, M],\ p \geq 0. \tag{17}$$

(2) Let X be a manifold over a CW-complex and an element $x \in \pi_n(X)$, $n \geq 1$. Let $Y = X \bigcup_{\varphi^n} D^{n+1}$ be the CW-complex obtained from X by attaching an $(n+1)$-cell with map $\varphi^n : S^n \to X$, with $x = [\varphi^n] \in \pi_n(X)$. The operation of attaching the $(n+1)$-cell is said to *kill x*.

Theorem 10. (CW-substitute)

(1) *For any manifold, M, we can construct a CW-complex X and a weak homotopy equivalence $f : X \to M$, (i.e., the induced maps $f_* : \pi_r(X') \to \pi_r(X)$ on the Hurewicz homotopy groups are bijective for $r \geq 0$).*[11] *Then X' is called the CW-substitute of X. This is unique up to homotopy.*

(2) *Furthermore if $h : X \to Y$ is a continuous map between manifolds, and (X', f), (Y', g) are the corresponding CW-substitutes, then we can find a cellular map $h : X' \to Y'$ so that the following diagram is commutative:*

[11] Note that an homotopy equivalence is an weak homotopy equivalence, but the vice versa is not true. Recall that two pointed topological spaces (X, x_0) and (Y, y_0) have the same homotopy type if $\pi_1(X, x_0) \cong \pi_1(Y, y_0)$, and $\pi_n(X, x_0)$ and $\pi_n(Y, y_0)$ are isomorphic as modules over $\mathbb{Z}[\pi_1(X, x_0)]$ for $n \geq 2$. A *simply homotopy equivalence* between m-dimensional manifolds, (or finite CW complexes), is a homotopy equivalence $f : M \approxeq N$ such that the Whitehead

$$\begin{tikzcd} X' \arrow[r, "f"] \arrow[d, "h'"'] & X \arrow[d, "h"] \\ Y' \arrow[r, "g"'] & Y \end{tikzcd} \tag{18}$$

h' is unique up to homotopy.

Definition 7. An *n-dimensional bordism* $(W; M_0, f_0; M_1, f_1)$ consists of a compact manifold W of dimension n, and closed $(n-1)$-dimensional manifolds M_0, M_1, such that $\partial W = N_0 \sqcup N_1$, and diffeomorphisms $f_i : M_i \cong N_i$, $i = 0, 1$. An *n-dimensional h-bordism* (resp. *s-bordism*) is a n-dimensional bordism as above, such that the inclusions $N_i \hookrightarrow W$, $i = 0, 1$, are homotopy equivalences (resp. simply homotopy equivalences).[12] W is a *trivial h-bordism* if $W \cong M_0 \times [0, 1]$. In such a case M_0 is diffeomorphic to M_1: $M_0 \cong M_1$

We will simply denote also by $(W; M_0, M_1)$ a n-dimensional bordism.

If $\varphi^p : S^{p+1} \times D^{n-p-1} \to M_1$ is an embedding, then

$$W + (\varphi^p) \equiv W \bigcup_{\varphi^p} D^p \times D^{n-p} \equiv W \bigcup h^p \tag{19}$$

is said obtained from W by *attaching a handle*, $h^p \equiv D^p \times D^{n-p}$, of *index p* by φ^p.[13] Put $\partial(W + \varphi^p)_0 = M_0$, $\partial(W + \varphi^p)_1 = \partial(W + \varphi^p) - M_0$.

Theorem 11 (CW-substitute of manifold and Hurewicz morphisms). *For any manifold M we can construct a CW-complex M' and a weak homotopy equivalence*

torsion $\tau(f) \in Wh(\pi_1(M))$, where $Wh(\pi_1(M))$ is the Whitehead group of $\pi_1(M)$. With this respect, let us recall that if A is an associative ring with unity, such that A^m is isomorphic to A^n iff $m = n$, put $GL(A) \equiv \bigcup_{n-1} GL_n(A)$, the *infinite general linear group of A* and $E(A) \equiv [GL(A), GL(A)] \lhd GL(A)$. $E(A)$ is the normal subgroup generated by the elementary matrices $\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$. The *torsion group* $K_1(A)$ is the abelian group $K_1(A) = GL(A)/E(A)$. Let $A^\bullet$ denote the multiplicative group of units in the ring A. For a commutative ring A, the inclusion $A^\bullet \hookrightarrow K_1(A)$ splits by the determinant map $\det : K_1(A) \to A^\bullet$, $\tau(\varphi) \mapsto \det(\varphi)$ and one has the splitting $K_1(A) = A^\bullet \bigoplus SK_1(A)$, where $SK_1(A) = \ker(\det : K_1(A) \to A^\bullet)$. If A is a field, then $K_1(A) \cong A^\bullet$ and $SK_1(A) = 0$. The *torsion* $\tau(f)$ of an isomorphism $f : L \cong K$ of finite generated free A-modules of rank n, is the torsion of the corresponding invertible matrix $(f^i_j) \in GL_n(A)$, i.e., $\tau(f) = \tau(f^j_i) \in K_1(A)$. The isomorphism is *simple* if $\tau(f) = 0 \in K_1(A)$. The Whitehead group of a group G is the abelian group $Wh(G) \equiv K_1(\mathbb{Z}[G])/\{\tau(\mp g)|g \in G\}$. $Wh(G) = 0$ in the following cases: (a) $G = \{1\}$; (b) $G = \pi_1(M)$, with M a surface; (c) $G = \mathbb{Z}^m$, $m \geq 1$. There is a conjecture, (Novikov) that extends the case (b) also to m-dimensional compact manifolds M with universal cover $\widetilde{M} = \mathbb{R}^m$. This conjecture has been verified in many cases [6].

[12] Let us emphasize that to state that the inclusions $M_i \hookrightarrow W$, $i = 0, 1$, are homotopy equivalences is equivalent to state the M_i are deformation retracts of W.

[13] In general $W \bigcup h^p$ is not a manifold but a CW-complex.

$f : M' \to M$. Then M' is called the CW-substitute *of M. M' is unique up to homotopy. Then the homotopy groups of M and M' are isomorphic, i.e., one has the top horizontal exact short sequence reported in the commutative diagram (20). There the vertical lines represent the Hurewicz morphisms relating homotpy groups and homology groups.*

$$\begin{tikzcd}
0 \arrow[r] & \pi_p(M) \arrow[r] \arrow[d, "a"'] & \pi_p(M')) \arrow[r] \arrow[d, "a'"'] & 0 \\
0 \arrow[r] & H_p(M) \arrow[r] & H_p(M') \arrow[r] & 0
\end{tikzcd} \tag{20}$$

If M is $(n-1)$-connected, $n \geq 2$, then the morphisms a, a', become isomorphisms for $p \leq n$ and epimorphisms for $p = n + 1$.

We call the morphisms a and a' the Hurewicz morphisms *of the manifold M and M' respectively.*

Definition 8. A *p-surgery* on a manifold M of dimension n is the procedure of construction a new n-dimensional manifold[14]:

$$N \equiv \overline{(M \setminus S^p \times D^{n-p})} \bigcup_{S^p \times S^{n-p-1}} D^{p+1} \times S^{n-p-1}. \tag{21}$$

Example 5. Since for the n-dimensional sphere S^n we can write

$$\begin{aligned} S^n &= \partial D^{n+1} = \partial(D^{p+1} \times D^{n-p}) \\ &= S^p \times D^{n-p} \bigcup D^{p+1} \times S^{n-p-1} \end{aligned} \tag{22}$$

it follows that the surgery removing $S^p \times D^{n-p} \subset S^n$ converts S^n into the product of two spheres

$$D^{p+1} \times S^{n-p-1} \bigcup_{S^p \times S^{n-p-1}} D^{p+1} \times S^{n-p-1} = S^{p+1} \times S^{n-p-1}. \tag{23}$$

Theorem 12 (Surgery and Euler characteristic).

(1) *Let M be a $2n$-dimensional smooth manifold and let apply to N obtained by M with a p-surgery as defined in (21). Then the Euler characteristic of N is related to the M one, by the relation reported in (24).*

$$\chi(N) = \begin{cases} \chi(M) + 2 & p = \text{odd} \\ \chi(M) - 2 & p = \text{even.} \end{cases} \tag{24}$$

[14] We say also that a p-surgery removes a *framed p-embedding* $g : S^p \times D^{n-p} \hookrightarrow M$. Then it kills the homotopy class $[g] \in \pi_p(M)$ of the core $g = g| : S^p \times \{0\} \hookrightarrow M$.

(2) *Let $M = 2n+1$, $n \geq 0$. If $M = \partial V$, then V can be chosen a manifold with $\chi(V) = 0$, i.e., having the same Euler characteristic of M.*
(3) *Let $M = 2n$, $n \geq 0$. If $M = \partial V$, then $\chi(M) = 2\chi(V)$.*

Proof. (1) Let us first note that we can write $M = \overline{(M \setminus S^p \times D^{n-p})} \bigcup (S^p \times D^{2n-p})$, hence we get

$$\begin{cases} \chi(M) = \chi(\overline{(M \setminus S^p \times D^{n-p})}) + \chi(S^p \times D^{2n-p}) \\ \qquad\quad = \chi(\overline{M \setminus S^p \times D^{n-p}}) + (1 + (-1)^p). \end{cases} \tag{25}$$

From (25) we get

$$\chi(\overline{M \setminus S^p \times D^{n-p}}) = \chi(M) - (1 + (-1)^p) = \begin{cases} \chi(M) & p = \text{odd} \\ \chi(M) - 2 & p = \text{even}. \end{cases} \tag{26}$$

On the other hand one has

$$\begin{cases} \chi(N) = \chi(\overline{M \setminus S^p \times D^{n-p}}) + \chi(D^{p+1} \times S^{2n-p-1}) - \chi(S^p \times S^{2n-p-1}) \\ \qquad = \chi(\overline{M \setminus S^p \times D^{n-p}}) + \begin{Bmatrix} 2 & p = \text{odd} \\ 0 & p = \text{even} \end{Bmatrix}. \end{cases} \tag{27}$$

Then from (26) and (27) we get

$$\chi(N) = \begin{cases} \chi(M) + 2 & p = \text{odd} \\ \chi(M) - 2 & p = \text{even}. \end{cases} \tag{28}$$

(2) In fact, if $n = 1$ then M can be considered the boundary of a Möbius strip M_{ob}, that has just $\chi(M_{ob}) = 0$. If $n \geq 3$, and $\chi(V) = 2q$, we can add to V q times p-surgeries with p even in order to obtain a manifold V' that has the same dimension and boundary of V but with Euler characteristic zero. Furthermore, if $\chi(V) = 2q + 1$, we consider the manifold $V'' = V \sqcup \mathbb{R}P^{2n+1}$ that has the same dimension and boundary of V, but $\chi(V'')$ is even. Then we can proceed as before on V''.
(3) Let us consider $V' = V \bigcup_M V$. Then one has $\chi(V') = 0 = 2\chi(V) - \chi(M)$.

Example 6. A *connected sum* of connected n-dimensional manifolds M and N is the connected n-dimensional manifold

$$M \sharp N = (M \setminus D^n) \bigcup (S^{n-1} \times D^1) \bigcup (N \setminus D^n). \tag{29}$$

$M \sharp N$ is the effect of the 0-surgery on the disjoint union $M \sqcup N$ which removes the framed 0-embedding $S^0 \times D^n \hookrightarrow M \times N$ defined by the disjoint union of the embeddings $D^n \hookrightarrow M$, $D^n \hookrightarrow N$.

Example 7. Given a $(n+1)$-dimensional manifold with boundary $(M, \partial M)$ and an embedding $S^{i-1} \times D^{n-i+1} \hookrightarrow \partial M$, $0 \leq i \leq n+1$, we define the $(n+1)$-dimensional manifold $(W, \partial W)$ obtained from M by attaching a i-handle:

$$W = M \bigcup_{S^{i-1} \times D^{n-i+1}} D^i \times D^{n-i+1} = M \bigcup h^i. \tag{30}$$

Then ∂W is obtained from ∂M by an $(i-1)$-surgery:

$$\partial W = (\partial M \setminus S^{i-1} \times D^{n-i+1}) \bigcup_{S^{i-1} \times S^{n-i}} D^i \times S^{n-1}. \tag{31}$$

Definition 9. An *elementary* $(n+1)$*-dimensional bordism of index* i is the bordism $(W; M, N)$ obtained from $M \times D^1$ by attaching a i-handle at $S^{i-1} \times D^{n-i+1} \hookrightarrow M \times \{1\}$. The *dual of an elementary* $(n+1)$*-dimensional bordism* $(W; M, N)$ of index i is the elementary $(n+1)$-dimensional bordism $(W; N, M)$ of index $(n-i+1)$, obtained by reversing the ends and regarding the i-handle attached to $M \times D^1$ as a $(n-i+1)$-handle attached to $N \times D^1$.

Theorem 13 (Handle decomposition of bordisms in the category $\mathfrak{M}_\infty$).

(1) *Every bordism* $(W; M, N)$*,* $\dim W = n+1$*,* $\dim M = \dim N = n$*, has a handle decomposition of the union of a finite sequence*

$$(W; M, N) = (W_1; M_0, M_1) \bigcup (W_2; M_1, M_2) \bigcup \cdots \bigcup (W_k; M_{k-1}, M_k) \tag{32}$$

of adjoining elementary bordisms $(W_s; M_{s-1}, M_s)$ *with index* (i_s) *such that* $0 \leq i_1 \leq i_2 \leq \cdots \leq i_k \leq n+1$.

(2) *Closed* n*-dimensional manifolds* M*,* N *are bordant iff* N *can be obtained from* M *by a sequence of surgeries.*

(3) *Every closed* n*-dimensional manifold* M *can be obtained from* $\varnothing$ *by attaching handles:*

$$M = h^{i_0} \bigcup h^{i_1} \bigcup \cdots \bigcup h^{i_k}. \tag{33}$$

Furthermore, M *has a Morse function* $f : M \to \mathbb{R}$ *with critical points* $\{x_{i_0}, x_{i_1}, \cdots, x_{i_k},\}$*, where* x_λ *is a critical point with index* λ*, and the corresponding vector field* $\zeta = \operatorname{grad} f : M \to TM$ *has zero-value only at such critical points. (See Fig. 1.)*

Proof. In fact any n-dimensional manifold can be characterized by means of its corresponding CW-substitute.

Example 8 (Sphere S^2). In this case one has the following handle decomposition: $S^2 = h^0 \bigcup h^2$, with $h^0 = D^0 \times D^2 = \{0\} \times D^2$, the *south hemisphere*, and $h^2 = D^2 \times D^0 = D^2 \times \{1\}$, the *north hemisphere*.

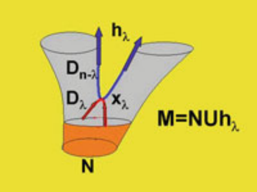


Fig. 1 Passing through critical point $x_\lambda \in M$, (index λ), of Morse function $f : M \to \mathbb{R}$, identified by attaching handle to a manifold N. (Separatrix diagram)

Example 9 (Torus $T^2 = S^1 \times S^1$). This 2-dimensional manifold has the following CW-complex structure: $T^2 = h^0 \bigcup h^1 \bigcup h^1 \bigcup h^2$, with $h^0 = \{0\} \times D^2$, $h^1 = D^1 \times D^1$, $h^2 = D^2 \times D^0 = D^2 \times \{1\}$.

Remark 1. One way to prove whether two manifolds are diffeomorphic is just to suitably use bordism and surgery techniques. (See, e.g., [48, 49].) In fact we should first prove that they are bordant and then see if some bordism can be modified by successive surgeries on the interior to become an s-bordism.

Theorem 14 (Homology properties in $\mathfrak{M}_\infty$.). *Let M be a n-dimensional manifold. One has the following homology structures.*

Let $\{C_\bullet(M;A) \cong A \otimes_{\mathbb{R}} C_\bullet(M;\mathbb{R}), \partial\}$ be the chain complex extension of the singular chain complex of M. Then one has the exact commutative diagram (34).

$$\begin{tikzcd}
 & & 0 \arrow[d] & 0 \arrow[d] & & \\
 & 0 \arrow[r] & B_\bullet(M;A) \arrow[r] \arrow[d] & Z_\bullet(M;A) \arrow[r] \arrow[d] & H_\bullet(M;A) \arrow[r] & 0 \\
 & & C_\bullet(M;A) \arrow[r, equal] \arrow[d] & C_\bullet(M;A) \arrow[d] & & \\
0 \arrow[r] & {}^{A}\underline{\Omega}_{\bullet,s}(M) \arrow[r] & Bor_\bullet(M;A) \arrow[r] \arrow[d] & Cyc_\bullet(M;A) \arrow[r] \arrow[d] & 0 & \\
 & & 0 & 0 & &
\end{tikzcd} \tag{34}$$

where:

$$\begin{cases}
B_\bullet(M;A) = \ker(\partial|_{\bar{C}_\bullet(M;A)});\ Z_\bullet(M;A) = \mathrm{im}\,(\partial|_{C_\bullet(M;A)});\\
H_\bullet(M;A) = Z_\bullet(M;A)/B_\bullet(M;A),\\
b \in [a] \in Bor_\bullet(M;A) \Rightarrow a - b = \partial c,\ c \in C_\bullet(M;A);\\
b \in [a] \in Cyc_\bullet(M;A) \Rightarrow \partial(a-b) = 0;\\
b \in [a] \in {}^{A}\underline{\Omega}_{\bullet,s}(M) \Rightarrow \left\{ \begin{array}{l} \partial a = \partial b = 0 \\ a - b = \partial c, \quad c \in C_\bullet(M;A) \end{array} \right\}.
\end{cases}$$

Furthermore, one has the following canonical isomorphism: ${}^{A}\underline{\Omega}_{\bullet,s}(M) \cong H_\bullet(M;A)$. As $C_\bullet(M;A)$ is a free two-sided projective A-module, one has the unnatural isomorphism: $Bor_\bullet(M;A) \cong {}^{A}\underline{\Omega}_{\bullet,s}(M) \bigoplus Cyc_\bullet(M;A)$.

Proof. It follows from standard results in homological algebra and homology in topological spaces. (For more details about see also [28].)

$$\begin{tikzcd}
 & & 0 \arrow[d] & 0 \arrow[d] & & \\
 & 0 \arrow[r] & B^\bullet(M;A) \arrow[r] \arrow[d] & Z^\bullet(M;A) \arrow[r] \arrow[d] & H^\bullet(M;A) \arrow[r] & 0 \\
 & & C^\bullet(M;A) \arrow[r, equal] \arrow[d] & C^\bullet(M;A) \arrow[d] & & \\
0 \arrow[r] & {}^A\underline{\Omega}^\bullet_s(M) \arrow[r] & Bor^\bullet(M;A) \arrow[r] \arrow[d] & Cyc^\bullet(M;A) \arrow[r] \arrow[d] & 0 & \\
 & & 0 & 0 & &
\end{tikzcd} \tag{35}$$

Theorem 15 (Cohomology properties in $\mathfrak{M}_\infty$). *Let M be a n-dimensional manifold. One has the following cohomology structures.*

Let $\{C^\bullet(M;A) \equiv Hom_A(C_\bullet(M;A);A) \cong Hom_{\mathbb{R}}(C_\bullet(M;\mathbb{R});A), \delta\}$ be the dual of the chain complex $C_\bullet(M;A)$ considered in above theorem. Then one has the exact commutative diagram (35).

where:

$$\begin{cases}
B^\bullet(M;A) = \ker(\delta|_{\bar{C}^\bullet(M;A)});\ Z^\bullet(M;A) = \operatorname{im}(\delta|_{C^\bullet(M;A)});\\
H^\bullet(M;A) = Z^\bullet(M;A)/B^\bullet(M;A);\\
b \in [a] \in Bor^\bullet(M;A) \Rightarrow a - b = \delta c;\ c \in C^\bullet(M;A);\\
b \in [a] \in Cyc^\bullet(M;A) \Rightarrow \delta(a-b) = 0;\\
b \in [a] \in {}^A\underline{\Omega}^\bullet_s(M) \Rightarrow \left\{\begin{array}{ll} \delta a = \delta b = 0 & \\ a - b = \delta c, & c \in C^\bullet(M;A) \end{array}\right\}.
\end{cases}$$

Furthermore, one has the following canonical isomorphism: ${}^A\underline{\Omega}^\bullet_s(M) \cong H^\bullet(M;A)$. As $C^\bullet(M;A)$ is a free two-sided projective A-module, one has the unnatural isomorphism: $Bor_\bullet(M;A) \cong {}^A\underline{\Omega}^\bullet_s(M) \bigoplus Cyc^\bullet(M;A)$.

Proof. It follows from standard results in cohomological algebra and cohomology in topological spaces. (See also [28].)

Definition 10. We say that a manifold M is *cohomologically trivial* if all the cohomology groups $H^r(M;A)$ vanish for $r \geq 1$.

Theorem 16. *Let M be a n-dimensional manifold. The following propositions are equivalent.*

(i) *M is cohomologically trivial.*
(ii) *$H^r(M;\mathbb{K}) = 0$, $\forall r \geq 1$.*
(iii) *$H_r(M;A) \cong H_r(M;\mathbb{K}) = 0$, $\forall r \geq 1$.*
(iv) *The complex $\{C^\bullet(M;A), \delta\}$ is acyclic, i.e., the sequence*

$$0 \longrightarrow Z^0 \xrightarrow{\epsilon} C^0(M;A) \xrightarrow{\delta} C^1(M;A) \xrightarrow{\delta} \cdots \xrightarrow{\delta} C^n(M;A) \xrightarrow{\delta} 0 \tag{36}$$

is exact.

Proof. (i)⇔(ii) since $H^r(M;A) \cong H^r(M;\mathbb{K}) \bigotimes_{\mathbb{K}} A$.
(i)⇔(iii) since $H^r(M;A) \cong Hom_A(H_r(M;A);A)$ and considering the following isomorphism $H_r(M;A) \cong H_r(M;\mathbb{K}) \bigotimes_{\mathbb{K}} A$.
(i)⇔(iv) since the exactness of the sequence (36) is equivalent to $H^r(M;A) = 0$, for $r \geq 1$.

Theorem 17. *Let M be a n-dimensional manifold modeled on the algebra A. The following propositions are equivalent.*

(i) *M is cohomologically trivial.*
(ii) *$H_r(M;A) = 0$, $r \geq 1$.*
(iii) *$H_r(M;\mathbb{K}) = H^r(M;\mathbb{K}) = 0$, $r \geq 1$.*

Example 10. A manifold contractible to a point is cohomologically trivial.

Theorem 18 (h-Cobordism groups).

(1) *If $(W;X_0,X_1)$, $\partial W = X_0 \sqcup X_1$, is a h-cobordant and X_i, $i = 0,1$, are simply connected, then W is a trivial h-cobordism. (For $n = 4$ the h-cobordism is true topologically.)*
(2) *A simply connected manifold M is h-cobordant to the sphere S^n iff M bounds a contractible manifold.*
(3) *If M is a homotopy sphere, then $M\sharp(-M)$ bounds a contractible manifold.*
(4) *If a homotopy sphere of dimension $2k$ bounds an stably-parallelizable manifold M then it bounds a contractible manifold M_1. (See Definition 12.)*
(5) *Let Θ_n denote the collection of all h-cobordism classes of homotopy n-spheres. Θ_n is an additive group with respect the connected sum,[15] where the sphere S^n serves as zero element. The opposite of an element X is the same manifold with reversed orientation, denoted by $-X$. In Table 8 are reported the expressions of some calculated groups Θ_n.[16]*

Proof. There are topological manifolds that have not smooth structure. Furthermore, there are examples of topological manifolds that have smooth structure

[15] The *connected sum* of two connected n-dimensional manifolds X and Y is the n-dimensional manifold $X\sharp Y$ obtained by excising the interior of embedded discs $D^n \subset X$, $D^n \subset Y$, and joining the boundary components $S^{n-1} \subset \overline{X \setminus D^n}$, $S^{n-1} \subset \overline{Y \setminus D^n}$, by $S^{n-1} \times I$.

[16] It is interesting to add that another related notion of cobordism is the *H-cobordism* of n-dimensional manifold, $(V;M,N)$, $\partial V = M \sqcup N$, with $H_\bullet(M) \cong H_\bullet(N) \cong H_\bullet(V)$. An n-dimensional manifold Σ is a *homology sphere* if $H_\bullet(\Sigma) = H_\bullet(S^n)$. Let Θ_n^H be the abelian group of H-cobordism classes of n-dimensional homology spheres, with addition by connected sum. (Kervaire's theorem.) For $n \geq 4$ every n-dimensional homology sphere Σ is H-cobordant to a homology sphere and the forgethful map $\Theta_n \to \Theta_n^H$ is an isomorphism.

everywhere except a single point. If a neighborhood of that point is removed, the smooth boundary is a homotopy sphere. Any smooth manifold may be *triangulated*, i.e., admits a PL structure, and the underlying PL manifold is unique up to a PL isomorphism.[17] The vice versa is false. No all topological or PL manifolds have at least one smooth structure.

The h-cobordism classes Θ_n of n-dimensional homotopy spheres are trivial for $1 \leq n \leq 6$, i.e., $\Theta_n \cong 0$, and $[S^n]=0$. For $n=1,2$ this follows from the fact that each of such topological manifolds have a unique smooth structure uniquely determined by its homology. For $n=3$ this follows from the proof of the Poincaré conjecture (see [40]). Furthermore, each topological 3-manifold has an unique differential structure [21, 23, 50]. Therefore, since from the proof of the Poincaré conjecture it follows that all 3-homotpy spheres are homeomorphic to S^3, it necessarily follows that all 3-homotpy spheres are diffeomorphic to S^3 too. Furthermore, for $n = 4$, the triviality of Θ_4 follows from the works by Freedman [7]. (See also Cerf [4].)

Lemma 8 (Freedman's theorem [7]). *Two closed simply connected* 4*-manifolds are homeomorphic iff they have the same symmetric bilinear form $\sigma : H^2(M;\mathbb{Z}) \otimes H^2(M;\mathbb{Z}) \to H^4(M;\mathbb{Z}) \cong \mathbb{Z}$, (with determinant* ± 1, *induced by the cup product), and the same* Kirby-Siebermann invariant κ.[18] *Any σ can be realized by such a manifold. If $\sigma(x \otimes x)$ is odd for some $x \in H^2(M;\mathbb{Z})$, then either value of κ can be realized also. However, if $\sigma(x \otimes x)$ is always even, then κ is determined by σ, being congruent to $\frac{1}{8}\sigma$.*

In particular, if M is homotopy sphere, then $H^2(M,\mathbb{Z}) = 0$ and $\kappa \equiv 0$, so M is homeomorphic to S^4.[19]

The cases $n = 5, 6$, can be proved by using surgery theory and depend by the Smale's h-cobordism theorem.[20]

Lemma 9 (Smale's h-cobordism theorem). *[45] Any n-dimensional simply connected h-cobordism W, $n > 5$, with $\partial W = M \sqcup (-N)$, is diffeomorphic to*

[17] A topological manifold M is piecewise linear, i.e., admits a PL structure, if there exists an atlas $\{U_\alpha, \varphi_\alpha\}$ such that the composities $\varphi_\alpha \circ \varphi_{\alpha'}^{-1}$, are piecewise linear. Then there is a polyehdron $P \subset \mathbb{R}^s$, for some s and a homeomorphism $\varphi : P \approx M$, (*triangulation*), such that each composite $\varphi_\alpha \circ \varphi$ is piecewise linear.

[18] κ is $\mathbb{Z}_2$-valued and vanishes iff the product manifold $M \times \mathbb{R}$ can be given a differentiable structure.

[19] It is not known which 4-manifolds with $\kappa = 0$ actually possess differentiable structure, and it is not known when this structure is essentially unique.

[20] There exists also a s-cobordism version of such a theorem for non-simply connected manifolds. More precisely, an $(n+1)$-dimensional h-cobordism $(V; N, M)$ with $n \geq 5$, is trivial iff it is an s-cobordism. This means that for $n \geq 5$ h-cobordant n-dimensional manifolds are diffeomorphic iff they are s-cobordant. Since the Whitehead group of the trivial group is trivial, i.e., $Wh(\{1\}) = 0$, it follows that h-cobordism theorem is the simply-connected special case of the s-cobordism.

$M \times [1,0]$. *(All manifolds are considered smooth and oriented.* $-N$ *denotes the manifold* N *with reversed orientation.)*[21]

If $n \geq 5$ *any two homotopy* n*-sphere are PL homeomorphic, and diffeomorphic too except perhaps at a single point. (If* $n = 5$*, (resp.* $n = 6$*), then any homotopy* n*-sphere* Σ *bounds a contractible* 6*-manifold, (resp.* 7*-manifold), and is diffeomorphic to* S^5*, (resp.* S^6*). Every smooth manifold* M *of dimension* $n > 4$*, having the homotopy of a sphere is a* twisted sphere*, i.e.,* M *can be obtained by taking two disks* D^n *and gluing them together by a diffeomorphism* $f : S^{n-1} \cong S^{n-1}$ *of their boundaries. More precisely one has the isomorphism* $\pi_0(Diff_+(S^n)) \cong \Theta_{n+1}$*,* $[f] \mapsto \Sigma_f \equiv D^{n+1} \bigcup_f (-D^{n+1})$*, where* $\pi_0(Diff_+(S^n))$ *denotes the group of isotopy classes of oriented preserving diffeomorphisms of* S^n*.*

See the paper by Kervaire and Milnor [18] and the following ones by Smale [45,46]. In the following we shall give a short summary of this proof for $n \geq 5$. This is really an application of the Browder-Novikov theorem.

Lemma 10. *Let* Ξ_n *denote the set of smooth* n*-dimensional manifolds homeomorphic to* S^n*. Let* $\sim_d$ *denote the equivalence relation in* Ξ_n *induced by diffeomorphic manifolds. Put* $\Gamma_n \equiv \Xi_n / \sim_d$*.*[22] *Then the operation of connected sum makes* Γ_n *an abelian group for* $n \geq 1$*.*

Proof. Let us first remark that since we are working in Ξ_n, the operation of connected sum there must be considered in smooth sense. Then it is easy to see that the for $M_1, M_2, M_3 \in \Xi_n$, one has $M_1 \sharp M_2 \cong M_2 \sharp M_1$ and that $(M_1 \sharp M_2) \sharp M_3 \cong M_1 \sharp (M_2 \sharp M_3)$. Therefore, it is well defined the commutative and associative composition map $+ : \Gamma_n \times \Gamma_n \to \Gamma_n$, $[M_1]+[M_2] = [M_1 \sharp M_2]$. The zero of this composition is the equivalence class $[S^n] \in \Gamma_n$. In fact, since $\overline{S^n \setminus D^n} \bigcup_{S^{n-1}} (S^{n-1} \times I) \cong D^n$, we get $M \sharp S^n \cong \overline{M \setminus D^n} \bigcup_{S^{n-1}} (\overline{S^n \setminus D^n} \bigcup_{S^{n-1}} (S^{n-1} \times I)) \cong \overline{M \setminus D^n} \bigcup_{S^{n-1}} D^n \cong M$. Therefore, $[S^n] = 0 \in \Gamma_n$. Furthermore, each element $M \in \Xi_n$ admits, up to diffeomorphisms, an unique opposite $M' \in \Xi_n$. In fact, since $\overline{M \setminus D^n} \cong D^n$, it follows that $M \cong D^n \bigcup_\lambda D^n \cong D^n \bigcup_{S^{n-1}} D^n$, where

[21] The proof utilizes Morse theory and the fact that for an h-cobordism $H_\bullet(W, M) \cong H_\bullet(W, N) \cong 0$, gives $W \cong M \times [0,1]$. The motivation to work with dimensions $n \geq 5$ is in the fact that it is used the *Whitney embedding theorem* that states that a map $f : N \to M$, between manifolds of dimension n and m respectively, such that either $2n + 1 \leq m$ or $m = 2n \geq 6$ and $\pi_1(M) = \{1\}$, is homotopic to an embedding.

[22] Γ_n can be identified with the set of *twisted* n*-spheres* up to orientation-preserving diffeomorphisms, for $n \neq 4$. One has the exact sequences given in (37).

$$\pi_0(Diff^+(D^n)) \longrightarrow \pi_0(Diff^+(S^{n-1})) \longrightarrow \Gamma_n \longrightarrow 0 \tag{37}$$

If the used diffeomorphism $S^{n-1} \to S^{n-1}$ to obtain a twisted n-sphere by gluing the corresponding boundaries of two disks D^n, is not smoothly isotopic to the identity, one obtains an exotic n-sphere. For $n > 4$ every exotic n-sphere is a twisted sphere. For $n = 4$, instead, twisted spheres are standard ones [4].

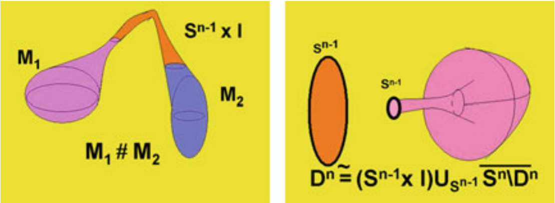


Fig. 2 Connected sum representations. $M_1\sharp M_2 = (M_1 \setminus D^n)\bigcup(S^{n-1}\times D^1)\bigcup(M_2 \setminus D^n)$

$\lambda : S^{n-1} \to S^{n-1}$ is a given diffeomorphism that identifies the two copies of S^{n-1}. Then M' is defined by $M' = D^n \bigcup_{\lambda^{-1}} D^n$. In fact one has $M\sharp M' \cong (D^n \bigcup_\lambda D^n)\sharp(D^n \bigcup_{\lambda^{-1}} D^n) \cong D^n \bigcup_1 D^n \cong S^n$, where $1 = \lambda\circ\lambda^{-1} : S^{n-1} \to S^{n-1}$. (See Fig. 2.)

Lemma 11. *One has the group isomorphisms $\Gamma_n \cong \Theta_n$ for any $n \geq 1$ and $n \neq 4$. So these groups classify all possible differentiable structures on S^n, up to orientation preserving diffeomorphisms.*

Remark 2 (Strange phenomena on dimension four). It is well known that on $\mathbb{R}^4$ there are uncountably many inequivalent differentiable structures, i.e., one has *exotic* $\mathbb{R}^4$, say $\widetilde{\mathbb{R}^4}$. (This is a result by Freedman [7], starting from some results by Donaldson [5].) On the other hand by the fact that $\Gamma_4 \cong \Theta_4 = 0$ it follows for any 4-dimensional homotopy sphere $\Sigma \cong S^4$. So taking in to account that $S^4\setminus\{pt\} \cong \mathbb{R}^4$, it natural arises the question: *Do exotic 4-sphere $\widetilde{\Sigma}$ exist such that $\widetilde{\Sigma} \setminus \{pt\} \cong \widetilde{\mathbb{R}^4}$?* The answer to this question, conjecturing the isomorphism $\Gamma_4 \cong \Theta_4 = 0$, should be in the negative. This means that all exotic $\widetilde{\mathbb{R}^4}$ collapse on the unique one S^4 by the process of one point compactification ! However this is generally considered an open problem in geometric topology and called the *smooth Poincaré conjecture*. (See, e.g., [8].)

Lemma 12 (Hirsch and Munkres [16, 23]). *The obstructions to the existence of a smooth structure on a n-dimensional combinatorial (or PL) manifold lie in the groups $H^{k+1}(M;\Gamma_k)$; while the obstruction to the uniqueness of such a smooth structures, when it exists, are elements of $H^k(M;\Gamma_k)$.*

Lemma 13 (Kirby and Siebenman [19]). *For a topological manifold M of dimension $n \geq 5$, there is only one obstruction to existence of a PL-structure, living in $H^4(M;\mathbb{Z}_2)$, and only one obstruction to the uniqueness of this structure (when it exists), living in $H^3(M;\mathbb{Z}_2)$.*[23]

[23] This result by Kirby and Siebenman does not exclude that every manifold of dimension $n > 4$ can possess some triangulation, even if it cannot be PL-homeomorphic to Euclidean space.

Definition 11 (Intersection form and signature of manifold).

(1) The *intersection form* of a $2n$-dimensional topological manifold with boundary $(M, \partial M)$ is the $(-1)^n$-symmetric form λ over the $\mathbb{Z}$-module $H \equiv H^n(M, \partial)/$ *torsion*, $\lambda(x, y) =< x \cup y, [M] > \in \mathbb{Z}$.[24]

(2) (Milnor's plumbing theorem.) For $n \geq 3$ every $(-1)^n$-quadratic form (H, λ, μ) over $\mathbb{Z}$ is realized by an $(n-1)$-connected $2n$-dimensional framed manifold with boundary $(V, \partial V)$ with $H_n(V) = H$. The form (H, λ, μ) is nonsingular iff $H_\bullet(\partial V) = H_\bullet(S^{2n-1})$. Let (H, λ) be a nonsingular (-1)-symmetric form over $\mathbb{Z}$ and $\{b_j, c_j\}_{1 \leq j \leq p}$ a basis for the $\mathbb{Z}$-module H, such that $\lambda(b_r, b_s) = 0$, $\lambda(c_r, c_s) = 0$, $\lambda(b_r, c_s) = 0$, for $r \neq s$, and $\lambda(b_r, c_r) = 1$. Let $\mu : H \to Q_{-1}(\mathbb{Z}) = \mathbb{Z}_2$ be a (-1)-quadratic function associated to (H, λ). Then the *Arf invariant* of a nonsingular (-1)-quadratic form (H, λ, μ) over $\mathbb{Z}$ is $Arf(H, \lambda, \mu) = \sum_{1 \leq j \leq p} \mu(b_j)\mu(c_j) \in \mathbb{Z}_2 \equiv \{0, 1\}$.
If $\partial M = \varnothing$, or $H_\bullet(\partial M) = H_\bullet(S^{2n-1})$, then λ is nonsingular.

(3) The *signature* $\sigma(M)$ of a $4k$-dimensional manifold $(M, \partial M)$ is $\sigma(M) = \sigma(\lambda) \in \mathbb{Z}$, where λ is the symmetric form over the $\mathbb{Z}$-module $H_{2k}(M)/torsion$.

(4) Let M be an oriented manifold with empty boundary, $\partial M = \varnothing$, $\dim M = n = 4k$. Then $H^i(M; \mathbb{Z})$ is finitely generated for each i and $H^{2k}(M; \mathbb{Z}) \cong \mathbb{Z}^s \bigoplus Tor$, where Tor is the torsion subgroup. Let $[M] \in H_{4k}(M; \mathbb{Z})$ be the orientation class of M. Let $<,>$: $H^{2k}(M; \mathbb{Z}) \times H^{2k}(M; \mathbb{Z}) \to \mathbb{Z}$, be the symmetric bilinear form given by $(a, b) \mapsto < a, b > = < a \cup b, [M] > \in \mathbb{Z}$. This form vanishes on the torsion subgroup, hence it factors on $H^{2k}(M; \mathbb{Z})/Tor \times H^{2k}(M; \mathbb{Z})/Tor \cong \mathbb{Z}^s \times \mathbb{Z}^s \to \mathbb{Z}$. This means that the adjoint map $\varphi : H^{2k}(M; \mathbb{Z})/Tor \to Hom_{\mathbb{Z}}(H^{2k}(M; \mathbb{Z})/Tor; \mathbb{Z}) = (H^{2k}(M; \mathbb{Z})/Tor))^*$, $a \mapsto \varphi(a)(b) =< a, b >$, is an isomorphism. This is a just a direct consequence of Poincaré duality: $H_{2k}(M; \mathbb{Z})/Tor = (H^{2k}(M; \mathbb{Z})/Tor)^*$ and $a(b \cap [M]) =< a \cup b, [M] >$. Then the signature of this bilinear form is the usual signature

[24] Let R be a commutative ring and H a finite generated free R-module. A *ϵ-symmetric form over* H is a bilinear mapping $\lambda : H \times H \to R$, such that $\lambda(x, y) = \epsilon\lambda(y, x)$, with $\epsilon \in \{+1, -1\}$. The form λ is *nonsingular* if the R-module morphism $H \to H^* \equiv Hom_R(H; R)$, $x \mapsto (y \mapsto \lambda(x, y))$ is an isomorphism. A *ϵ-quadratic form* associated to a ϵ-symmetric form λ over H, is a function $\mu : H \to Q_\epsilon(R) \equiv \operatorname{coker}(1 - \epsilon : R \to R)$, such that: (i) $\lambda(x, y) = \mu(x + y) - \mu(x) - \mu(y)$; (ii) $\lambda(x, x) = (1 + \epsilon)\mu(x) \in \operatorname{im}(1 + \epsilon : R \to R) \subseteq \ker(1 - \epsilon : R \to R)$, $\forall x, y \in H$, $a \in R$. If $R = \mathbb{Z}$ and $\epsilon = 1$, we say *signature of* λ, $\sigma(\lambda) = p - q \in \mathbb{Z}$, where p and q are respectively the number of positive and negative eigenvalues of the extended form on $\mathbb{R} \bigotimes_{\mathbb{Z}} H$. Then λ has a 1-quadratic function $\mu : H \to Q_{+1}(\mathbb{Z})$ iff λ has even diagonal entries, i.e., $\lambda(x, x) \equiv 0 \pmod 2$, with $\mu(x) = \lambda(x, x)/2$, $\forall x \in H$. If λ is nonsingular then $\sigma(\lambda) \equiv 0 \pmod 8$. Examples. (1) $R = H = \mathbb{Z}$, $\lambda = 1$, $\sigma(\lambda) = 1$.

(2) $R = \mathbb{Z}$, $H = \mathbb{Z}^8$, $\lambda = E^8$-form, given by $\left(\lambda_{ij}\right) = \begin{pmatrix} a_{rs} & b_{rs} \\ c_{rs} & d_{rs} \end{pmatrix}$ with $(a_{rs}) = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{pmatrix}$,

$(b_{rs}) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$, $(c_{rs}) = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$, $(d_{rs}) = \begin{pmatrix} 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \end{pmatrix}$.

of the form after tensoring with the rationals $\mathbb{Q}$, i.e., of the symmetrix matrix associated to the form, after choosing a basis for $\mathbb{Q}^s$. Hence the *signature* is the difference between the number of $+1$ eigenvalues with the number of -1 eigenvalues of such a matrix. Let us denote by $\sigma(M)$ the signature of the above nondegenerate symmetric bilinear form, and call it *signature* of M.

Theorem 19 (R. Thom's properties of signature).

(1) *If M has* $\dim M = 4k$*, and it is a boundary, then* $\sigma(M) = 0$.

(2) $\sigma(-M) = -\sigma(M)$.

(3) *Let M and L be two $4k$-dimensional closed, compact, oriented manifolds without boundary. Then we have $\sigma(M \sqcup L) = \sigma(M) + \sigma(L)$ and $\sigma(M \times L) = \sigma(M).\sigma(L)$, where the orientation on $M \times L$ is $[M \times L] = [M] \otimes [L]$.*

(4) (Rohlin's signature theorem). *The signature of a closed oriented $4k$-dimensional manifold is an oriented cobordism invariant, i.e., if $\partial W = M \sqcup N$, it follows that $\sigma(M) = \sigma(N) \in \mathbb{Z}$. More precisely, the signature for oriented boundary $4k$-dimensional manifolds is zero and it defines a linear form $\sigma : {}^+\Omega_{4k} \to \mathbb{Z}$. Furthermore, let M and N be $4k$-dimensional manifolds with differentiable boundaries: $\partial M = \bigcup_j X_j$, $\partial N = \bigcup_i Y_i$, such that $X_1 = Y_1$. Then one has the formula (38).*

$$\sigma(M \bigcup_{X_1 = Y_1} N) = \sigma(M) + \sigma(N). \tag{38}$$

(5) (Hirzebruch's signature theorem (1952)). *The signature of a closed oriented $4k$-dimensional manifold M is given by*

$$\sigma(M) =< \mathcal{L}_k(p_1, \cdots, p_k), [M] >\in \mathbb{Z} \tag{39}$$

with $\mathcal{L}_k(p_1, \cdots, p_k)$ polynomial in the Pontrjagin classes p_j of M, i.e., $p_j(M) \equiv p_j(TM) \in H^{4j}(M)$, representing the $\mathcal{L}$ genus, i.e., the genus of the formal power series given in (40).[25]

[25] A *genus* for closed smooth manifolds with some X-structure, is a ring homomorphism $\Omega^X_\bullet \to R$, where R is a ring. For example, if the X-structure is that of oriented manifolds, i.e., $X = SO$, then the signature of these manifolds just identifies a genus $\sigma : {}^+\Omega_\bullet = \Omega^{SO}_\bullet \to \mathbb{Z}$, such that $\sigma(1) = 1$, and $\sigma : {}^+\Omega_p \to 0$ if $p \neq 4q$. Therefore the genus identifies also a $\mathbb{Q}$-algebra homomorphism $\Omega^{SO}_\bullet \bigotimes_{\mathbb{Q}} \mathbb{Q} \to \mathbb{Q}$. More precisely, let $A \equiv \mathbb{Q}[t_1, t_2, \cdots]$ be a graded commutative algebra, where t_i has degree i. Set $\mathcal{A} \equiv A[[a_0, a_1, \cdots]]$, where $a_i \in A$ is homogeneous of degree i, i.e., the elements of $\mathcal{A}$ are infinite formal sums $a \equiv a_0 + a_1 + a_2 + \cdots$. Let $\mathcal{A}^\bullet \subset \mathcal{A}$ denote the subgroup of the multiplicative group of $\mathcal{A}$ of elements with leading term 1. Let $K_1(t_1)$, $K_2(t_1, t_2)$, $K_3(t_1, t_2, t_3), \cdots \in A$, be a sequence of polynomials of A, where K_n is homogeneous of degree n. For $a = a_0 + a_1 + a_2 + \cdots \in \mathcal{A}^\bullet$, we define $K(a) \in \mathcal{A}^\bullet$ by $K(a) = 1 + K_1(a_1) + K_2(a_1, a_2) + \cdots$. We say that K_n form a *multiplicative sequence* if $K(ab) = K(a)K(b)$, $\forall a, b \in \mathcal{A}^\bullet$. An example is with $K_n(t_1, \cdots, t_n) = \lambda^n t_n$, $\lambda \in \mathbb{Q}$. Another example is given by the formal power series given in (40) with $\lambda_k = (-1)^{k-1} \frac{2^{2k} B_{2k}}{(2k)!}$.

Table 4 Polynomials $s_I(\sigma_1, \cdots, \sigma_n)$, for $0 \leq n \leq 4$

n	$s_I(\sigma_1, \cdots, \sigma_n)$
0	$s = 1$
1	$s_{(1)}(\sigma_1) = \sigma_1$
2	$s_{(2)}(\sigma_1, \sigma_2) = \sigma_1^2 - 2\sigma_2$
	$s_{(1,1)}(\sigma_1, \sigma_2) = \sigma_2$
3	$s_{(3)}(\sigma_1, \sigma_2, \sigma_3) = \sigma_1^3 - 3\sigma_1\sigma_2 + 3\sigma_3$
	$s_{(1,2)}(\sigma_1, \sigma_2, \sigma_3) = \sigma_1\sigma_2 - 3\sigma_3$
	$s_{(1,1,1)}(\sigma_1, \sigma_2, \sigma_3) = \sigma_3$
4	$s_{(4)}(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = \sigma_1^4 - 4\sigma_1^2\sigma_2 + 2\sigma_2^2 + 4\sigma_1\sigma_3 - 4\sigma_4$
	$s_{(1,3)}(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = \sigma_1^2\sigma_2 - 2\sigma_2^2 - \sigma_1\sigma_3 + 4\sigma_4$
	$s_{(2,2)}(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = \sigma_2^2 - 2\sigma_1\sigma_3 + 2\sigma_4$
	$s_{(1,1,2)}(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = \sigma_1\sigma_3 - 4\sigma_4$
	$s_{(1,1,1,1)}(\sigma_1, \sigma_2, \sigma_3, \sigma_4) = \sigma_4$

$$\frac{\sqrt{z}}{\tanh(\sqrt{z})} = \sum_{k \geq 0} \frac{2^{2k} B_{2k} z^k}{(2k)!} = 1 + \frac{z}{3} - \frac{z^2}{45} + \cdots \tag{40}$$

where the numbers B_{2k} are the Bernoulli numbers.[26]

(6) *The intersection form of a $4k$-dimensional manifold M has a 1-quadratic function iff it has $2k^{th}$ Wu class $v_{2k}(M) = 0 \in H^{2k}(M; \mathbb{Z}_2)$, in which case* $\sigma(M) \equiv 0 \pmod 8$.

Definition 12. An n-dimensional manifold M is *parallelizable* if its tangent n-plane bundle $\tau_M : M \to BO(n)$ is trivial, i.e., isomorphic to $M \times \mathbb{R}^n \to M \equiv \epsilon^n$.

For any partition $I = (i_1, i_2, \cdots, i_k)$ of n, set $\lambda_I = \lambda_{i_1}\lambda_{i_2} \cdots \lambda_{i_k}$. Now define polynomials $\mathcal{L}_n(t_1, \cdots, t_n) \in A$ by $\mathcal{L}_n(t_1, \cdots, t_n) = \sum_I \lambda_I s_I(t_1, \cdots, t_n)$, where the sum is over all partitions of n and s_I is the unique polynomial belonging to $\mathbb{Z}[t_1, \cdots, t_n]$ such that $s_I(\sigma_1, \cdots, \sigma_n) = \sum t^I$, where $\sigma_1, \cdots, \sigma_n$ are the elementary symmetric functions that form a polynomial basis for the ring $\mathcal{S}_n$ of symmetric functions in n variables. ($\mathcal{S}_n$ is the graded subring of $\mathbb{Z}[t_1, \cdots, t_n]$ of polynomials that are fixed by every permutation of the variables. Therefore we can write $\mathcal{S}_n = \mathbb{Z}[\sigma_1, \cdots, \sigma_n]$, with σ_i of degree i. In Table 4 are reported the polynomials $s_I(\sigma_1, \cdots, \sigma_n)$, for $0 \leq n \leq 4$.) $\mathcal{L}_n$ form a multiplicative sequence. In fact $\mathcal{L}(ab) = \sum_I s_I(ab) = \sum_I \lambda_I \sum_{I_1 I_2 = I} s_{I_1}(a) s_{I_2}(b) = \sum_{I_1 I_2 = I} \lambda_{I_1} s_{I_1}(a) \lambda_{I_2} s_{I_2}(b) = \mathcal{L}(a)\mathcal{L}(b)$. Then for an n-dimensional manifold M one defines $\mathcal{L}$-*genus*, $\mathcal{L}[M] = 0$ if $n \neq 4k$, and $\mathcal{L}[M] =< K_k(p_1(TM), \cdots, p_k(TM)), \mu_M >$ if $n = 4k$, where μ_M is the rational fundamental class of M and $K_k(p_1(TM), \cdots, p_k(TM)) \in H^n(M; \mathbb{Z})$.

[26] Formula (39) is a direct consequence of Thom's computation of ${}^+\Omega_\bullet \bigotimes_{\mathbb{Z}} \mathbb{Q} \cong \mathbb{Q}[y_{4k} | k \geq 1]$, with $y_{4k} = [\mathbb{C}P^{2k}]$. (${}^+\Omega_j \bigotimes_{\mathbb{Z}} \mathbb{Q} = 0$ for $j \neq 4k$.) In Fact, one has $p_j(\mathbb{C}P^n) = p_j(T\mathbb{C}P^n) = \binom{n+1}{j} \in H^{4j}(\mathbb{C}P^n) = \mathbb{Z}$, $0 \leq j \leq \frac{n}{2}$. For $n = 2k$ the evaluation $< \mathcal{L}_k, [\mathbb{C}P^{2k}] >= 1 \in \mathbb{Z}$ coincides with the signature of $\mathbb{C}P^{2k}$: $\sigma(\mathbb{C}P^{2k}) = \sigma(H^{2k}(\mathbb{C}P^{2k}), \lambda) = \sigma(\mathbb{Z}, 1) = 1 \in \mathbb{Z}$. Therefore, the signature identifies a $\mathbb{Q}$-algebra homomorphism $\Omega_\bullet^{SO} \bigotimes_{\mathbb{Z}} \mathbb{Q} \to \mathbb{Q}$. So the Hirzebruch signature theorem states that this last homomorhism induced by the signature, coincides with the one induced by the genus. In Table 5 are reported some Hirzebruch's polynomials for $\mathbb{C}P^{2k}$. In Table 6 are reported also the Bernoulli numbers B_n, with the Kronecker's formula, and explicitly calculated for $0 \leq n \leq 18$.

Table 5 $\mathcal{L}_k$-polynomials for $\mathbb{C}P^{2k}$, with $k = 1, 2, 3, 4$ and related objects

k	$\mathcal{L}_k$	$p_j(\mathbb{C}P^{2k})$	$\sigma(\mathbb{C}P^{2k})$
1	$\mathcal{L}_1(p_1) = \frac{1}{3}$	$p_1(\mathbb{C}P^2) = 3$	$\sigma(\mathbb{C}P^2) = \frac{1}{3} \times 3 = 1$
2	$\mathcal{L}_2(p_1, p_2) = \frac{1}{45}(7p_2 - p_1^2)$	$p_1(\mathbb{C}P^4) = 5$	$\sigma(\mathbb{C}P^4) = \frac{1}{45}(7 \times 10 - 25) = 1$
		$p_2(\mathbb{C}P^4) = 10$	
3	$\mathcal{L}_3(p_1, p_2, p_3)$	$p_1(\mathbb{C}P^6) = 7$	$\sigma(\mathbb{C}P^6) = \frac{1}{945}(62 \times 35 - 13 \times 21$
	$= \frac{1}{945}(62p_3 - 13p_2p_1 + 2p_1^3)$	$p_2(\mathbb{C}P^6) = 21$	$\times 7 + 2 \times 343) = 1$
		$p_3(\mathbb{C}P^6) = 35$	
4	$\mathcal{L}_3(p_1, p_2, p_3, p_4) =$	$p_1(\mathbb{C}P^8) = 9$	$\sigma(\mathbb{C}P^8) = \frac{1}{14175}(48006 - 53676$
	$\frac{1}{14175}(381p_4 - 71p_3p_1 - 19p_2^2$	$p_2(\mathbb{C}P^8) = 36$	$-24624 + 64152 - 19683) = 1$
	$+22p_2p_1^2 - 3p_1^4)$	$p_3(\mathbb{C}P^8) = 84$	
		$p_4(\mathbb{C}P^8) = 126$	

Table 6 Bernoulli numbers $B_n = -\sum_{1 \leq k \leq n+1} \frac{(-1)^k}{k}\binom{n+1}{k} = \sum_{1 \leq j \leq k} j^n \in \mathbb{Q}, n \geq 0$

n	Numerator	Denominator
0	1	1
1	−1	2
2	1	6
4	−1	30
6	1	42
8	−1	30
10	5	66
12	−691	2,730
14	7	6
16	−3,617	510
18	43,862	798

$B_n = 0, n = \text{odd} > 1$

Two vector bundles ξ_1 and ξ_2 over a same base M are *stably isomorphic* if $\xi_1 \bigoplus \epsilon_1 \cong \xi_2 \bigoplus \epsilon_1$, where ϵ_i, $i = 1, 2$ are vector bundles over M with dimensions such that if M is a complex of dimension r, the total fiber dimensions of the Whitney sums exceeds r. Such bundles are said to be in *stable range*.

Proposition 1. *A connected compact n-manifold M with non-trivial boundary, is parallelizable iff it is stably parallelizable.*

Proof. In fact M has the homotopy type of an $(n-1)$-complex and thus TM is in the stable range.

Proposition 2. *The set of framed n-manifolds properly contains the set of parallelizable n-manifolds.*

Table 7 Parallelizable S^n

n	1	2	3	4	5	6	7	>7
parallelizable	Yes	No	Yes	No	No	No	Yes	No

Example 11 (Bott-Milnor [3]).

(1) The sphere S^n is framed with $S^n \times \mathbb{R} \subset \mathbb{R}^{n+1}$, $TS^n \bigoplus \epsilon \cong \epsilon^{n+1}$, but not necessarily parallelizable. (See Table 7.)
(2) All spheres are stably parallelizable.
(3) Every homotopy sphere Σ^n is stably parallelizable.

Definition 13 (Pontrjagin-Thom construction framed-cobordism). If (M, φ) is any n-manifold with *framing* $\varphi : \nu(M) \cong M \times \mathbb{R}^{k-n}$ of the normal bundle in $\mathbb{R}^k$, the same definition yields a map $p(M, \varphi) \in \pi^s_n$. If $(M_1, \varphi_1) \sqcup (M_2, \varphi_2) \subset \mathbb{R}^k$ form the framed boundary of a $(n+1)$-manifold $(W, \partial W, \Phi) \subset (\mathbb{R}^k \times [0, \infty), \mathbb{R}^k \times \{0\})$, we say that are *framed cobordant*. The framed cobordism is an equivalence relation and the corresponding set of framed cobordism classes is denoted by Ω^{fr}_n. This is an abelian additive group with respect the operation of disjoint union $\sqcup$. The class $[\varnothing]$ is the zero of Ω^{fr}_n. Then one has the canonical isomorphism given in (41).[27]

$$\left\{ \begin{array}{l} \Omega^{fr}_n \cong \pi^s_n \\ M \mapsto f_M : S^{n+k} \to S^{n+k}/(S^{n+k} \setminus (M \times \mathbb{R}^k)) = \Sigma^k M^+ \to S^k \end{array} \right. \tag{41}$$

where $M^+ \equiv M \bigcup \{pt\}$, and Σ^k is the k-suspension functor for spectra.

Example 12 (Smale's paradox and framed cobordism). Let us consider the so-called *Smale's paradox* turning a sphere $S^2 \subset \mathbb{R}^3$ inside out. (See also [42].) Let us denote by ${}^-S^2$ the sphere S^2 with reversed orientation. Let us first note that these surfaces are characterized by the same generalized curvature integra. It is useful to recall here some definitions and results about this topological invariant. The *generalized Gauss map* of an n-dimensional framed manifold M with $f : M \hookrightarrow \mathbb{R}^{n+k}$, $\nu_f \cong \epsilon^k$, is the map $c : M \to V_{n+k,k}$, $x \mapsto ((\nu_f)_x = \mathbb{R}^k \hookrightarrow T_{f(x)}\mathbb{R}^{n+k} = \mathbb{R}^{n+k})$ and classifies the tangent n-planes bundle $\tau_M : M \to BO(n)$, with the k-stable trivialization $TM \bigoplus \epsilon^k \cong TM \bigoplus \nu_f = \epsilon^{n+k} = T\mathbb{R}^{n+k}|_M$.[28] The *generalized curvatura integra* of M is the degree of the generalized Gauss map.

[27] In Table 8 are reported the n-stems for $0 \leq n \leq 17$.

[28] $V_{n+k,k} \cong O(n+k)/O(n)$ is the *Stiefel space* of orthonormal k-frames in $\mathbb{R}^{n+k}$, or equivalently of isometries $\mathbb{R}^k \to \mathbb{R}^{n+k}$. $V_{n+k,k}$ is $(n-1)$-connected with $H_n(V_{n+k,k}) = \mathbb{Z}$, if $n \equiv 0 (\mathrm{mod}\ 2)$ or if $k = 1$, and $H_n(V_{n+k,k}) = \mathbb{Z}_2$, if $n \equiv 1\ (\mathrm{mod}\ 2)$ and $k > 1$. One has $G_{n+k,k} = V_{n+k,k}/O(k)$, where $G_{n+k,k}$ is the Grassmann space of k-dimensional subspaces of $\mathbb{R}^{n+k}$. Then the classifying space for n-planes is $BO(n) = \varinjlim_{k} G_{n+k,k}$, and the corresponding stable classifying space is $BO = \varinjlim_{n} BO(n)$.

$$c_*[M] \in H_n(V_{n+k,k}) = \left\{ \begin{array}{l} \mathbb{Z},\ \text{if}\ n \equiv 0\ (\text{mod}\ 2) \\ \mathbb{Z}_2,\ \text{if}\ n \equiv 1\ (\text{mod}\ 2) \end{array} \right\}\ (k > 1). \tag{42}$$

The curvatura integra of an n-dimensional framed manifold M can be expressed with the *Kervaire's formula* given in (43).

$$c_*[M] = Hopf(M) + \left\{ \begin{array}{l} \chi(M)/2 \in \mathbb{Z},\ \text{if}\ n \equiv 0\ (\text{mod}\ 2) \\ \chi_{1/2}(M) \in \mathbb{Z}_2,\ \text{if}\ n \equiv 1\ (\text{mod}\ 2) \end{array} \right. \tag{43}$$

where

$$\chi_{1/2}(M) = \sum_{0 \le j \le (n-1)/2} \dim_{\mathbb{Z}_2} H_j(M;\mathbb{Z}) \in \mathbb{Z}_2 \tag{44}$$

is called the *Kervaire semicharacteristic*, and

$$Hopf[M] = \left\{ \begin{array}{l} 0 \in \mathbb{Z},\ \text{if}\ n \equiv 0\ (\text{mod}\ 2) \\ H_2(F) \in \mathbb{Z}_2,\ \text{if}\ n \equiv 1\ (\text{mod}\ 2) \end{array} \right. \tag{45}$$

with $F : S^{n+k} \to S^k$ the Pontrjagin-Thom map, and $H_2(F)$ determined by the *mod 2 Hopf invariant*. This is the morphism

$$\left\{ \begin{array}{l} H_2 : \pi_{n+k}(S^k) \to \mathbb{Z} \\ (F : S^{n+k} \to S^k) \mapsto H_2(F),\ (m \ge 1) \end{array} \right. \tag{46}$$

determined by the Steenrod square in the mapping cone $X = S^k \bigcup_F D^{n+k+1}$. If $a = 1 \in H^k(X;\mathbb{Z}_2) = \mathbb{Z}_2, b = 1 \in H^{n+k+1}(X;\mathbb{Z}_2) = \mathbb{Z}_2$, then $S_q^{n+1}(a) = H_2(F)b \in H^{n+k+1}(X;\mathbb{Z}_2)$. One has (Adams) that $H_2 = 0$ for $n\ \ /= 1,3,7$. $Hopf(M)$ is a framed cobordism invariant.

Taking into account that $\chi(S^2) = \chi(^-S^2) = 2$, and that $Hopf(S^2) = Hopf(^-S^2) = 0$, we get $c_*[S^2] = c_*[^-S^2] = 1$. Furthermore one has $\Omega_2^{fr} \cong \pi_2^s =, \mathbb{Z}_2 \cong \Omega_2$. In order to see that S^2 is cobordant with $^-S^2$ it is enough to prove that $^-S^2$ can be obtained by S^2 by a sequence of surgeries. (See Theorem 13.) In fact, we can write the oriented S^2 as $S^2 = D_W^2 \bigcup_{S^1} D_E^2$, where D_W^2 and D_E^2 are two oriented discs in such a way that S^2 is oriented with outgoing normal unitary vector field. (Fig. 3a.) By a surgery we can remove D_E^2 and smoothly add another D_E^2 on the left of D_W^2. (Fig. 3b.) Next by an orientation preserving diffeomorphisms, we get $^-S^2$, the surface represented in Fig. 3c.

In conclusion $S^2 \sqcup {}^-S^2 = \partial W$, where $W \cong (S^2 \times \{0\}) \times I \cong (^-S^2 \times \{1\}) \times I \subset \mathbb{R}^3 \times [0,\infty)$. Therefore, S^2 is framed cobordant with $^-S^2$. Furthermore $S^2 \cong {}^-S^2$ (diffeomorphism reversing orientation), that agrees with the well known result in differential topology that two connected, compact, orientable surfaces are diffeomorphic iff they have the same genus, the same Euler characteristics and the same number of boundaries. (See, e.g., [17].)

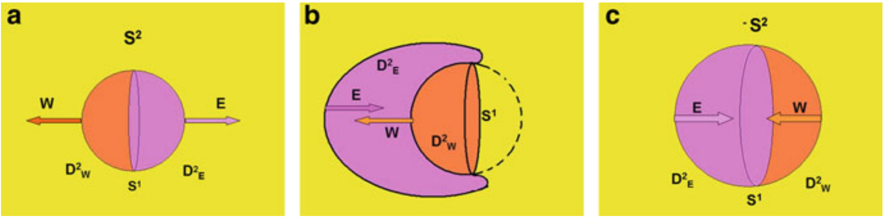


Fig. 3 Surgery and Smale's paradox turning a sphere $S^2\subset\mathbb{R}^3$ inside out

Another proof that S^2 and $-S^2$ are cobordant can be obtained by the Arf invariant. Let us recall that if M is a framed surface $M\times\mathbb{R}^k\subset\mathbb{R}^{k+2}$, the intersection form $(H^1(M),\lambda)$ has a canonical (-1)-quadratic function $\mu{:}H^1(M)=H_1(M)\to Q_{-1}(\mathbb{Z})=\Omega_1^{fr}=\mathbb{Z}_2$, given by $x\mapsto(x:S^1\hookrightarrow M)$, sending each $x\in H^1(M)$ to an embedding $x:S^1\hookrightarrow M$ with a corresponding framing $S^1\times\mathbb{R}^{k+1}\subset\mathbb{R}^{k+2}$, $\delta\nu_x:\nu_x\bigoplus\epsilon^k\cong\epsilon^{k+1}$. Then one has the isomorphism $Arf:\Omega_2^{fr}=\pi_2^s\cong\mathbb{Z}_2$, $[M]\mapsto Arf(H^1(M),\lambda,\mu)$. In the particular case that $M=S^2\sqcup -S^2$, we get $H^1(M)=0$ and $Arf(H^1(M),\lambda,\mu)=0$, hence must necessarily be $[M=S^2\sqcup -S^2]=0\in\Omega_2^{fr}$. This agrees with the fact that $\Omega_2^{fr}=\mathbb{Z}_2=\Omega_2$, and that both surfaces S^2 and $-S^2$ belong to $0\in\Omega_2$ since are orientable ones.

Definition 14. An n-dimensional manifold V with boundary ∂V, is *almost framed* if the open manifold $V\setminus\{pt\}$ framed: $(V\setminus\{pt\})\times\mathbb{R}^k\subset\mathbb{R}^{n+k}$ (for k large enough).

Theorem 20 (Properties of almost framed manifold).

(1) *An almost framed manifold V, with $\partial V\neq\varnothing$ is a framed manifold and a parallelizable manifold.*

(2) *If V is an almost framed manifold with $\partial V=\varnothing$, then there is a* framing obstruction

$$\mathfrak{o}(V)\in\ker(J:\pi_{n-1}(O)\to\pi_{n-1}^s)\qquad(47)$$

in the sense that V is framed iff $\mathfrak{o}(V)=0$.

(3) (Kervaire invariant for almost framed manifolds). Let $(M,\partial M)$ be a $(4k+2)$-dimensional almost framed manifold with boundary such that either $\partial M=\varnothing$ or $H_\bullet(M)=H_\bullet(S^{4k+1})$, so that $(H^{2k+1}(M;\mathbb{Z}_2),\lambda\,\mu)$ is a nonsingular quadratic form over $\mathbb{Z}_2$. The *Kervaire* of M is defined in (48).

$$Kervaire(M)=Arf(H^{2k+1}(M;\mathbb{Z}_2),\lambda\,\mu).\qquad(48)$$

One has the following propositions.

(i) If $\partial M=\varnothing$ and $M=\partial N$ is the boundary of a $(4k+3)$-dimensional almost framed manifold N, then $Kervaire(M)=0\in\mathbb{Z}_2$.

(ii) The Kervaire of a manifold identifies a framed cobordism invariant, i.e., it defines a map $Kervaire:\Omega_{4k+2}^{fr}=\pi_{4k+2}^s\to\mathbb{Z}_2$ that is 0 if $k\neq 2^i-1$.

(iii) There exist $(4k+2)$-dimensional framed manifolds M with *Kervaire* $(M) = 1$, for $k = 0, 1, 3, 7$. For $k = 0, 1, 3$ can take $M = S^{2k+1} \times S^{2k+1}$.

(4) *(Kervaire-Milnor's theorem on almost framed manifolds). Let us denote by Ω_n^{afr} the cobordism group of closed n-dimensional almost framed manifolds. One has the exact sequence in (49).*

$$\begin{tikzcd} \Omega_n^{afr} \arrow[r, "\mathfrak{o}"] & \pi_{n-1}(O) \arrow[r, "J"] & \pi_{n-1}^s \end{tikzcd} \tag{49}$$

- *For a $4k$-dimensional almost framed manifold V one has the framing obstruction reported in (50).*[29]

$$\left\{\begin{array}{l} \mathfrak{o}(V) = p_k(V)/(a_k(2k-1)!) \\ \qquad \ker(J : \pi_{4k-1}(O) \to \pi_{4k-1}^s) \\ \qquad = j_k\mathbb{Z} \subset \pi_{4k-1}(O) = \mathbb{Z} \end{array}\right\}$$

$$p_k(V) \in H^{4k}(V) = \mathbb{Z} \text{ (Pontryagin class)} \tag{50}$$

$$a_k = \left\{\begin{array}{l} 1 \, \textit{for } k \equiv 0 \pmod 2 \\ 2 \, \textit{for } k \equiv 1 \pmod 2 \end{array}\right\} \; j_k = \text{den}\left(\frac{B_k}{4k}\right)$$

(5) (Kervaire-Milnor's theorem on almost framed manifolds-2). Let P_n be the cobordism group of n-dimensional framed manifolds with homotopy sphere boundary. (P_n is called the *n-dimensional simply-connected surgery obstruction group*.) For $n \geq 4$ Θ_n is finite, with the short exact sequence given in (51).

$$\begin{tikzcd} 0 \arrow[r] & \text{coker}\,(a : \Omega_{n+1}^{afr} \to P_{n+1}) \arrow[r, "b"] & \Theta_n \arrow[r, "c"] & \ker(a : \Omega_n^{afr} \to P_n) \arrow[r] & 0 \end{tikzcd} \tag{51}$$

and

$$\ker(a) \subseteq \text{coker}\,(J : \pi_n \to \pi_n^s) = \ker(\mathfrak{o} : \Omega_n^{afr} \to \pi\pi_{n-1}(O)).$$

- In (52) are reported the calculated groups P_n.

$$\left\{\begin{array}{l} P_{2n+1} = 0 \\ P_n = \left\{\begin{array}{l} \mathbb{Z} \text{ if } n \equiv 0 \pmod 4 \\ 0 \text{ if } n \equiv 1 \pmod 4 \\ \mathbb{Z}_2 \text{ if } n \equiv 2 \pmod 4 \\ 0 \text{ if } n \equiv 3 \pmod 4 \end{array}\right\} \end{array}\right. \tag{52}$$

[29] For $k = 1$ one has $j_1 = 24$; $\mathfrak{o}(V) = p_1(V)/2 \in 24\mathbb{Z} \subset \pi_3(O) = \mathbb{Z}$.

(6) *(Kervaire-Milnor's braid $n \geq 5$). For $n \geq 5$ there is the exact commutative braid diagram given in (53).*

$$\begin{tikzcd}
 & \pi_{n+1}(G/PL)=P_{n+1} \arrow[rr, bend left, "b"] \arrow[dr] & & \pi_n(PL/O)=\Theta_n \arrow[rr, bend left, "0"] \arrow[dr, "c"] & & \pi_{n-1}(O) \\
\pi_{n+1}(G/O)=\Omega^{afr}_{n+1} \arrow[ur, "a"] \arrow[dr, "o"'] & & \pi_n(PL) \arrow[ur] \arrow[dr] & & \pi_n(G/O)=\Omega^{afr}_n \arrow[ur, "o"] \arrow[dr, "a"] & \\
 & \pi_n(O) \arrow[ur] \arrow[rr, bend right, "J"'] & & \pi_n(G)=\pi^s_n=\Omega^{fr}_n \arrow[ur] \arrow[rr, bend right] & & \pi_n(G/PL)=P_n
\end{tikzcd} \tag{53}$$

In (53) the mappings a, b and c are defined in (54).

$$\left\{\begin{array}{l}
a:\Omega^{afr}_{2n}\to P_{2n},\ a(M)=\left\{\begin{array}{l}\frac{1}{8}\sigma(M)\in\mathbb{Z}\text{if } n\equiv 0 \ (\text{mod } 2)\\ Kervaire(M)\in\mathbb{Z}_2\text{if } n\equiv 1 \ (\text{mod } 2)\end{array}\right\}\in P_{2n}.\\
b: P_{2n}\to\Theta_{2n-1},\ b(M)=\text{plumbing construction } \Sigma=\partial M.\\
c:\Theta_n\to\Omega^{afr}_n,\ c(\Sigma)=[\Sigma]\in\Omega^{afr}_n
\end{array}\right. \tag{54}$$

The image of b is denoted $bP_n \lhd \Theta_{n-1}$. Then if $\Sigma \in bP_n$, then $\Sigma = \partial V$, where V is a n-dimensional framed differentiable manifold. Furthermore, by considering the mapping c as $c : \Theta_n \to \pi_n(G/O)$, it sends an n-dimensional exotic sphere Σ to its fibre-homotopy trivialized stable normal bundle.

(7) (Kervaire-Milnor's braid $n = 4k + 2 \geq 5$). *For $n = 4k + 2 \geq 5$ the exact commutative braid diagram given in (53) becomes the one reported in (55).*

$$\begin{tikzcd}
P_{4k+3}=0 \arrow[rr, bend left, "b"] \arrow[dr] & & \Theta_{4k+2} \arrow[rr, bend left, "0"] \arrow[dr, "c"] & & \pi_{4k+1}(O) \arrow[rr, bend left, "J"] \arrow[dr, "c"] & & \pi_{4k+1}(G) \\
 & \pi_{4k+2}(PL) \arrow[ur] \arrow[dr] & & \pi_{4k+2}(G/O) \arrow[ur, "o"] \arrow[dr, "a"] & & \pi_{4k+1}(PL) \arrow[ur] \arrow[dr] & \\
\pi_{4k+2}(O)=0 \arrow[ur] \arrow[rr, bend right, "J"'] & & \pi_{4k+2}(G) \arrow[ur] \arrow[rr, bend right, "K"'] & & P_{4k+2}=\mathbb{Z}_2 \arrow[ur] \arrow[rr, bend right, "b"'] & & \Theta_{4k+1}
\end{tikzcd} \tag{55}$$

K is the Kervaire invariant on the $(4k+2)$-dimensional stable homotopy group of spheres

$$\left\{\begin{array}{rl}
K:\pi_{4k+2}(G)=\pi^s_{4k+2}= & \varinjlim_j \pi_{j+4k+2}(S^j)\\
= & \Omega^{fr}_{4k+2}\to P_{4k+2}=\mathbb{Z}_2.
\end{array}\right. \tag{56}$$

- *K is the surgery obstruction: $K = 0$ iff every $(4k + 2)$-dimensional framed differentiable manifold is framed cobordant to a framed exotic sphere.*

- *The exotic sphere group Θ_{4k+2} fits into the exact sequence (57).*

$$\begin{tikzcd} 0 \arrow[r] & \Theta_{4k+2} \arrow[r] & \pi_{4k+2}(G) \arrow[r, "K"'] & \mathbb{Z}_2 \arrow[r] & \ker(\pi_{4k+1}(PL) \to \pi_{4k+1}(G)) \arrow[r] & 0 \end{tikzcd} \tag{57}$$

- *$a : \pi_{4k+2}(G/O) \to \mathbb{Z}_2$ is the surgery obstruction map sending a normal map $(f, b) : M \to S^{4k+2}$ to the Kervaire invariant of M.*
- *$b : P_{4k+2} = \mathbb{Z}_2 \to \Theta_{4k+1}$ sends the generator $1 \in \mathbb{Z}_2$ to the boundary $b(1) = \Sigma^{4k+1} = \partial W$ of the Milnor plumbing W of two copies of TS^{2k+1} using the standard rank 2 quadratic form $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ over $\mathbb{Z}$ with Arf invariant 1. The subgroup $bP_{4k+2} \lhd \Theta_{4k+1}$ represents the $(4k+1)$-dimensional exotic spheres $\Sigma^{4k+1} = \partial V$ that are boundaries of framed $(4k + 2)$-dimensional differentiable manifolds V. If k is such that $K = 0$ (e.g., $k = 2$) then $bP_{4k+2} = \mathbb{Z}_2 \lhd \Theta_{4k+1}$ and if $\Sigma^{4k+1} = 1 \in bP_{4k+2}$, then the $(4k + 2)$-dimensional manifold $M = V \bigcup_{\Sigma^{4k+1}} D^{4k+2}$ is a PL manifold without a differentiable structure.*
- *For any $k \geq 1$ the following propositions are equivalent.*

  (i) *$K : \pi_{4k+2}(G) = \pi^s_{4k+2} \to \mathbb{Z}_2$ is $K = 0$.*
  (ii) *$\Theta_{4k+2} \cong \pi_{4k+2}(G)$.*
  (iii) $\ker(\pi_{4k+1}(PL) \to \pi_{4k+1}(G)) \cong \mathbb{Z}_2$.
  (iv) *Every simply-connected $(4k + 2)$-dimensional Poincaré complex X with a vector bundle reduction $\widetilde{\nu}_x : X \to BO$ of the Spivak normal fibration $\nu_x : X \to BG$ is homotopy equivalent to a closed $(4k + 2)$-dimensional differentiable manifold.*

Theorem 21 (Pontrjagin, Thom, Kervaire-Milnor).

(1) *Let bP_{n+1} denote the set of those h-cobordism classes of homotopy spheres which bound parallelizable manifolds.*[30] *For $n \neq 3$, there is a short exact sequence*

$$\begin{tikzcd} 0 \arrow[r] & bP_{n+1} \arrow[r] & \Theta_n \arrow[r] & \Theta_n/bP_{n+1} \arrow[r] & 0 \end{tikzcd} \tag{58}$$

where the left hand group is finite cyclic. Furthermore, there exists an homomorphism $J : \pi_n(SO) \to \pi^s_n$ such that Θ_n/bP_{n+1} injects into $\pi^s_n/J(\pi_n(SO))$ via the Pontrjagin-Thom construction. When $n \neq 2^j - 2$, the right hand group is isomorphic to $\pi^s_n/J(\pi_n(SO))$.

(2) *If Σ^n bounds a parallelizable manifold, it bounds a parallelizable manifold W such that $\pi_j(W) = 0$, $j < n/2$.*

(3) *For any $k \geq 1$, $bP_{2k+1} = 0$.*

[30] bP_{n+1} is a subgroup of Θ_n. If $\Xi_1, \Xi_2 \in bP_{n+1}$, with bounding parallelizable manifolds W_1 and W_2 respectively, then $\Xi_1 \sharp \Xi_2$ bounds the parallelizable manifold $W_1 \sharp W_2$, (commutative sum along the boundary).

Proof. For any manifold M with stably trivial normal bundle with framing φ, there is a homotopy class $p(M,\varphi)$, depending on the framed cobordism class of (M,φ). If $p(M) \subset \pi_n^s$ is the set of all $p(M,\varphi)$ where φ ranges over framings of the normal bundle, it follows that $0 \in p(M)$ iff M bounds a parallelizable manifold. (This is a result by Pontrjagin and Thom.) In particular, the set $p(S^n)$ has an explicit description. More precisely, the Whitehead J-homomorphism : $\pi_n(SO(r)) \to \pi_{n+r}(S^r)$ is defined by $J : (\alpha : S^n \to SO(r)) \mapsto (J(\alpha) : S^{n+r} \to S^r)$ that is the Pontrjagin-Thom map of $S^n \subset S^{n+r}$, with the framing $b_\alpha : S^n \times D^r \subset S^{n+r} = S^n \times D^r \bigcup D^{n+1} \times S^{r-1}$, $(x,y) \mapsto (x, \alpha(x)(y))$. Therefore, the map $J(\alpha) : S^{n+r} \to S^r$, is obtained by considering $S^{n+r} = (S^n \times D^r) \bigcup (D^{n+1} \times S^{r-1})$ and sending $(x,y) \in D^n \times D^r$ to $\alpha(x)y \in D^r/\partial D^r = S^r$ and $D^{n+1} \times S^{r-1}$ to the collapsed ∂D^r. Then $J : \pi_n(SO) = \lim_{\overrightarrow{r}} \pi_n(SO(r)) \to \lim_{\overrightarrow{r}} \pi_{n+r}(S^r) = \pi_n^s$ is the stable limit of the maps $J(\alpha)$ as $r \to \infty$, and $p(S^n)$ is the image $J(\pi_n(SO)) \subset \pi_n^s = \Omega_n^{fr}$, hence one has that to α there corresponds the framed cobordism class (S^n, b_α).

The characterization of global solutions of a PDE $E_k \subseteq J_n^k(W)$, in the category $\mathfrak{M}_\infty$, can be made by means of its integral bordism groups $\Omega_p^{E_k}$, $p \in \{0,1,\dots, n-1\}$. Let us shortly recall some fundamental definitions and results about.

Definition 15. Let $f_i : X_i \to E_k$, $f_i(X_i) \equiv N_i \subset E_k$, $i = 1,2$, be p-dimensional admissible compact closed smooth integral manifolds of E_k. The admissibility requires that N_i should be contained into some solution $V \subset E_k$, identified with a n-chain, with coefficients in A. Then, we say that they are *E_k-bordant* if there exists a $(p+1)$-dimensional smooth manifolds $f : Y \to E_k$, such that $\partial Y = X_1 \sqcup X_2$, $f|_{X_i} = f_i$, $i = 1,2$, and $V \equiv f(Y) \subset E_k$ is an admissible integral manifold of E_k of dimension $(p+1)$. We say that N_i, $i=1,2$, are *E_k-bordant* if there exists a $(p+1)$-dimensional smooth manifolds $f : Y \to J^k_{m|n}(W)$, such that $\partial Y = X_1 \sqcup X_2$, $f|_{X_i} = f_i$, $i = 1,2$, and $V \equiv f(Y) \subset J_n^k(W)$ is an admissible integral manifold of $J_n^k(W)$ of dimension $(p+1)$. Let us denote the corresponding bordism groups by $\Omega_p^{E_k}$ and $\Omega_p(E_k)$, $p \in \{0,1,\dots,n-1\}$, called respectively *p-dimensional integral bordism group* of E_k and *p-dimensional quantum bordism group* of E_k. Therefore these bordism groups work, for $p = (n-1)$, in the category of manifolds that are solutions of E_k, and $(J_n^k(W), E_k)$. Let us emphasize that singular solutions of E_k are, in general, (piecewise) smooth manifolds into some prolongation $(E_k)_{+s} \subset J_n^{k+s}(W)$, where the set, $\Sigma(V)$, of *singular points* of a solution V is a non-where dense subset of V. Here we consider *Thom-Boardman singularities*, i.e., $q \in \Sigma(V)$, if $(\pi_{k,0})_*(T_qV) \not\cong T_qV$. However, in the case where E_k is a differential equation of finite type, i.e., the symbols $g_{k+s} = 0$, $s \geq 0$, then it is useful to include also in $\Sigma(V)$, discontinuity points, $q, q' \in V$, with $\pi_{k,0}(q) = \pi_{k,0}(q') = a \in W$, or with $\pi_k(q) = \pi_k(q') = p \in M$, where $\pi_k = \pi \circ \pi_(k,0) : J_n^k(W) \to M$. We denote such a set by $\Sigma(V)_S$, and, in such cases we shall talk more precisely of *singular boundary* of V, like $(\partial V)_S = \partial V \setminus \Sigma(V)_S$. Such singular solutions are also called *weak solutions.*

Remark 3. Let us emphasize that weak solutions are not simply exotic solutions, introduced in Mathematical Analysis in order to describe "non-regular phenomena". But their importance is more fundamental in a theory of PDE's. In fact, by means of such solutions we can give a full algebraic topological characterization of PDE's. This can be well understood in Theorem 22 below, where it is shown the structural importance played by weak solutions. In this respect, let us, first, define some notation to distinguish between some integral bordisms group types.

$$\begin{tikzcd}
 & 0 \arrow[d] & 0 \arrow[d] & 0 \arrow[d] & \\
0 \arrow[r] & K^{E_k}_{n-1,w/(s,w)} \arrow[r] \arrow[d] & K^{E_k}_{n-1,w} \arrow[r] \arrow[d] & K^{E_k}_{n-1,s,w} \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & K^{E_k}_{n-1,s} \arrow[r] \arrow[d] & \Omega^{E_k}_{n-1} \arrow[r] \arrow[d] & \Omega^{E_k}_{n-1,s} \arrow[r] \arrow[d] & 0 \\
 & 0 \arrow[r] & \Omega^{E_k}_{n-1,w} \arrow[r] \arrow[d] & \Omega^{E_k}_{n-1,w} \arrow[r] \arrow[d] & 0 \\
 & & 0 & 0 &
\end{tikzcd} \tag{59}$$

Definition 16. Let $\Omega^{E_k}_{n-1}$, (resp. $\Omega^{E_k}_{n-1,s}$, resp. $\Omega^{E_k}_{n-1,w}$), be the *integral bordism group* for $(n-1)$-dimensional smooth admissible regular integral manifolds contained in E_k, borded by smooth regular integral manifold-solutions, (resp. piecewise-smooth or singular solutions, resp. singular-weak solutions), of E_k.

Theorem 22. *Let $E_k \subset J^k_n(W)$ be a PDE on the fiber bundle $\pi : W \to M$, with* $\dim(W) = m + n$ *and* $\dim M = n$.

(1) *One has the exact commutative diagram (59). Therefore, one has the canonical isomorphisms:*

$$\begin{cases} K^{E_k}_{n-1,w/(s,w)} \cong K^{E_k}_{n-1,s}; \qquad \Omega^{E_k}_{n-1}/K^{E_k}_{n-1,s} \cong \Omega^{E_k}_{n-1,s}; \\ \Omega^{E_k}_{n-1,s}/K^{E_k}_{n-1,s,w} \cong \Omega^{E_k}_{n-1,w}; \; \Omega^{E_k}_{n-1}/K^{E_k}_{n-1,w} \cong \Omega^{E_k}_{n-1,w}. \end{cases} \tag{60}$$

If E_k is formally integrable, then one has the following isomorphisms:

$$\Omega^{E_k}_{n-1} \cong \Omega^{E_\infty}_{n-1} \cong \Omega^{E_\infty}_{n-1,s}; \quad \Omega^{E_k}_{n-1,w} \cong \Omega^{E_\infty}_{n-1,w}. \tag{61}$$

(2) *Let $E_k \subset J^k_n(W)$ be a quantum super PDE that is formally integrable, and completely integrable. We shall assume that the symbols $g_{k+s} \neq 0$, $s = 0, 1$. (This excludes the case $k = \infty$.) Then one has the following isomorphisms: $\Omega^{E_k}_{p,s} \cong \Omega^{E_k}_{p,w} \cong \Omega_p(E_k)$, with $p \in \{0, \dots, n-1\}$.*

(3) *Let $E_k \subset J^k_n(W)$ be a PDE, that is formally integrable and completely integrable. One has the following isomorphisms: $\Omega^{E_k}_{n-1,w} \cong \Omega_{n-1}(E_k) \cong \Omega^{E_{k+h}}_{n-1,w} \cong \Omega^{E_\infty}_{n-1,w} \cong \Omega_{n-1,w}(E_{k+h}) \cong \Omega_{n-1}(E_\infty)$.*

Proof. See [30, 40].

In order to distinguish between manifolds V representing singular solutions, where $\Sigma(V)$ has no discontinuities, and integral manifolds where $\Sigma(V)$ contains discontinuities, we can also consider "conservation laws" valued on integral manifolds N representing the integral bordism classes $[N]_{E_k} \in \Omega^{E_k}_p$.

Definition 17. Set

$$\left\{\begin{array}{l} \Im(E_k) \equiv \bigoplus_{p\geq 0} \dfrac{\Omega^p(E_k) \cap d^{-1}(C\Omega^{p+1}(E_k))}{d\Omega^{p-1}(E_k) \oplus \{C\Omega^p(E_k) \cap d^{-1}(C\Omega^{p+1}(E_k))\}} \\ \qquad\quad \equiv \bigoplus_{p\geq 0} \Im(E_k)^p. \end{array}\right. \tag{62}$$

Here $C\Omega^p(E_k)$ denotes the space of all Cartan quantum p-forms on E_k. Then we define *integral characteristic numbers* of N, with $[N]_{E_k} \in \Omega^{E_k}_p$, the numbers $i[N] \equiv< [N]_{E_k}, [\alpha] >\in \mathbb{R}$, for all $[\alpha] \in \Im(E_k)^p$.

Then, one has the following theorems.

Theorem 23. *Let us assume that $\Im(E_k)^p \neq 0$. One has a natural homomorphism:*

$$\left\{\begin{array}{l} \underline{j}_p : \Omega^{E_k}_p \to Hom(\Im(E_k)^p; \mathbb{R}), \quad [N]_{E_k} \mapsto \underline{j}_p([N]_{E_k}), \\ \underline{j}_p([N]_{E_k})([\alpha]) = \int_N \alpha \equiv< [N]_{E_k}, [\alpha] > . \end{array}\right. \tag{63}$$

Then, a necessary condition that $N' \in [N]_{E_k}$ is the following: $i[N] = i[N']$, $\forall [\alpha] \in \Im(E_k)^p$. Furthermore, if N is orientable then above condition is sufficient also in order to say that $N' \in [N]_{E_k}$.

Proof. See [27, 28, 30, 31].

Corollary 24. *Let $E_k \subseteq J^k_n(W)$ be a PDE. Let us consider admissible p-dimensional, $0 \leq p \leq n-1$, orientable integral manifolds. Let $N_1 \in [N_2]_{E_k} \in \Omega^{E_k}_p$, then there exists a $(p+1)$-dimensional admissible integral manifold $V \subset E_k$, such that $\partial V = N_1 \sqcup N_2$, where V is without discontinuities iff the integral numbers of N_1 and N_2 coincide.*

Above considerations can be generalized to include more sophisticated solutions of PDEs.

Definition 18. Let $E_k \subset J^k_n(W)$ be a PDE and let B be an algebra. Let us consider the following chain complex *(bar chain complex of E_k)*: $\{\bar{C}_\bullet(E_k; B), \partial\}$, induced by B on the corresponding bar chain complex of E_k, i.e., $\{\bar{C}_\bullet(E_k; B), \partial\}$. (See [27, 28, 30].) More precisely $\bar{C}_p(E_k; B)$ is the free two-sided B-module of formal linear combinations with coefficients in B, $\sum \lambda_i c_i$, where c_i is a singular p-chain $f : \triangle^p \to E_k$, that extends on a neighborhood $U \subset \mathbb{R}^{p+1}$, such that f on U is differentiable and $Tf(\triangle^p) \subset \mathbf{E}^k_n$, where $\mathbf{E}^k_n$ is the Cartan distribution of E_k.

Theorem 25. *The homology $\bar{H}_\bullet(E_k; B)$ of the bar chain complex of E_k is isomorphic to* (closed) bar integral singular (p)-bordism groups*, with coefficients in B, of E_k: ${}^B\bar{\Omega}^{E_k}_{p,s} \cong \bar{H}_q(E_k; B) \cong (\bar{\Omega}^{E_k}_{p,s} \otimes_{\mathbb{R}} B)$, $p \in \{0, 1, \dots, n-1\}$. (If $B = \mathbb{R}$ we omit the apex B). The relation between closed bordism and bordism, is given by the following unnatural isomorphism*[31]*:*

$$Bor_\bullet(E_k; B) \cong {}^B\underline{\Omega}_{\bullet,s}(E_k) \bigoplus Cyc_\bullet(E_k; B). \tag{64}$$

Proof. It follows from above results, and the following exact commutative diagram naturally associated to the bar quantum chain complex of E_k.

$$\begin{tikzcd}
 & & 0 \arrow[d] & 0 \arrow[d] & & \\
 & 0 \arrow[r] & \bar{B}_\bullet(E_k;B) \arrow[r] \arrow[d] & \bar{Z}_\bullet(E_k;B) \arrow[r] \arrow[d] & \bar{H}_\bullet(E_k;B) \arrow[r] & 0 \\
 & & \bar{C}_\bullet(E_k;B) \arrow[r, equal] \arrow[d] & \bar{C}_\bullet(E_k;B) \arrow[d] & & \\
0 \arrow[r] & {}^B\underline{\Omega}_{\bullet,s}(E_k) \arrow[r] & \bar{B}or_\bullet(E_k;B) \arrow[r] \arrow[d] & \bar{C}yc_\bullet(E_k;B) \arrow[r] \arrow[d] & 0 & \\
 & & 0 & 0 & &
\end{tikzcd} \tag{65}$$

where $\bar{B}_\bullet(E_k; B) = \ker(\partial|_{\bar{C}_\bullet(E_k;B)})$, and $\bar{Z}_\bullet(E_k; B) = \mathrm{im}\,(\partial|_{\bar{C}_\bullet(E_k;B)})$, $\bar{H}_\bullet(E_k; B) = \bar{Z}_\bullet(E_k; B)/\bar{B}_\bullet(E_k; B)$. Furthermore,

[31] Note that if X is a compact space with boundary ∂X, the boundary of $X \times I$, $I \equiv [0, 1] \subset \mathbb{R}$, is $\partial(X \times I) = (X \times \{0\}) \bigcup (\partial X \times I) \bigcup (X \times \{1\}) \equiv X_0 \bigcup P \bigcup X_1$, with $X_0 \equiv X \times \{0\}$, $X_1 \equiv X \times \{1\}$, $P \equiv \partial X \times I$. One has $\partial P = (\partial X \times \{0\}) \bigcup (\partial X \times \{1\}) = \partial X_0 \bigcup \partial X_1$. On the other hand, whether X is closed, then $\partial(X \times I) = X_0 \bigcup X_1$. Furthermore we shall denote by $[N]_{E_k}$ the equivalence class of the integral admissible bordism of $N \subset E_k$, even if N is not necessarily closed. So, if N is closed one has $[N]_{E_k} \in {}^B\Omega^{E_k}_{\bullet,s}$, and if N is not closed one has $[N]_{E_k} \in \bar{B}_\bullet(E_k; B)$.

$$\begin{cases} b \in [a] \in \bar{B}or_{\bullet}(E_k; B) \Rightarrow a - b = \partial c, \quad c \in \bar{C}_{\bullet}(E_k; B), \\ b \in [a] \in \bar{C}yc_{\bullet}(E_k; B) \Rightarrow \partial(a - b) = 0, \\ b \in [a] \in {}^{A}\underline{\Omega}_{\bullet,s}(E_k) \Rightarrow \left\{ \begin{array}{l} \partial a = \partial b = 0 \\ a - b = \partial c, \quad c \in \bar{C}_{\bullet}(E_k; B) \end{array} \right\} . \end{cases}$$

Furthermore, one has the following canonical isomorphism: ${}^{B}\underline{\Omega}_{\bullet,s}(E_k) \cong \bar{H}_{\bullet}(E_k; B)$. As $\bar{C}_{\bullet}(E_k; B)$ is a free two-sided projective B-module, one has the unnatural isomorphism: $\bar{B}or_{\bullet}(E_k; B) \cong {}^{B}\underline{\Omega}_{\bullet,s}(E_k) \bigoplus \bar{C}yc_{\bullet}(E_k; B)$.

The spaces of conservation laws of PDEs, identify *Hopf algebras*. (Hopf algebras considered here are generalizations of usual Hopf algebras [28].)

Definition 19. The *full space of p-conservation laws*, (or *full p-Hopf algebra*), of E_k is the following one: $\mathbf{H}_p(E_k) \equiv \mathbb{R}^{\Omega_p^{E_k}}$. We call *full Hopf algebra*, of E_k, the following: $\mathbf{H}_{n-1}(E_\infty) \equiv \mathbb{R}^{\Omega_{n-1}^{E_\infty}}$.

Definition 20. The *space of (differential) conservation laws* of $E_k \subset J^k_n(W)$, is $\mathfrak{C}ons(E_k) = \mathfrak{I}(E_\infty)^{n-1}$.

Theorem 26. *The full p-Hopf algebra of a PDE $E_k \subset J^k_n(W)$ has a natural structure of Hopf algebra (in extended sense). Furthermore, the space of conservation laws of E_k has a canonical representation in $\mathbf{H}_{n-1}(E_\infty)$.*

Proof. See [27, 28].

Theorem 27. *Set: $\mathbf{H}_{n-1}(E_k) \equiv \mathbb{R}^{\Omega_{n-1}^{E_k}}$, $\mathbf{H}_{n-1,s}(E_k) \equiv \mathbb{R}^{\Omega_{n-1,s}^{E_k}}$, $\mathbf{H}_{n-1,w}(E_k) \equiv \mathbb{R}^{\Omega_{n-1,w}^{E_k}}$. One has the exact and commutative diagram reported in (66), that define the following spaces: $\mathbf{K}^{E_k}_{n-1,w/(s,w)}$, $\mathbf{K}^{E_k}_{n-1,w}$, $\mathbf{K}^{E_k}_{n-1,s,w}$, $\mathbf{K}^{E_k}_{n-1,s}$.*

$$\begin{tikzcd}
 & 0 & 0 & 0 & \\
0 & \mathbf{K}^{E_k}_{n-1,w/(s,w)} \arrow[l] \arrow[u] & \mathbf{K}^{E_k}_{n-1,w} \arrow[l] \arrow[u] & \mathbf{K}^{E_k}_{n-1,s,w} \arrow[l] \arrow[u] & 0 \arrow[l] \\
0 & \mathbf{K}^{E_k}_{n-1,s} \arrow[l] \arrow[u] & \mathbf{H}_{n-1}(E_k) \arrow[l] \arrow[u] & \mathbf{H}_{n-1,s}(E_k) \arrow[l] \arrow[u] & 0 \arrow[l] \\
 & 0 \arrow[u] & \mathbf{H}_{n-1,w}(E_k) \arrow[l] \arrow[u] & \mathbf{H}_{n-1,w}(E_k) \arrow[l] \arrow[u] & 0 \arrow[l] \\
 & & 0 \arrow[u] & 0 \arrow[u] &
\end{tikzcd} \tag{66}$$

More explicitly, one has the following canonical isomorphisms:

$$\left\{\begin{array}{l}\mathbf{K}^{E_k}_{n-1,w/(s,w)} \cong \mathbf{K}^{K^{E_k}_{n-1,s}};\\ \mathbf{K}^{E_k}_{n-1,w}/\mathbf{K}^{E_k}_{n-1,s,w} \cong \mathbf{K}^{K^{E_k}_{n-1,w/(s,w)}};\\ \mathbf{H}_{n-1}(E_k)/\mathbf{H}_{n-1,s}(E_k) \cong \mathbf{K}^{E_k}_{n-1,s};\\ \mathbf{H}_{n-1}(E_k)/\mathbf{H}_{n-1,w}(E_k) \cong \mathbf{K}^{E_k}_{n-1,w}\\ \cong \mathbf{H}_{n-1,s}(E_k)/\mathbf{H}_{n-1,w}(E_k) \cong \mathbf{K}^{E_k}_{n-1,s,w}.\end{array}\right. \tag{67}$$

Furthermore, under the same hypotheses of Theorem 4.44(2) one has the following canonical isomorphism: $\mathbf{H}_{n-1,s}(E_k) \cong \mathbf{H}_{n-1,w}(E_k)$. *Furthermore, we can represent differential conservation laws of* E_k *in* $\mathbf{H}_{n-1,w}(E_k)$.

Proof. The proof follows directly for duality from the exact commutative diagram (59).

Definition 21. We define *crystal obstruction* of E_k the following quotient algebra: $cry(E_k) \equiv \mathbf{H}_n((E_k)_\infty)/\mathbb{R}^{\Omega_n}$. We say that E_k is a 0-*crystal PDE* if $cry(E_k) = 0$.

Remark 4. An extended 0-crystal equation $E_k \subset J^k_n(W)$ does not necessitate to be a 0-crystal PDE. In fact E_k is an extended 0-crystal PDE if $\Omega^{E_k}_{n,w} = 0$. This does not necessarily imply that $\Omega^{E_k}_n = 0$.

Corollary 28. *Let* $E_k \subset J^k_n(W)$ *be a* 0-*crystal PDE. Let* $N_0, N_1 \subset E_k$ *be two closed compact* $(n-1)$-*dimensional admissible integral manifolds of* E_k *such that* $X \equiv N_0 \sqcup N_1 \in [0] \in \Omega_n$. *Then there exists a smooth solution* $V \subset E_k$ *such that* $\partial V = X$. *(See also [34–37, 40].)*

Let us consider, now, the interaction between surgery and global solutions in PDE's of the category $\mathfrak{M}_\infty$. Since the surgery is a proceeding to obtain manifolds starting from other ones, or eventually from $\varnothing$, in any theory of PDE's, where we are interested to characterize nontrivial solutions, surgery is a fundamental tool to consider. We have just seen that integral bordism groups are the main structures able to characterize global solutions of PDE's. On the other hand surgery is strictly related to bordism groups, as it is well known in algebraic topology. Therefore, in this section, we shall investigate as integral surgery interacts with integral bordism groups.

Definition 22. Let $\pi : W \to M$ be a smooth fiber bundle of dimension $m+n$ over a n-dimensional manifold M. Let $E_k \subset J^k_n(W)$ be a PDE of order k for n-dimensional submanifolds of W. Let $N \subset E_k$ be an admissible integral manifold of dimension $p \in \{0, 1, \cdots, n-1\}$. Therefore, there exists a solution $V \subset E_k$ such that $N \subset V$. An *admissible integral i-surgery*, $0 \leq i \leq n-1$, on N is the procedure of constructing a new p-dimensional admissible integral manifold N':

$$N' \equiv \overline{N \setminus S^i \times D^{p-1-i}} \bigcup_{S^i \times S^{p-2-i}} D^{i+1} \times S^{p-2-i} \tag{68}$$

such that $D^{i+1} \times S^{p-2-i} \subset V$. Here $\overline{Y}$ is the closure of the topological subspace $Y \subset X$, i.e., the intersection of all closed subsets $Z \subset X$, with $Y \subset Z$.

Theorem 29. *Let $N_1, N_0 \subset E_k$ be two integral compact (non-necessarily closed) admissible p-dimensional submanifolds of the PDE $E_k \subset J^k_n(W)$, such that there is an admissible $(p+1)$-dimensional integral manifold $V \subset E_k$, such that $\partial V = N_0 \sqcup N_1$. Then it is possible to identify an integral admissible manifold N'_1, obtained from N_1 by means of an integral i-surgery, iff N'_1 is integral bording with N_0, i.e., $N'_1 \in [N_0]_{E_k}$.*

Proof. As it is well known N'_1 is bording with N_0, i.e., there exists a $(p+1)$-dimensional manifold Y, with $\partial Y = N_0 \sqcup N'_1$. More precisely we can take $Y = N_0 \times I \bigcup D^{i+1} \times D^{p-1-i}$. By the way, in order N'_1 should be integral admissible, it is necessary that should be contained into a solution passing from N_1. Then N'_1 is integral bording with N_1, hence it is also integral bording with N_0.

Theorem 30 (Integral h-cobordism in Ricci flow PDE).

The generalized Poincaré conjecture, for any dimension $n \geq 1$ is true, i.e., any n-dimensional homotopy sphere M is homeomorphic to S^n: $M \approx S^n$.

For $1 \leq n \leq 6$ one has also that M is diffeomorphic to S^n: $M \cong S^n$. But for $n \geq 7$, it does not necessitate that M is diffeomorphic to S^n. This happens when the Ricci flow equation, under the homotopy equivalence full admissibility hypothesis, *(see below for definition), becomes a 0-crystal.*

Moreover, under the sphere full admissibility hypothesis*, the Ricci flow equation becomes a 0-crystal in any dimension $n \geq 1$.*

Proof. Let us first consider the following lemma.

Lemma 14. *Let $N_0, N_1 \subset (RF)$ be two space-like connected smooth compact Cauchy manifolds at two different instant $t_0 \neq t_1$, identified respectively with two different Riemannian structures (M, γ_0) and (M, γ_1). Then one has $N_0 \approxeq N_1$.*

Proof. In fact this follows directly from the fact the diffeomorphisms $(M, \gamma_i) \cong N_i$, $i = 0, 1$, and from the fact that any Riemannian metric on M can be continuously deformed into another one. More precisely we shall prove that there exists two continuous functions $f : N_0 \to N_1$ and $h : N_1 \to N_0$, such that $h \circ f \simeq 1_{N_0}$ and $f \circ h \simeq 1_{N_1}$. Realy we can always find homotopies $F, G : I \times M \to M$, that continuosly deform γ_1 into γ_0 and vice versa. More precisely $F_0 = id_M$, $G_0 = id_M$, $F_1^* \gamma_1 = \gamma_0$, and $G_1^* \gamma_0 = \gamma_1$. Therefore, we get $G_1 \circ F_1 \simeq G_0 \circ F_0 = 1_M$ and $F_1 \circ G_1 \simeq F_0 \circ G_0 = 1_M$. Thus we can identify f with F_1 and h with G_1.

Lemma 15. *Let $N_0, N_1 \subset (RF)_{+\infty}$ be two space-like, smooth, compact closed, homotopy equivalent Cauchy n-manifolds, corresponding to two different times $t_0 \neq t_1$. Then N_0 and N_1, have equal all the the integral characteristic numbers.*

Proof. Since we have assumed $N_0 \approxeq N_1$, there are two mappings $f : N_0 \to N_1$ and $h : N_1 \to N_0$, such that $h \circ f \simeq 1_{N_0}$ and $f \circ h \simeq 1_{N_1}$. These mappings for functorial property induce canonical homomorphisms between the groups $\pi_p(N_i)$, $H_p(N_i)$, $H^p(N_i)$, $i = 0, 1$, that we shall simply denote by f_* and h_* or f^* and h^* according to respectively the direct or inverse character of functoriality. Then one has the following properties $f_* \circ h_* = 1$, $h_* \circ f_* = 1$ and similarly for the controvariant cases, i.e., $f^* \circ h^* = 1$, $h^* \circ f^* = 1$. These relations means that the induced morphisms f_* and f^* are isomorphisms with inverse h_* and h^* respectively. In other words one has the isomorphisms $\pi_p(N_0) \cong \pi_p(N_1)$, $H_p(N_i) \cong H_p(N_i)$, $H^p(N_i) \cong H^p(N_i)$. As a by-product we get also the commutative diagram reported in (69).

$$\begin{tikzcd}
H_n(N_0;\mathbb{R}) \times H^n(N_0;\mathbb{R}) \arrow[dd, equal, "(f_*{,}(f^{-1})^*)"', "\wr"] \arrow[dr, "<{,}>"] & \\
& \mathbb{R} \\
H_n(N_1;\mathbb{R}) \times H^n(N_1;\mathbb{R}) \arrow[ur, "<{,}>"'] &
\end{tikzcd} \tag{69}$$

Since $H_n(N_i;\mathbb{R}) \cong \mathbb{R} \cong H^n(N_i;\mathbb{R})$, $i = 0, 1$, let the isomorphism f_* be identified with a non-zero number $\lambda \in \mathbb{R} \setminus \{0\}$, then $(f^{-1})^* = 1/\lambda$, and we get that $< f_*[N_0], (f^{-1})^*\alpha >=< \lambda, \mu/\lambda >= \mu$ where μ is the number that represents the n-differential form α. On the other hand one has $< [N_0], \alpha >= 1.\mu = \mu$.

Lemma 16. *Under the same hypotheses of Lemma 15, let us add that we assume admissible orientable Cauchy manifolds only. Then $N_0 \in [N_1] \in \Omega_n^{(RF)+\infty}$. In other words, $N = \partial V$, where V is a smooth solution, iff $< [\alpha], [N] >= 0$ for all the conservation laws α.*

Proof. If we assume admissible only orientable Cauchy manifolds, then the canonical homomorphism $j_n : \Omega_n^{(RF)+\infty} \to (\mathcal{I}(RF)^n_{+\infty})^*$ is injective, (see [27]), therefore $N_0 \in [N_1] \in \Omega_n^{(RF)+\infty}$ iff N_0 and have equal all integral characteristic numbers.

Since Lemma 16 is founded on the assumption that the space of conservation laws of (RF) is not zero, in the following lemma we shall prove that such an assumption is true.

Lemma 17. *The space $\mathcal{I}(RF)^n_{+\infty} \cong E_1^{0,n}$ of conservation laws of the (RF) is not zero. In fact any differential n-form given in (70) is a conservation law of (RF).*[32]

[32] $E_1^{0,n}$ is the spectral term, in the Cartan spectral sequence of a PDE $E_k \subset J^k_n(W)$, just representing the conservation laws space of E_k. (See e.g., [27, 28, 31].)

$$\omega = Tdx^1 \wedge \cdots \wedge dx^n + X^p(-1)^p dt \wedge dx^1 \wedge \cdots \wedge \widehat{dx^p} \wedge \cdots \wedge dx^n \tag{70}$$

with

$$\left\{ \begin{array}{l} T = g_{ij}\varphi^{ij} \\ X^p = \kappa \int \left(R_{ij}(g)\,\varphi^{ij} - R^{ij}(\varphi)\,g_{ij} \right)\,dx^p + c^p \end{array} \right\} : \left\{ \begin{array}{l} \varphi^{ij}{}_{,t} - \kappa R^{ij}(\varphi) = 0 \\ g_{ij,t} + \kappa R_{ij}(g) = 0 \end{array} \right\}. \tag{71}$$

$c^p \in \mathbb{R}$ are arbitrary constants and φ^{ij}, $1 \leq i, j \leq n$, are functions on $\mathbb{R} \times M$, symmetric in the indexes, solutions of the equation given in (71).

Proof. Let us prove that $d\omega|_V = 0$ for any solution V of (RF). In fact, by a direct calculation we get

$$d\omega = \left[(g_{ij,t} + \kappa R_{ij}(g))\varphi^{ij} + g_{ij}(\varphi^{ij}{}_{,t} - \kappa R^{ij}(\varphi))\right] dt \wedge dx^1 \wedge \cdots \wedge dx^n. \tag{72}$$

Therefore, the conservation laws in (70) are identified with the solutions of the PDE given in (71) for φ_{ij}. This is an equation of the same type of the Ricci flow equation, hence its set of solutions is not empty.

Now, let M belong to the same integral bordism class of S^n in (RF): $M \in [S^n] \in \Omega_n^{(RF)}$. It follows from Theorems 3 and 2, that M is necessarily homeomorphic to S^n. However, if $n \geq 4$, the smooth solution V such that $\partial V = M \sqcup S^n$ does not necessitate to be a trivial h-cobordism. This happens iff the homotopy equivalence $f : M \approxeq S^n$ is such that $f \simeq 1_{S^n}$. (See Theorem 7.) This surely is the case at low dimensions $n = 1, 2, 3, 5, 6$, and also for $n = 4$, if the smooth Poincaré conjecture holds. But for $n \geq 7$ an homotopy sphere may have different structures with respect to this property. (See Table 8, Lemmas 10 and 11.) In fact it is well known that there are homotopy spheres characterized by rational Pontrjagin numbers. Since rational Pontrjagin classes $p_q \in H^{4q}(M; \mathbb{Q})$ are homeomorphic invariants, such manifolds cannot admit a differentiable structure, taking into account the fact that the signature is a topological invariant. Such homotopy spheres are obtained by gluing a disk D^n, along its boundary S^{n-1}, with the boundary of a disk-D^q-fiber bundle over a sphere S^m, $E \to S^m$, such that $q = n - m$. When the $(n-1)$-dimensional boundary ∂E is diffeomorphic to S^{n-1}, gives to $\widetilde{E} \equiv E \bigcup D^n$ a differentiable structure. But whether $\partial E \approx S^{n-1}$, Σ^n cannot, in general, have a differentiable structure, since it is characterized by rational Pontrjagin numbers. Therefore there are exotic spheres, (for example $\Sigma^{n-1} \equiv \partial E$), that are homeomorphic, but not diffeomorphic to S^{n-1}. In such cases the solution V of the Ricci flow equation such that $\partial V = \Sigma^{n-1} \sqcup S^{n-1}$, cannot be, in general, a trivial h-cobordism.

By conclusion we get that not all n-dimensional homotopy spheres M can, in general, belong to the same integral boundary class of $[S^n] \in \Omega_n^{(RF)}$, even if there exist singular solutions $V \subset (RF)$ such that $\partial V = M \sqcup S^n$. In fact, it does not necessitate, in general, that V should be a trivial h-cobordism, i.e., that the homotopy equivalence between M and S^n should be a diffeomorphism

Table 8 Calculated groups Θ_n for $1 \leq n \leq 20$ and some related groups

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Θ_n	0	0	0	0	0	0	$\mathbb{Z}_{28}$	$\mathbb{Z}_2$	$\mathbb{Z}_8$	$\mathbb{Z}_6$	$\mathbb{Z}_{992}$	0	$\mathbb{Z}_3$	$\mathbb{Z}_2$	$\mathbb{Z}_{16256}$	$\mathbb{Z}_2$	$\mathbb{Z}_{16}$	$\mathbb{Z}_{16}$	$\mathbb{Z}_{523264}$	$\mathbb{Z}_{24}$
bP_{n+1}	0	0	0	0	0	0	$\mathbb{Z}_{28}$	0	$\mathbb{Z}_2$	0	$\mathbb{Z}_{992}$	0	0	0	$\mathbb{Z}_{8128}$	0	$\mathbb{Z}_2$	0	$\mathbb{Z}_{261632}$	0
Θ_n/bP_{n+1}	0	0	0	0	0	0	0	$\mathbb{Z}_2$	$\mathbb{Z}_4$	$\mathbb{Z}_6$	0	0	$\mathbb{Z}_3$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_8$	$\mathbb{Z}_{16}$	$\mathbb{Z}_2$	$\mathbb{Z}_{24}$
π_n^s/J	0	$\mathbb{Z}_2$	0	0	0	$\mathbb{Z}_2$	0	$\mathbb{Z}_2$	$\mathbb{Z}_4$	$\mathbb{Z}_6$	0	0	$\mathbb{Z}_3$	$\mathbb{Z}_4$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_8$	$\mathbb{Z}_{16}$	$\mathbb{Z}_2$	$\mathbb{Z}_{24}$
π_n^s	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}_{24}$	0	0	$\mathbb{Z}_2$	$\mathbb{Z}_{240}$	$\mathbb{Z}_4$	$\mathbb{Z}_8$	$\mathbb{Z}_6$	$\mathbb{Z}_{504}$	0	$\mathbb{Z}_3$	$\mathbb{Z}_4$	$\mathbb{Z}_{960}$	$\mathbb{Z}_4$	$\mathbb{Z}_{16}$	—	—	—
J	$\mathbb{Z}_2$	0	$\mathbb{Z}_{24}$	0	0	0	$\mathbb{Z}_{240}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	0	$\mathbb{Z}_{504}$	0	0	0	$\mathbb{Z}_{480}$	$\mathbb{Z}_2$	$\mathbb{Z}_2$	—	—	—

$\dim_{\mathbb{Z}} \Theta_n =$ number of differential structures on the n-dimensional homotopy sphere

$bP_{n+1} \lhd \Theta_n$: subgroup of n-dimensional homotopy spheres bounding parallelizable manifolds

bP_{n+1} is a finite cyclic group that vanishes if n is even

Kervaire-Milnor formula: $\dim_{\mathbb{Z}} bP_{4n} = \frac{3-(-1)^n}{2} 2^{2n-2}(2^{2n-1}-1)Numerator(\frac{B_{4n}}{4n})$, $n \geq 2$, with B_{4n} Bernoulli numbers

$\dim_{\mathbb{Z}} bP_{4n+2} = 0$, $n = 0, 1, 3, 7, 15$; $\dim_{\mathbb{Z}} bP_{4n+2} = 0$, or $\mathbb{Z}_2$, $n = 31$; $\dim_{\mathbb{Z}} bP_{4n+2} = \mathbb{Z}_2$ otherwise

$\pi_n^s \equiv \underset{\overrightarrow{k}}{\lim} \pi_{n+k}(S^k)$: stable homotopy groups or n-stems. (Serre's theorem. The groups π_n^s are finite)

There is an injective map $\Theta_n/bP_{n+1} \to \pi_n^s/J$ where J is the image of Whitehead's J-homomorphisms $\pi_n(SO) \to \pi_n^s$

For $n = 0$ one has $\pi_n^s = \mathbb{Z}$ and $J = 0$

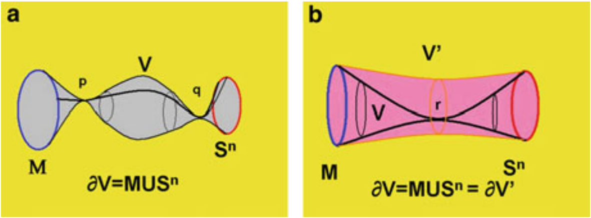


Fig. 4 Neck-pinching singular solutions type, V, $\partial V = M \sqcup S^n$, in Ricci flow equations, with singular points p, q in (**a**) and r in (**b**). In (**b**) is reported also a smooth solution V', bording M and S^n as well as a neck-pinching singular solution V bording the same manifolds

of S^n isotopic to the identity. This has, as a by-product, that in general M is only homeomorphic to S^n but not diffeomorphic to S^n. In order to better understand this aspect in the framework of PDE's algebraic topology, let us first show how solutions with neck-pinching singular points are related to smooth solutions. Recall the commutative diagram in Theorem 2.24 in [27], here adapted in (73) to $(RF)_{+\infty}$ and in dimension $p = n$.

$$\begin{tikzcd}
 & & 0 \arrow[d] & \\
 & \Omega_n^{(RF)_{+\infty}} \arrow[r, "i_n"] \arrow[d, "j_n"'] & \Omega_{n,s}^{(RF)_{+\infty}} \cong \bar{H}_n((RF)_{+\infty};\mathbb{R}) \arrow[d] & \\
0 \arrow[r] & (\mathcal{I}((RF)_{+\infty}))^* \arrow[r] & (\bar{H}_n((RF)_{+\infty};\mathbb{R}))^* \arrow[r] & 0
\end{tikzcd} \tag{73}$$

There $\bar{H}^n((RF)_{+\infty};\mathbb{R})$ is the n-dimensional bar de Rham cohomology of $(RF)_{+\infty}$. The isomorphism $\Omega_{n,s}^{(RF)_{+\infty}} \cong \bar{H}_n((RF)_{+\infty};\mathbb{R})$ is a direct by-product of the exact commutative diagram in Definition 4.8(3) in [28]. Then, taking into account that in solutions of $(RF)_{+\infty}$ cannot be present Thom-Boardman singularities, it follows that solutions bording smooth Cauchy manifolds in the bordism classes of $\widehat{\Omega_n^{(RF)_{+\infty}}} \equiv i_n(\Omega_n^{(RF)_{+\infty}}) \triangleleft \Omega_{n,s}^{(RF)_{+\infty}}$, can have singularities of neck-pinching type. (See Fig. 4a.)

From Corollary 2.5 in [27] it follows that if M is an homotopy sphere belonging to the integral bordism class $[S^n] \in \widehat{\Omega_n^{(RF)_{+\infty}}}$, one has surely $M \sqcup S^n = \partial V'$, for some smooth solution V' of (RF), but can be also $M \sqcup S^n = \partial V$ for some solution $V \subset (RF)_{+\infty}$ having some neck-pinching singularity. (See Fig. 4b.) In general V cannot be considered isotopic to V'. However M is diffeomorphic to S^n, (the diffeomorphism is that induced by the smooth solution V'), and all the singular points, in the neck-pinching singular solutions, bording M with S^n, are "solved" by the smooth bordism V'. Let us also emphasize that if an n-dimensional homotopy sphere $M \in [S^n] \in \Omega_n^{(RF)_{+\infty}}$, i.e., there exists a smooth solution $V \subset (RF)_{+\infty}$ such that $\partial V = M \sqcup S^n$, then the characteristic flow on V is without singular points, hence from Theorem 2 it follows that $V \cong M \times I$, and $V \cong S^n \times I$, hence $M \cong S^n$. If this happens for all n-dimensional homotopy sphere, then $\Theta_n = 0$ and vice versa. However, it is well known that there are homotopy spheres of dimensions $n \geq 7$ for the which $\Theta_n \neq 0$. (For example the Milnor spheres.) This is equivalent to say that $\pi_0(Diff_+(S^{n-1})) \neq 0$, since $\Theta_n \cong \pi_0(Diff_+(S^{n-1}))$ (Smale). This happens when there are homotopy spheres that bound non-contractible manifolds. In fact, if there exists a trivial h-bordism V bording S^n with M, then $W = V \bigcup_{S^n} D^{n+1} \cong D^{n+1}$ and $\partial W = M$. However, since the conservation laws of (RF) depend on a finite derivative order (second order), the fact that all n-dimensional homotopy spheres have the same integral characteristic numbers of the sphere S^n, implies that there are smooth integral manifolds bording them at finite order. There the symbols of the Ricci flow equation, and its finite order prolongations, are not trivial ones, hence in general such solutions present Thom-Boardman singular points. As a by-product of Theorem 2.25 in [27], (see also [28]), and Theorems 2.1, 2.12 and 3.6 in [30] between such solutions, there are ones that are not smooth, but are topological solutions inducing the homeomorphisms between M and S^n: $M \approx S^n$. Therefore, if we consider admissible in (RF) only space-like Cauchy integral manifolds, corresponding to homotopy spheres, (*homotopy equivalence full admissibility hypothesis*), then one has the short exact sequence (74).

$$\begin{tikzcd} 0 \arrow[r] & K_{n,s}^{(RF)} \arrow[r] & \Omega_n^{(RF)} \arrow[r] & \Omega_{n,s}^{(RF)} = 0 \arrow[r] & 0 \end{tikzcd} \qquad (74)$$

We get

$$\Omega_n^{(RF)} \cong K_{n,s}^{(RF)} = \{[M] | M = \partial V,\ V = \text{singular solution of } (RF)\} \qquad (75)$$

and $M \in [S^n] \in \Omega_n^{(RF)}$ iff $M \cong S^n$. Furthermore, two n-dimensional homotopy spheres $'\Sigma^n$ and Σ^n belong to the same bordism class in $\Omega_n^{(RF)}$ iff $'\Sigma^n \cong \Sigma^n$. Therefore we get the following canonical mapping $\Gamma_n \to \Omega_n^{(RF)}$, $[\Sigma^n]_{\Gamma_n} \mapsto [\Sigma^n]_{\Omega_n^{(RF)}}$, such that $0 \{\in \Gamma_n\} \mapsto [S^n]_{\Omega_n^{(RF)}}$. (For $n \neq 4$, one can take $\Gamma_n = \Theta_n$.) This mapping is not an isomorphism. Therefore, in the full admissibility hypothesis, and in the case that $\Gamma_n = 0$, we get that the Ricci flow equation becomes a 0-crystal PDE, so all homotopy spheres are diffeomorphic to S^n. This is the case, for example, of $n = 3$, corresponding to the famous Poincaré conjecture. Finally, if we consider admissible in (*RF*) only space-like Cauchy integral manifolds, corresponding to manifolds diffeomorphic to spheres, (*sphere full admissibility hypothesis*), then

$\Omega_n^{(RF)} \cong K_{n,s}^{(RF)} \cong \Omega_{n,s}^{(RF)} = 0$ and one has $cry(RF) = 0$, i.e., (RF) becomes a 0-crystal for any dimension $n \geq 1$.[33]

Acknowledgements I would like thank Editors for their kind invitation to contribute with a my paper to this book, dedicated to Stephen Smale in occasion of his 80th birthday.
Work partially supported by MIUR Italian grants "PDE's Geometry and Applications".

References

1. R.P. Agarwal, A. Prástaro, Geometry of PDE's.III(I): webs on PDE's and integral bordism groups. The general theory. Adv. Math. Sci. Appl. **17**(1), 239–266 (2007); Geometry of PDE's.III(II): webs on PDE's and integral bordism groups. Applications to Riemannian geometry PDE's. Adv. Math. Sci. Appl. **17**(1), 267–281 (2007)
2. R.P. Agarwal, A. Prástaro, Singular PDE's geometry and boundary value problems. J. Nonlinear Conv. Anal. **9**(3), 417–460 (2008); On singular PDE's geometry and boundary value problems. Appl. Anal. **88**(8), 1115–1131 (2009)
3. R. Bott, J.W. Milnor, On the parallelizability of spheres. Bull. Am. Math. Soc. **64**, 87–89 (1958)
4. J. Cerf, *Sur les difféomorphismes de la sphére de dimension trois ($\Gamma_4 = 0$)*. Lecture notes in mathematics, vol. 53 (Springer, Berlin/New York, 1968)
5. S.K. Donaldson, Self-dual connections and the topology of smooth 4-manifolds. Bull. Am. Math. Soc. **8**, 81–83 (1983)
6. S. Ferry, A.A. Ranicki, J. Rosenberg, Novikov conjecture, rigidity and index theorems, in *Proceedings of 1993 Oberwolfach Conference*. London mathematical society lecture note, vols. 226, 227 (Cambridge University Press, Cambridge, 1995)
7. M. Freedman, The topology of four-dimensional manifolds. J. Differ. Geom. **1**(3), 357–453 (1982)
8. M. Freedman, R. Gompf, S. Morrison, K. Walker, Man and machine thinking about the smooth 4-dimensional Poincaré conjecture (2009). arXiv:09065.5177[math.GT]
9. H. Goldshmidt, Integrability criteria for systems of non-linear partial differential equations. J. Differ. Geom. **1**, 269–307 (1967)
10. M. Gromov, *Partial Differential Relations* (Springer, Berlin, 1986)
11. R.S. Hamilton, Three-manifolds with positive Ricci curvature. J. Differ. Geom. **17**, 255–306 (1982)
12. R.S. Hamilton, Four-manifolds with positive Ricci curvature operator. J. Differ. Geom. **24**, 153–179 (1986)
13. R.S. Hamilton, Eternal solutions to the Ricci flow. J. Differ. Geom. **38**, 1–11 (1993)
14. R.S. Hamilton, The formation of singularities in the Ricci flow, in *Surveys in Differential Geometry*, vol. 2 (International Press, Cambridge, MA, 1995), pp. 7–136

[33] From this theorem we get the conclusion that the Ricci flow equation for n-dimensional Riemannian manifolds, admits that starting from a n-dimensional sphere S^n, we can dynamically arrive, into a finite time, to any n-dimensional homotopy sphere M. When this is realized with a smooth solution, i.e., solution with characteristic flow without singular points, then $S^n \cong M$. The other homotopy spheres Σ^n, that are homeomorphic to S^n only, are reached by means of singular solutions. So the titles of this paper and its companion [42] are justified now ! Results of this paper agree with previous ones by Cerf [4], Freedman [7], Kervaire and Milnor [18, 20], Moise [21, 22] and Smale [45–47], and with the recent proofs of the Poincaré conjecture by Hamilton [11–15], Perelman [24, 25], and Prástaro [1, 40].

15. R.S. Hamilton, A compactness property for solutions of the Ricci flow on three-manifolds. Comm. Anal. Geom. **7**, 695–729 (1999)
16. M. Hirsch, Obstruction theories for smoothing manifolds and mappings. Bull. Am. Math. Soc. **69**, 352–356 (1963)
17. M. Hirsch, *Differential Topology* (Springer, New York, 1976)
18. M.A. Kervaire, J.W. Milnor, Groups of homotopy spheres: I. Ann. Math. **77**(3), 504–537 (1963)
19. R.C. Kirby, L.C. Siebenman, On the triangulation of manifolds and the Hauptveruntumg. Bull. Am. Math. Soc. **75**, 742–749 (1969)
20. J. Milnor, On manifolds homeomorphic to the 7-sphere. Ann. Math. **64**(2), 399–405 (1956)
21. E. Moise, Affine structures in 3-manifolds. V. The triangulation theorem and Hauptvermuntung. Ann. Math. Sec. Ser. **56**, 96–114 (1952)
22. E. Moise, *Geometric Topology in Dimension* 2 *and* 3 (Springer, Berlin, 1977)
23. J.R. Munkres, Obstructions to smoothing a piecewise differential homeomorphisms. Ann. Math. **72**, 521–554 (1960); Obstructions to imposing differentiable structures. Ill. J. Math. **8**, 361–376 (1964)
24. G. Perelman, The entropy formula for the Ricci flow and its geometry applications (2002). http://arxiv.org/abs/0211159
25. G. Perelman, Ricci flow with surgery on three-mainfolds (2003). http://arxiv.org/abs/0303109
26. A. Prástaro, Quantum geometry of PDE's. Rep. Math. Phys. **30**(3), 273–354 (1991)
27. A. Prástaro, Quantum and integral (co)bordisms in partial differential equations. Acta Appl. Math. **51**, 243–302 (1998)
28. A. Prástaro, (Co)bordism groups in PDE's. Acta Appl. Math. **59**(2), 111–202 (1999)
29. A. Prástaro, *Quantized Partial Differential Equations* (World Scientific, Singapore, 2004)
30. A. Prástaro, Geometry of PDE's. I: integral bordism groups in PDE's. J. Math. Anal. Appl. **319**, 547–566 (2006)
31. A. Prástaro, Geometry of PDE's. II: variational PDE's and integral bordism groups. J. Math. Anal. Appl. **321**, 930–948 (2006)
32. A. Prástaro, (Un)stability and bordism groups in PDE's. Banach J. Math. Anal. **1**(1), 139–147 (2007)
33. A. Prástaro, Geometry of PDE's. IV: Navier-Stokes equation and integral bordism groups. J. Math. Anal. Appl. **338**(2), 1140–1151 (2008)
34. A. Prástaro, On the extended crystal PDE's stability. I: the n-d'Alembert extended crystal PDE's. Appl. Math. Comput. **204**(1), 63–69 (2008)
35. A. Prástaro, On the extended crystal PDE's stability. II: entropy-regular-solutions in MHD-PDE's. Appl. Math. Comput. **204**(1), 82–89 (2008)
36. A. Prástaro, Extended crystal PDE's stability. I: the general theory. Math. Comput. Model. **49**(9–10), 1759–1780 (2009)
37. A. Prástaro, Extended crystal PDE's stability. II: the extended crystal MHD-PDE's. Math. Comput. Model. **49**(9–10), 1781–1801 (2009)
38. A. Prástaro, Surgery and bordism groups in quantum partial differential equations. I: the quantum Poincaré conjecture. Nonlinear Anal. Theory Method Appl. **71**(12), 502–525 (2009)
39. A. Prástaro, Surgery and bordism groups in quantum partial differential equations. II: variational quantum PDE's. Nonlinear Anal. Theory Method Appl. **71**(12), 526–549 (2009)
40. A. Prástaro, Extended crystal PDEs (2008). http://arxiv.org/abs/0811.3693
41. A. Prástaro, Quantum extended crystal super PDEs (2009). http://arxiv.org/abs/0906.1363
42. A. Prástaro, Exotic heat PDE's. Commun. Math. Anal. **10**(1), 64–81 (2011). http://arxiv.org/abs/1006.4483
43. A. Prástaro, Exotic heat PDEs (2010). II http://arxiv.org/abs/1009.1176
44. A. Prástaro, Th. M. Rassias, Ulam stability in geometry of PDE's. Nonlinear Funct. Anal. Appl. **8**(2), 259–278 (2003)
45. S. Smale, Generalized Poincaré conjecture in dimension greater than four. Ann. Math. **74**(2), 391–406 (1961)
46. S. Smale, On the structure of manifolds. Am. J. Math. **84**, 387–399 (1962)
47. S. Smale, Differentiable dynamical systems. Bull. Am. Math. Soc. **73**, 747–817 (1967)

48. C.T.C. Wall, Determination of the cobordism ring. Ann. Math. **72**, 292–311 (1960)
49. C.T.C. Wall, *Surgery on Compact Manifolds*. London Mathematical Society Monographs, vol. 1 (Academic Press, New York, 1970); 2nd edn, ed. by A.A. Ranicki, American mathematical surveys and monographs, vol. 69 (American Mathematical Society, Providence, 1999)
50. J.H.C. Whitehead, Manifolds with transverse fields in Euclidean spaces. Ann. Math. **73**, 154–212 (1961)

Topology at a Scale in Metric Spaces

Nat Smale

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract This is an expository article that discusses some developments in joint work with Laurent Bartholdi, Thomas Schick and Steve Smale in [1] and also in [10]. Recently, in various contexts, there has been interest in the topology of certain spaces (even finite data sets) at a "scale", for example, in reconstruction of manifolds or other spaces from a discrete sample as in [8] and [4], and also in connection with learning theory [9, 11] and [7]. In persistence homology, [3, 5] mathematicians have been computing topological features at a range of scales, to find the fundamental structures of spaces and data sets. See also [2]. In this paper, we will first give an explicit description of homology at a scale, for a compact metric space. We will then describe a Hodge theory for the corresponding cohomology when the space has a Borel probability measure.

1 The Čech Complex of a Metric Space at a Fixed Scale

Let (X, d) be a compact metric space. The closed ball of radius $r > 0$ centered at p in X is denoted by $B_r(p) = \{q \in X : d(p,q) \leq r\}$. Now, let $\alpha > 0$ be any positive number. This will be what we call the scale, and the goal is to in some sense understand the homology of X at the scale α. We will want to consider two points to be in some sense connected, if they are close relative to α. For $k \geq 0$ an integer, we will denote by X^{k+1}, the $k+1$-fold Cartesian product of X, that is all points of

N. Smale (✉)
Department of Mathematics, University of Utah, 155 S. 1400 E. Rm 223 Salt Lake City,
Utah 84112-0090, USA
e-mail: smale@math.utah.edu

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_16, © Springer-Verlag Berlin Heidelberg 2012

the form $(x_0, \dots x_k)$ where $x_i \in X$ for each i. We will denote the diagonal of X^{k+1} by Δ^{k+1}. Thus Δ^{k+1} will denote all points in X^{k+1} of the form $(p, \dots, p)$ where $p \in X$. We will make X^{k+1} into a metric space by defining the distance between two points $x = (x_0, \dots, x_k)$ and $y = (y_0, \dots, y_k)$ in X^ℓ by

$$d_{k+1}(x, y) = \max\{d(x_i, y_i) : i = 0, \dots, k\}$$

The distance from a point $x \in X^{k+1}$ to Δ^{k+1} is just

$$\operatorname{dist}(x, \Delta) = \min\{d_{k+1}(x, y) : y \in \Delta^{k+1}\}$$

Thus, for example if $x = (x_0, \dots, x_k) \in X^{k+1}$ satisfies $\operatorname{dist}(x, \Delta^{k+1}) \leq \alpha$, then there exists (by compactness of X^{k+1}) $q = (t, \dots, t) \in \Delta^{k+1}$ with $d_{k+1}(x, q) \leq \alpha$. This is equivalent to saying there is a $t \in X$ such that $d(x_i, t) \leq \alpha$ for $i = 0, \dots, k$. Another way of stating this is that $\operatorname{dist}(x, \Delta) \leq \alpha$ if and only if $\cap_{i=0}^k B_\alpha(x_i)$ is nonempty. Of fundamental importance in scaled homology is the closed α neighborhood of the diagonal

$$U_\alpha^{k+1} = \{x \in X^{k+1} : \operatorname{dist}(x, \Delta^{k+1}) \leq \alpha\}$$

Thus, for a point $p = (p_0, \dots, p_k) \in U_\alpha^{k+1}$, there exists $t \in \cap_{i=0}^k B_\alpha(p_i)$. Any such t is sometimes called a “witness” for p or a witness for $p_0, \dots, p_k$. For $k = 0$, $U_\alpha^{k+1} = X$.

We will consider points in U_α^{k+1} as k-dimensional simplices. Thus a point $(x_0, x_1) \in U_\alpha^2$ will be thought of as an edge (a 1-dimensional simplex) between x_0 and x_1, and a point $(x_0, x_1, x_2) \in U_\alpha^3$ will be thought of as a triangle (a 2-dimensional simplex) with vertices x_0, x_1 and x_2. Of course, if $(x_0, \dots, x_k) \in U_\alpha^{k+1}$, and τ is any permutation of of the numbers $0, 1, \dots, k$, then, intuitively, $(x_{\tau(0)}, \dots, x_{\tau(k)})$ should be the same simplex, as it has the same vertices. However, it will be important to impose an orientation on our simplices. We will thus think of a point (x_0, x_1) in U_α^2 as an edge with a direction from x_0 towards x_1. In general k-simplices will have two possible orientations. Formally, we will put an equivalence relation $\sim$ on U_α^{k+1} as follows. Two points $x = (x_0, \dots, x_k)$ and $y = (y_0, \dots, y_k)$ in U_α^{k+1} are equivalent, $x \sim y$, if there is an even permutation (the composition of an even number of transpositions) τ on the numbers $0, \dots, k$ such that $(y_0, \dots, y_k) = (x_{\tau(0)}, \dots, x_{\tau(k)})$. Thus (x_0, x_1) and (x_1, x_0) will be in different equivalence classes. The points (x_0, x_1, x_2) and (x_2, x_0, x_1) are in the same equivalence class in U_α^3 but (x_0, x_1, x_2) and (x_2, x_1, x_0) are in different equivalence classes. We will denote the set of equivalence classes in U_α^{k+1} by $\hat{U}_\alpha^{k+1}$ and the equivalence class of a point $(x_0, \dots, x_k)$ will be denoted by $[x_0, \dots, x_k]$, and these will be referred to as simplices (or more precisely as k-dimensional simplices at scale α). If $[x_0, \dots, x_k] \in \hat{U}_\alpha^{k+1}$ such that $x_i = x_j$ for some $i \neq j$, then $[x_0, \dots, x_k]$ will be called a degenerate simplex. Of course, elements in $\hat{U}_\alpha^1$ are just vertices (points in X). Note that if $[x_0, \dots, x_k] \in \hat{U}_\alpha^{k+1}$, then if we delete $k - j$ $(j < k)$ vertices to get the point $(x_{i_0}, \dots, \dots, x_{i_j})$ then $[x_{i_0}, \dots, \dots, x_{i_j}]$ is in U_α^{j+1} because if t is a witness for $(x_0, \dots, x_k)$, it is also a witness for $(x_{i_0}, \dots, \dots, \dots, x_{i_j})$. Thus, the collection $\hat{U}_\alpha^1, \hat{U}_\alpha^2, \dots$ gives rise to an abstract simplicial complex on X.

We will now construct a vector space out of the simplices $\hat{U}_\alpha^{k+1}$. For $k \geq 0$, the set of k chains, denoted by $C_{k,\alpha}(X)$ is the set of all formal linear combinations over $\mathbf{R}$ of elements in $\hat{U}_\alpha^{k+1}$. For a simplex $\sigma = [x_0, \dots, x_k]$ we define it's additive inverse to be $-\sigma = [x_{\tau(0)}, \dots, x_{\tau(k)}]$ where τ is any odd permutation of $0, \dots, k$. Thus for example, $[x_0, x_1] = -[x_1, x_0]$. With this rule, it is easy to check that $C_{k,\alpha}(X)$ can be made into a vector (linear) space, with addition and scalar multiplication carried out in the obvious way. Of course the requirement that $[x_0, \dots, x_k] = -[x_{\tau(0)}, \dots, x_{\tau(k)}]$ forces degenerate simplices to be the zero element of $C_{k,\alpha}(X)$.

Remark 1. Instead of defining a chain to be a linear combination of simplices over $\mathbf{R}$, one sometimes defines a chain to simply be a sum of simplices. This makes $C_{k,\alpha}(X)$ into an abelian group. That is, elements in $C_{k,\alpha}(X)$ are then linear combinations of elements in $\hat{U}_\alpha^{k+1}$ with coefficients that are in $\mathbf{Z}$ (the integers). This makes certain aspects of the theory a little simpler, and others a little more difficult (the structure of abelian groups is in general more complicated than vector spaces).

Finally, the Čech complex of X at scale α, denoted by $C_\alpha(X)$, is just the union $\cup_{k \geq 0} C_{\ell,\alpha}(X)$. This is a simplicial complex with the usual simplicial boundary operator defined as follows. For $[x_0, \dots, x_k] \in \hat{U}_\alpha^{k+1}$ we define

$$\partial_k [x_0, \dots, x_k] = \sum_{i=0}^{k} (-1)^i [x_0, \dots, \hat{x}_i, \dots, x_k]$$

where $\hat{x}_i$ indicates that x_i is deleted. We will usually omit the k in the notation. Thus for $\sigma \in \hat{U}_\alpha^{k+1}$, $\partial\sigma \in C_{k-1,\alpha}(X)$. One can easily check that for $\sigma \in \hat{U}_\alpha^{k+1}$, $\partial(-\sigma) = -\partial(\sigma)$, and so ∂ extends to a linear map

$$\partial : C_{k,\alpha}(X) \to C_{k-1,\alpha}(X)$$

An easy computation shows that $\partial_{k-1} \circ \partial_k = 0$ or more simply $\partial^2 = 0$ and we have the chain complex

$$\cdots \xrightarrow{\partial_{k+1}} C_{k,\alpha}(X) \xrightarrow{\partial_k} C_{k-1,\alpha}(X) \xrightarrow{\partial_{k-1}} \cdots C_{0,\alpha}(X) \xrightarrow{\partial_0} 0 \tag{1}$$

We will denote by $B_{k,\alpha}$ the image of ∂_{k+1} (the k-boundaries), and $Z_{k,\alpha}$ the kernel of ∂_k (the k-cycles), both of which are linear subspaces of $C_{k,\alpha}(X)$. Since $\partial^2 = 0$, $B_{k,\alpha} \subset Z_{k,\alpha}$, and the quotient space

$$H_{k,\alpha}(X) = \frac{Z_{k\alpha}}{B_{k,\alpha}}$$

is called the homology of X at scale α and degree k. Thus, two cycles are equivalent (represent the same homology class) if their difference is a boundary (an element of $B_{k,\alpha}$). The zero element of $H_{k,\alpha}(X)$ is represented by any element of $B_{k,\alpha}$. A non-trivial cycle thus represents a "topological feature" of X that we can observe at scale α. It is shown in [1] that $H_{1,\alpha}(X)$ is always finite dimensional, though examples of X are given (due to Anthony Baker) where $H_{2,\alpha}(X)$ is infinite dimensional for specific scales.

Example 1. Let X be the set of 12 points on the unit circle $\pi/6$ radians apart $\{e^{\frac{n\pi i}{6}} : n = 0, 1, \dots, 11\}$ equipped with metric induced as a subspace of $\mathbf{R}^2$. We will investigate $H_{1,\alpha}(X)$. To simplify notation, let $e_n = e^{\frac{n\pi i}{6}}$. The distance between adjacent points e_n, e_{n+1} is $\sqrt{2-\sqrt{3}}$, and if $\alpha < \sqrt{2-\sqrt{3}}$, then $H_{1,\alpha}(X) = 0$ since there are no non-trivial 1-chains. No two distinct e_n, e_m have a common witness. For $\alpha = \sqrt{2-\sqrt{3}}$, the following 1-chain

$$\sigma = \sum_{n=0}^{10}[e_n, e_{n+1}] + [e_{11}, e_0]$$

is easily verified to be a cycle, and in fact one can show that it is non-trivial. Each $[e_n, e_{n+1}]$ has e_n and e_{n+1} as witnesses. More simply, the following chain

$$\gamma = \sum_{n=0}^{8}[e_n, e_{n+2}] + [e_{10}, e_0]$$

is also a non-trivial cycle, with e_{n+1} serving as a witness for $[e_n, e_{n+2}]$. If we define

$$\Gamma = \sum_{n=0}^{8}[e_n, e_{n+1}, e_{n+2}] + [e_{10}, e_{11}, e_0]$$

then $\Gamma \in C_{2,\alpha}(X)$, since e_{n+1} is a witness for $[e_n, e_{n+1}, e_{n+2}]$, and one can easily check that $\partial\Gamma = \sigma - \gamma$. Thus σ and γ represent the same homology class at scale α. For α large enough, these cycles become homologically trivial. In fact if $\alpha \geq \sqrt{3}$ (the distance between e_n and e_{n+4}) one can easily check that

$$\gamma = \partial([e_0, e_2, e_4] + [e_4, e_6, e_8] + [e_8, e_{10}, e_0] + [e_0, e_4, e_8])$$

For any compact metric space, $H_{k,\alpha}(X) = 0$ for $k > 0$, when α is big enough, for example if X is contained in a ball of radius α.

Note that if X is any finite metric space as in a data set, all of the spaces $C_{k,\alpha}$, $Z_{k,\alpha}$ and $B_{k,\alpha}$ are finite dimensional and the homology at scale α can be computed with linear algebra.

2 Cohomology and Hodge Theory

It is often useful to consider the cohomology of a space instead of homology, as there are additional structures that can give more insight. We first consider the simplicial cohomology of the Čech complex above. Thus (X, d) is a compact metric space and $\alpha > 0$. Consider the co-chain complex corresponding to (1)

$$0 \to C_\alpha^0(X) \xrightarrow{\delta_0} C_\alpha^1(X) \xrightarrow{\delta_1} \cdots \xrightarrow{\delta_{k-1}} C_\alpha^k(X) \xrightarrow{\delta_\ell} \cdots \tag{2}$$

That is, $C_\alpha^k(X)$ is the dual space to the k-chains, or the space of linear functions from $C_{k,\alpha}(X)$ to $\mathbf{R}$, and $\delta : C^{k,\alpha}(X) \to C^{k+1,\alpha}(X)$ is the co-boundary operator or the adjoint of ∂. Thus for $f \in C_\alpha^k(X)$, and $\sigma \in C_{k+1,\alpha}(X)$,

$$\delta f(\sigma) = f(\partial\sigma)$$

By functoriality, $\delta_{k+1} \circ \delta_k = 0$. Now, linear functions on $C_{k,\alpha}(X)$ are uniquely determined by their values on a basis, that is elements of $\hat{U}_\alpha^{k+1}$. Thus $C_\alpha^k(X)$ is equivalent to the vector space of all functions on $\hat{U}_\alpha^{k+1}$. On the other hand, a function f on $\hat{U}_\alpha^{k+1}$ which is linear on $C_{k,\alpha}(X)$ must satisfy $f(-\sigma) = -f(\sigma)$ for $\sigma \in \hat{U}_\alpha^{k+1}$, and therefore defines an alternating function on U_α^{k+1}. Thus $f(x_0, \ldots, x_k) = (-1)^\tau f(x_{\tau(0)}, \ldots, x_{\tau(k)})$ for $(x_0, \ldots, x_k) \in U_\alpha^{k+1}$, and for permutations τ, of k+1 elements, where $(-1)^\tau$ is the sign of the permutation. Thus we have identified $C_\alpha^k(X)$ with $F_a(U_\alpha^{k+1})$, the set of all alternating functions on U_α^{k+1}. We will rewrite the co-chain complex (1) as

$$0 \to F(X) \xrightarrow{\delta_0} F_a(U_\alpha^2) \xrightarrow{\delta_1} \cdots \xrightarrow{\delta_{k-1}} F_a(U_\alpha^{k+1}) \xrightarrow{\delta_\ell} \cdots \tag{3}$$

It is standard that the co-boundary operator $\delta : F_a(U_\alpha^{k+1}) \to F_a(U_\alpha^{k+2})$ is given by

$$\delta f(x_0, \ldots, x_{k+1}) = \sum_{i=0}^{k+1} (-1)^{i+1} f(x_0, \ldots, \hat{x}_i, x_{i+1}, \ldots, x_{k+1})$$

We define Z_α^k to be the kernel of δ_k (co-cycles at scale α and degree k) and B_α^k to be the image of δ_{k-1} (the co-boundaries at scale α). Since $\delta^2 = 0$, $B_\alpha^k \subset Z_\alpha^k$ and the quotient space

$$H_\alpha^k(X) = \frac{Z_\alpha^k}{B_\alpha^k}$$

is called the cohomology of X at scale α and degree k. Since the complexes (1) and (2) are over a field, $\mathbf{R}$, the Universal Coefficients theorem of algebraic topology implies that $H_{k,\alpha}(X)$ and H_α^k are isomorphic, and so H_α^k also describes in some sense the topology of X at scale α.

It turns out that one can impose further structures on the space of functions on U_α^{k+1} such as continuity, or square integrability with respect to a measure, and get co-chain complexes analogous to (3). Let μ be a Borel probability measure on X, and μ_{k+1} the corresponding product measure on X^{k+1}, $k \geq 0$. This induces a measure (which we will still call μ_{k+1}) on U_α^{k+1}. We will denote by $L_a^2(U_\alpha^{k+1})$ the space of alternating, real valued functions on U_α^{k+1} which are in L^2 with respect to μ_{k+1}. The following proposition was proved in [1].

Proposition 1. *The co-boundary operator $\delta : L_a^2(U_\alpha^{k+1}) \to L_a^2(U_\alpha^{k+2})$ defines a bounded linear map between Hilbert spaces, for all $k \geq 0$.*

To motivate our Hodge theory, let us recall the classical Hodge and de Rham theorems. Let X be a smooth, compact manifold of dimension n, and let, for $k \geq 0$, $\Omega^k(X)$ denote the space of smooth k-forms on X. The exterior derivative $d : \Omega^k(X) \to \Omega^{k+1}(X)$ is a first order, linear differential operator extending the notion of the differential of a function (a 0-form), and satisfies $d \circ d = 0$. The de Rham complex is the co-chain complex

$$0 \to \Omega^0(X) \xrightarrow{d} \Omega^1(X) \xrightarrow{d} \cdots \xrightarrow{d} \Omega^n(X) \xrightarrow{d} 0$$

The de Rham cohomology of degree k is the quotient space

$$H_{dR}^k(X) = \frac{\ker d : \Omega^k \to \Omega^{k+1}}{\operatorname{Im} d : \Omega^{k-1} \to \Omega^k}$$

and the de Rham Theorem states that $H_{dR}^k(X)$ is isomorphic to to the usual cohomology groups (singular, simplicial, or Čech) of X and is thus a topological invariant of X. Now assume that X has a Riemannian metric g. This induces an inner product on alternating k-tensors on the tangent space at each point, and thus on $\Omega^k(X)$ by integration of this with respect to the volume form. Then $d : \Omega^k(X) \to \Omega^{k+1}(X)$ has a formal adjoint $d^* : \Omega^{k+1}(X) \to \Omega^k(X)$, and the Hodge Laplacian is the second order, self adjoint, elliptic operator

$$\Delta_k = d^*d + dd^* : \Omega^k(X) \to \Omega^k(X)$$

When $k = 0$, this is the classical Laplacian on functions. The classical Hodge theorem [6] is a beautiful synthesis of analysis, topology and geometry.

Theorem 1 (Hodge theorem). *For $k = 0, \ldots, n$, we have the orthogonal, direct sum decomposition*

$$\Omega^k(X) = \mathit{Image}\,(d) \oplus \mathit{Image}\,(d^*) \oplus \mathit{Kernel}\,\Delta_k$$

and Kernel Δ_k is isomorphic to $H_{dR}^k(X)$.

For a compact metric space, and scale $\alpha > 0$ we view the co-chain complex

$$0 \to L^2(X) \xrightarrow{\delta_0} L^2_a(U^2_\alpha) \xrightarrow{\delta_1} \cdots \xrightarrow{\delta_{k-1}} L^2_a(U^{k+1}_\alpha) \xrightarrow{\delta_k} \cdots \tag{4}$$

as analogous to the de Rham complex. The cohomology of this complex is the quotient space

$$H^k_\alpha(X) = \frac{\ker \delta : L^2_a(U^{k+1}_\alpha) \to L^2_a(U^{k+2}_\alpha)}{\operatorname{Im} \delta : \delta : L^2_a(U^k_\alpha) \to L^2_a(U^{k+1}_\alpha)}$$

and we view this space as another form of topology at scale α. Now, from the above proposition, the cobounday operator has a bounded adjoint

$$\delta^* : L^2_a(U^{k+2}_\alpha) \to L^2_a(U^{k+1}_\alpha)$$

One can compute the explicit formula for δ^*

$$\delta^* f(x_0, \ldots, x_k) = (k+2) \int_{S_{x_0,\ldots,x_k}} f(t, x_0, \ldots, x_k)\, dt$$

where

$$S_{x_0,\ldots,x_k} = \{t \in X : (t, x_0, \ldots, x_k) \in U^{k+2}_\alpha\}$$

is the slice of of $x = (x_0, \ldots, x_k)$. The corresponding Hodge operator at scale α

$$\Delta_{k,\alpha} = \delta\delta^* + \delta^*\delta : L^2(U^{k+1}_\alpha) \to L^2(U^{k+1}_\alpha)$$

is a bounded, self adjoint positive operator of Hilbert spaces. A natural question in this context is

Hodge Question at Scale $\boldsymbol{\alpha}$. Under what conditions on X, d, μ, α do we have

$$L^2(U^{k+1}_\alpha) = \operatorname{Im} \delta \oplus \operatorname{Im} \delta^* \oplus \operatorname{Ker} \Delta_{k,\alpha}$$

and Ker $\Delta_{k,\alpha}$ is isomorphic to $H^k_{L^2,\alpha}$.

We call elements of Ker $\Delta_{k,\alpha}$ α-harmonic functions. Note that

$$< \Delta_{k,\alpha} f, f >=< \delta^*\delta f + \delta\delta^* f >= \|\delta f\|^2 + \|\delta^* f\|^2$$

and so $\Delta_{k,\alpha} f = 0$ if and only if $\delta f = 0$ and $\delta^* f = 0$.

The answer to the Hodge question is affirmative, precisely when δ has closed range in $L^2_\alpha(U^{k+1}_\alpha)$ (see the Hodge Lemma in [1]). In particular, this will hold when the image has finite codimension, that is $\dim H^k_{L^2,\alpha}(X) < \infty$. In [1] some sufficient conditions are given on α and the metric d. Roughly, the witness set $w_\alpha(x_0, \ldots, x_k) = \cap_{i=0}^k B_\alpha(x_i)$

$$w_\alpha : U^{k+1}_\alpha \to K(X)$$

must be continuous (here $K(X)$ is the metric space of compact subsets of X with the Hausdorff metric), and the radius of intersections of α balls must be controlled. As a special case, it is shown to be consistent with the classical Hodge-de Rham theory for Riemannian manifolds at small scales.

Theorem 2. *Let (X, g) be a compact Riemannian manifold. Then for $\alpha > 0$ sufficiently small, the answer to the Hodge question above is affirmative, and furthermore Ker $\Delta_{k,\alpha}$ is isomorphic to $H^k_{L^2,\alpha}$ as well as H^k_α and $H^k_{dR}(X)$.*

The proof is a bit lengthy and is carried out using a bi-complex argument.

2.1 An Explicit Isomorphism

The proof of the theorem in [1] does not give an explicit isomorphism between $H^k_{dR}(X)$ and $H^k_{L^2,\alpha}$. In [10], we construct a co-chain map (in the case of a Riemannian manifold and small α) between the de Rham complex and the L^2, α complex (4), which induces isomorphisms on cohomology. Let M be a compact Riemannian manifold, and let $\alpha > 0$ be small enough so that closed balls of radius 2α are strictly convex. We construct a co-chain map, that is for each k, a linear map

$$\Psi : \Omega^k(M) \to L^2_a(U^{k+1}_\alpha)$$

such that

$$\Psi \circ d = \delta \circ \Psi$$

For $(x_0, \ldots, x_k) \in U^{k+1}_\alpha$ we define a smooth k-simplex $S(x_0, \ldots, x_k)$ in M, inductively on k. $S(x_0, x_1)$ is just the minimizing geodesic from x_0 to x_1. $S(x_0, x_1, x_2)$ is the union of geodesics from x_2 to points on $S(x_0, x_1)$, and so on. We then define

$$\Psi_0 : \Omega^k(M) \to L^2(U^{k+1}_\alpha)$$

by

$$(\Psi_0\omega)(x_0, \ldots, x_k) = \int_{S(x_0,\ldots,x_k)} \omega$$

In general, $\Psi_0\omega$ will not be alternating, unless $k = 0, 1$, or M has constant curvature. We therefore alternate $\Psi_0\omega$ and define

$$\Psi\omega(x_0, \ldots, x_k) = \mathrm{Alt}\,(\Psi_0\omega)(x_0, \ldots, x_k)$$

Theorem 3. *Ψ is a co-chain map of co-chain complexes, and induces an isomorphism on cohomology.*

The proof that Ψ is a co-chain map is essentially Stoke's theorem. The proof that Ψ is an isomorphism on cohomology follows by constructing a left inverse for Ψ, and using the fact from [1] that the cohomology groups have the same dimension. To describe the left inverse, which we will call Φ, note that $\Psi\omega$ is actually a smooth alternating function. That is, Ψ is really a co-chain map into the sub-complex

$$0 \to C^\infty(X) \xrightarrow{\delta} C_a^\infty(U_\alpha^2) \xrightarrow{\delta} \cdots \xrightarrow{\delta} C_a^\infty(U_\alpha^k) \xrightarrow{\delta} \cdots$$

Where $C_a^\infty(U_\alpha^k)$ is the space of smooth, alternating functions on U_α^k. In [1] it was shown that the inclusion map from this complex to the L^2 complex induces an isomorphism on cohomology, thus it suffices to define the left inverse Φ on smooth alternating functions. In fact, if we define $\Phi : C_a^\infty(U_\alpha^{k+1}) \to \Omega^k(M)$ by

$$(\Phi f)(p)(v_1, \ldots, v_k) = D_1 D_2 \cdots D_k f(p, t_1, \ldots, t_k)(v_1, \ldots, v_k)$$

for $p \in M$ and $v_1, \ldots, v_k \in T_p M$ (D_i is the derivative of f taken at $t_i = p$), then it can be shown that $\Phi \circ \Psi = \mathrm{Id}$.

2.2 Further Results

It is shown that in general, for a harmonic 1 form ω, $\Psi\omega$ is harmonic in the abstract sense, and so Ψ is an isomorhism between classical harmonic 1 forms, and α-harmonic functions on U_α^2. For the flat n-dimensional torus, Ψ takes harmonic k-forms to harmonic functions on U_α^{k+1} for all k. We conjecture that this might be true for constant curvature in general.

For $k = 0$, we can compare the classical and abstract Hodge Laplacian since both act on functions on M. When appropriately scaled, the α-Laplacian is close to the classical Laplacian.

Theorem 4. *There is a universal constant c_n such that for a C^3 function f on M, we have*

$$\|\Delta f - c_n \alpha^{-n-2} \Delta_\alpha f\|_\infty \leq C \|f\|_{C^3} \alpha$$

References

1. L. Bartholdi, T. Schick, N. Smale, S. Smale, Hodge theory on metric spaces. Found. Comput. Math. 1–48 (2012). Springer
2. G. Carlsson, Topology and data. Bull. Am. Math. Soc. (N.S.) **46**(2), 255–308 (2009)
3. G. Carlsson, A. Zomorodian, Computing persistent homology. Discret. Comput. Geom. **33**(2), 249–274 (2005)
4. F. Chazal, S. Oudot, *Towards Persistence Based Reconstruction in Euclidean Spaces*. (ACM, New York, 2007)
5. H. Edelsbrunner, D. Letscher, A. Zomorodian, Topological persistence and simplification. Discret. Comput. Geom. **28**(4), 511–533 (2002)

6. W. V. D. Hodge *The Theory and Applications of Harmonic Integrals* (Cambridge University Press, Cambridge/England, 1941)
7. X. Jiang, L-H. Lim, Y. Yao, Y. Ye, Statistical ranking and combinatorial Hodge theory. arxiv.org/abs/0811.1067
8. P. Niyogi, S. Smale, S. Weinberger, Finding the homology of submanifolds with high confidence. Discret. Comput. Geom. **39**, 419–441 (2008)
9. P. Niyogi, S. Smale, S. Weinberger, A topological view of unsupervised learning fromnoisy data, University of Chicago Technical Report, 2008
10. N. Smale, S. Smale, Abstract And Classical Hodge-de Rham Theory. Anal. Appl. 91–111 (2012). World Scientific Publishing Company
11. S. Smale, D. X. Zhou, Geometry on probability spaces. Constr. Approx. 311–323 (2009). Springer

Riemann, Hurwitz and Hurwitz-Lerch Zeta Functions and Associated Series and Integrals

H.M. Srivastava

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract The main object of this article is to present a survey-cum-expository account of some recent developments involving the Riemann Zeta function $\zeta(s)$, the Hurwitz (or generalized) Zeta function $\zeta(s,a)$, and the Hurwitz-Lerch Zeta function $\Phi(z,s,a)$ as well as its various interesting extensions and generalizations. We first investigate the problems associated with the evaluations and representations of $\zeta(s)$ when $s \in \mathbb{N} \setminus \{1\}$, $\mathbb{N}$ being the set of natural numbers, emphasizing upon several interesting classes of rapidly convergent series representations for $\zeta(2n+1)$ $(n \in \mathbb{N})$ which have been developed in recent years. In two of many computationally useful special cases considered here, it is observed that $\zeta(3)$ can be represented by means of series which converge much more rapidly than that in Euler's celebrated formula as well as the series which was used more recently by Roger Apéry (1916–1994) in his proof of the irrationality of $\zeta(3)$. Symbolic and numerical computations using *Mathematica* (Version 4.0) for Linux show, among other things, that only 50 terms of one of these series are capable of producing an accuracy of seven decimal places. We also consider a variety of series and integrals associated with the Hurwitz-Lerch Zeta function $\Phi(z,s,a)$ as well as its various interesting extensions and generalizations.

1 Introduction, Definitions and Preliminaries

Throughout this article, we use the following standard notations:

H.M. Srivastava (✉)
Department of Mathematics and Statistics, University of Victoria, Victoria,
British Columbia V8W 3R4, Canada
e-mail: harimsri@math.uvic.ca

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_17, © Springer-Verlag Berlin Heidelberg 2012

$$\mathbb{N} := \{1, 2, 3, \cdots\}, \quad \mathbb{N}_0 := \{0, 1, 2, 3, \cdots\} = \mathbb{N} \cup \{0\}$$

and

$$\mathbb{Z}^- := \{-1, -2, -3, \cdots\} = \mathbb{Z}_0^- \setminus \{0\}.$$

Also, as usual, $\mathbb{Z}$ denotes the set of integers, $\mathbb{R}$ denotes the set of real numbers and $\mathbb{C}$ denotes the set of complex numbers.

Some rather important and potentially useful functions in *Analytic Number Theory* include (for example) the Riemann Zeta function $\zeta(s)$ and the Hurwitz (or generalized) Zeta function $\zeta(s, a)$, which are defined (for $\Re(s) > 1$) by

$$\zeta(s) := \begin{cases} \displaystyle\sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{1-2^{-s}} \sum_{n=1}^{\infty} \frac{1}{(2n-1)^s} & \big(\Re(s) > 1\big) \\ \displaystyle\frac{1}{1-2^{1-s}} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^s} & \big(\Re(s) > 0;\ s \neq 1\big) \end{cases} \tag{1}$$

and

$$\zeta(s, a) := \sum_{n=0}^{\infty} \frac{1}{(n+a)^s} \qquad \big(\Re(s) > 1;\ a \in \mathbb{C} \setminus \mathbb{Z}_0^-\big), \tag{2}$$

and (for $\Re(s) \leqq 1$; $s \neq 1$) by their *meromorphic* continuations (see, for details, the excellent work by Titchmarsh [54] and the monumental treatise by Whittaker and Watson [57]; see also [1, Chap. 23] and [43, Chap. 2]), so that (obviously)

$$\zeta(s, 1) = \zeta(s) = (2^s - 1)^{-1}\ \zeta\left(s, \frac{1}{2}\right) \quad \text{and} \quad \zeta(s, 2) = \zeta(s) - 1. \tag{3}$$

More generally, we have the following relationships:

$$\zeta(s) = \frac{1}{m^s - 1} \sum_{j=1}^{m-1} \zeta\left(s, \frac{j}{m}\right) \tag{4}$$

$$(m \in \mathbb{N} \setminus \{1\};\ \mathbb{N} := \{1, 2, 3, \ldots\})$$

and

$$\zeta(s, ma) = \frac{1}{m^s} \sum_{j=0}^{m-1} \zeta\left(s, a + \frac{j}{m}\right) \qquad (m \in \mathbb{N}). \tag{5}$$

A fascinatingly and tantalizingly large number of seemingly independent solutions of the so-called *Basler problem* of evaluating the Riemann Zeta function $\zeta(s)$ when $s = 2$, which was of vital importance to Leonhard Euler (1707–1783) and the Bernoulli brothers [Jakob Bernoulli (1654–1705) and Johann Bernoulli (1667–1748)], have appeared in the mathematical literature ever since Euler *first* solved this problem in the year 1736. In this context, one other remarkable *classical* result involving Riemann's ζ-function $\zeta(s)$ is the following elegant series representation for $\zeta(3)$:

$$\zeta(3) = -\frac{4\pi^2}{7} \sum_{k=0}^{\infty} \frac{\zeta(2k)}{(2k+1)(2k+2)\,2^{2k}}, \tag{6}$$

which was actually contained in Euler's 1772 paper entitled "*Exercitationes Analyticae*" (cf., e.g., Ayoub [6, pp. 1084–1085]). In fact, This 1772 result of Euler was rediscovered (among others) by Ramaswami [34] (see also a paper by Srivastava [37, p. 7, Eq. 2.23]) and (more recently) by Ewell [12]. Moreover, just as pointed out by (for example) Chen and Srivastava [4, pp. 180–181], another series representation:

$$\zeta(3) = \frac{5}{2} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^3 \binom{2k}{k}}, \tag{7}$$

which played a key rôle in the *celebrated* proof (see, for details, [3]) of the *irrationality* of $\zeta(3)$ by Roger Apéry (1916–1994), was derived *independently* by (among others) Hjortnaes [21], Gosper [17], and Apéry [3].

It is easily observed that Euler's series in (6) converges faster than the defining series for $\zeta(3)$, but obviously not as fast as the series in (7). Evaluations of such Zeta values as $\zeta(3)$, $\zeta(5)$, et cetera are known to arise naturally in a wide variety of applications such as those in Elastostatics, Quantum Field Theory, et cetera (see, for example, Tricomi [55], Witten [58], and Nash and O'Connor [29, 30]). On the other hand, in the case of *even* integer arguments, we already have the following computationally useful relationship:

$$\zeta(2n) = (-1)^{n-1} \frac{(2\pi)^{2n}}{2 \cdot (2n)!} B_{2n} \qquad (n \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}) \tag{8}$$

with the *well-tabulated* Bernoulli numbers defined by the following generating function:

$$\frac{z}{e^z - 1} = \sum_{n=0}^{\infty} B_n \frac{z^n}{n!} \qquad (|z| < 2\pi), \tag{9}$$

as well as the familiar recursion formula:

$$\zeta(2n) = \left(n + \frac{1}{2}\right)^{-1} \sum_{k=1}^{n-1} \zeta(2k)\, \zeta(2n-2k) \qquad (n \in \mathbb{N} \setminus \{1\}). \tag{10}$$

Our presentation in this article consistes of two major parts. First of all, motivated essentially by a genuine need (for computational purposes) for expressing $\zeta(2n+1)$ as a rapidly converging series for all $n \in \mathbb{N}$, we propose to present a rather systematic investigation of the various interesting families of rapidly convergent series representations for the Riemann $\zeta\,(2n+1)$ $(n \in \mathbb{N})$. Relevant connections of the results presented here with many other known series representations for $\zeta\,(2n+1)$ $(n \in \mathbb{N})$ are also briefly indicated. In fact, for two of the many computationally useful special cases considered here, we observe that $\zeta\,(3)$ can be represented by means of series which converge much more rapidly than that in Euler's celebrated formula (6) as well as that in the series (7) which was used recently by Apéry [3] in his proof of the *irrationality* of $\zeta\,(3)$. Symbolic and numerical computations using *Mathematica* (Version 4.0) for Linux show, among other things, that only 50 terms of one of these series are capable of producing an accuracy of seven decimal places. In the second part of this article, we consider a variety of series and integrals associated with the Hurwitz-Lerch Zeta function $\Phi(z,s,a)$ as well as its various interesting extensions and generalizations (see Sect. 6).

2 Series Representations for $\zeta\,(2n+1)$ $(n \in \mathbb{N})$

The following simple consequence of the binomial theorem *and* the definition (1):

$$\sum_{k=0}^{\infty} \frac{(s)_k}{k!}\,\zeta\,(s+k,a)\,t^k = \zeta\,(s,a-t) \qquad (|t| < |a|)\,, \tag{11}$$

yields, for $a = 1$ and $t = \pm 1/m$, a useful the series identity in the form:

$$\sum_{k=0}^{\infty} \frac{(s)_{2k}}{(2k)!}\,\frac{\zeta\,(s+2k)}{m^{2k}}$$

$$= \begin{cases} (2^s-1)\ \zeta\,(s) - 2^{s-1} & (m=2) \\ \dfrac{1}{2}\left[(m^s-1)\ \zeta\,(s) - m^s - \displaystyle\sum_{j=2}^{m-2} \zeta\left(s, \frac{j}{m}\right)\right] & (m \in \mathbb{N} \setminus \{1,2\})\,, \end{cases} \tag{12}$$

where $(\lambda)_\nu$ denotes the Pochhammer symbol or the *shifted factorial*, since

$$(1)_n = n! \qquad (n \in \mathbb{N}_0),$$

which is defined for $\lambda, \nu \in \mathbb{C}$, in terms of the familiar Gamma function, by

$$(\lambda)_\nu := \frac{\Gamma(\lambda+\nu)}{\Gamma(\lambda)} = \begin{cases} 1 & (\nu = 0;\ \lambda \in \mathbb{C} \setminus \{0\}) \\ \lambda(\lambda+1)\cdots(\lambda+n-1) & (\nu = n \in \mathbb{N};\ \lambda \in \mathbb{C}), \end{cases}$$

it being understood *conventionally* that $(0)_0 := 1$. (See, for details, [38, 43]).

Making use of the familiar harmonic numbers H_n given by

$$H_n := \sum_{j=1}^{n} \frac{1}{j} \qquad (n \in \mathbb{N}), \tag{13}$$

the following set of series representations for $\zeta(2n+1)$ $(n \in \mathbb{N})$ were proven by Srivastava [40] by appealing appropriately to the series identity (12) in its special cases when $m = 2, 3, 4$, and 6, and also to many other properties and characteristics of the Riemann Zeta function such as the familiar functional equation:

$$\zeta(s) = 2 \cdot (2\pi)^{s-1} \sin\left(\frac{1}{2}\pi s\right) \Gamma(1-s)\, \zeta(1-s) \tag{14}$$

or, equivalently,

$$\zeta(1-s) = 2 \cdot (2\pi)^{-s} \cos\left(\frac{1}{2}\pi s\right) \Gamma(s)\, \zeta(s), \tag{15}$$

the familiar derivative formula:

$$\begin{aligned}\zeta'(-2n) &= \lim_{\varepsilon \to 0} \left\{ \frac{\zeta(-2n+\varepsilon)}{\varepsilon} \right\} \\ &= \frac{(-1)^n}{2 \cdot (2\pi)^{2n}} (2n)!\, \zeta(2n+1) \qquad (n \in \mathbb{N})\end{aligned} \tag{16}$$

with, of course,

$$\zeta(0) = -\frac{1}{2}; \qquad \zeta(-2n) = 0 \qquad (n \in \mathbb{N}); \qquad \zeta'(0) = -\frac{1}{2}\log(2\pi), \tag{17}$$

and each of the following limit relationships:

$$\lim_{s \to -2n} \left\{ \frac{\sin\left(\frac{1}{2}\pi s\right)}{s+2n} \right\} = (-1)^n \frac{\pi}{2} \qquad (n \in \mathbb{N}) \tag{18}$$

and

$$\begin{aligned}\lim_{s \to -2n} \left\{ \frac{\zeta(s+2k)}{s+2n} \right\} &= \frac{(-1)^{n-k}}{2 \cdot (2\pi)^{2(n-k)}} (2n-2k)!\, \zeta(2n-2k+1) \\ &\qquad (k = 1, \ldots, n-1;\ n \in \mathbb{N} \setminus \{1\}).\end{aligned} \tag{19}$$

First Series Representation:

$$\zeta(2n+1) = (-1)^{n-1} \frac{(2\pi)^{2n}}{2^{2n+1}-1} \left[\frac{H_{2n} - \log \pi}{(2n)!} + \sum_{k=1}^{n-1} \frac{(-1)^k}{(2n-2k)!} \frac{\zeta(2k+1)}{\pi^{2k}} \right.$$
$$\left. +2 \sum_{k=1}^{\infty} \frac{(2k-1)!}{(2n+2k)!} \frac{\zeta(2k)}{2^{2k}} \right] \qquad (n \in \mathbb{N}). \tag{20}$$

Second Series Representation:

$$\zeta(2n+1) = (-1)^{n-1} \frac{2 \cdot (2\pi)^{2n}}{3^{2n+1}-1} \left[\frac{H_{2n} - \log\left(\frac{2}{3}\pi\right)}{(2n)!} + \sum_{k=1}^{n-1} \frac{(-1)^k}{(2n-2k)!} \frac{\zeta(2k+1)}{\left(\frac{2}{3}\pi\right)^{2k}} \right.$$
$$\left. +2 \sum_{k=1}^{\infty} \frac{(2k-1)!}{(2n+2k)!} \frac{\zeta(2k)}{3^{2k}} \right] \qquad (n \in \mathbb{N}). \tag{21}$$

Third Series Representation:

$$\zeta(2n+1) = (-1)^{n-1} \frac{2 \cdot (2\pi)^{2n}}{2^{4n+1}+2^{2n}-1} \left[\frac{H_{2n} - \log\left(\frac{1}{2}\pi\right)}{(2n)!} \right.$$
$$\left. + \sum_{k=1}^{n-1} \frac{(-1)^k}{(2n-2k)!} \frac{\zeta(2k+1)}{\left(\frac{1}{2}\pi\right)^{2k}} + 2 \sum_{k=1}^{\infty} \frac{(2k-1)!}{(2n+2k)!} \frac{\zeta(2k)}{4^{2k}} \right] \qquad (n \in \mathbb{N}). \tag{22}$$

Fourth Series Representation:

$$\zeta(2n+1) = (-1)^{n-1} \frac{2 \cdot (2\pi)^{2n}}{3^{2n}\left(2^{2n}+1\right)+2^{2n}-1} \left[\frac{H_{2n} - \log\left(\frac{1}{3}\pi\right)}{(2n)!} \right.$$
$$\left. + \sum_{k=1}^{n-1} \frac{(-1)^k}{(2n-2k)!} \frac{\zeta(2k+1)}{\left(\frac{1}{3}\pi\right)^{2k}} + 2 \sum_{k=1}^{\infty} \frac{(2k-1)!}{(2n+2k)!} \frac{\zeta(2k)}{6^{2k}} \right] \qquad (n \in \mathbb{N}). \tag{23}$$

Here, *as well as elsewhere in this presentation*, an empty sum is understood (as usual) to be zero.

The *first* series representation (20) is markedly different from each of the series representations for $\zeta(2n+1)$, which were given earlier by Zhang and Williams [60, p. 1590, Eq. 3.13] and (subsequently) by Cvijović and Klinowski [8, p. 1265, Theorem A] (see also [61, 62]). Since $\zeta(2k) \to 1$ as $k \to \infty$, the general term in the series representation (20) has the following *order estimate*:

$$O\left(2^{-2k} \cdot k^{-2n-1}\right) \qquad (k \to \infty;\ n \in \mathbb{N}),$$

whereas the general term in each of the aforecited *earlier* series representations has the order estimate given below:

$$O\left(2^{-2k}\cdot k^{-2n}\right) \qquad (k\to\infty;\ n\in\mathbb{N}).$$

In case we suitably combine (20) and (22), we readily obtain the following series representation:

$$\begin{aligned}\zeta(2n+1) = (-1)^{n-1}\,\frac{2\cdot(2\pi)^{2n}}{(2^{2n}-1)(2^{2n+1}-1)}\Bigg[&\frac{\log 2}{(2n)!}\\ &+\sum_{k=1}^{n-1}\frac{(-1)^k\left(2^{2k}-1\right)}{(2n-2k)!}\,\frac{\zeta(2k+1)}{\pi^{2k}}\\ &-2\sum_{k=1}^{\infty}\frac{(2k-1)!\left(2^{2k}-1\right)}{(2n+2k)!}\,\frac{\zeta(2k)}{2^{4k}}\Bigg] \qquad (n\in\mathbb{N}). \qquad (24)\end{aligned}$$

Moreover, in terms of the Bernoulli numbers B_n and the Euler polynomials $E_n(x)$ defined by the generating functions (9) and

$$\frac{2e^{xz}}{e^z+1} = \sum_{n=0}^{\infty} E_n(x)\,\frac{z^n}{n!} \qquad (|z|<\pi), \qquad (25)$$

respectively, it is known that (cf., e.g., [27, p. 29])

$$E_n(0) = (-1)^n E_n(1) = \frac{2\left(1-2^{n+1}\right)}{n+1}\,B_{n+1} \qquad (n\in\mathbb{N}). \qquad (26)$$

Thus, by combining (26) with the identity (8), we find that

$$E_{2n-1}(0) = \frac{4\cdot(-1)^n}{(2\pi)^{2n}}\,(2n-1)!\left(2^{2n}-1\right)\,\zeta(2n) \qquad (n\in\mathbb{N}). \qquad (27)$$

If we apply the relationship (27), the series representation (24) can immediately be put in the following *alternative* form:

$$\begin{aligned}\zeta(2n+1) = (-1)^{n-1}\,\frac{2\cdot(2\pi)^{2n}}{(2^{2n}-1)(2^{2n+1}-1)}\Bigg[&\frac{\log 2}{(2n)!}\\ &+\sum_{k=1}^{n-1}\frac{(-1)^k\left(2^{2k}-1\right)}{(2n-2k)!}\,\frac{\zeta(2k+1)}{\pi^{2k}}\\ &+\frac{1}{2}\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{(2n+2k)!}\left(\frac{\pi}{2}\right)^{2k} E_{2k-1}(0)\Bigg] \qquad (n\in\mathbb{N}), \qquad (28)\end{aligned}$$

which is a slightly *modified* and *corrected* version of a result proven, using a significantly different technique, by Tsumura [56, p. 383, Theorem B].

One other interesting combination of the series representations (20) and (22) leads us to the following variant of Tsumura's result (24) or (28):

$$\zeta(2n+1) = (-1)^{n-1} \frac{\pi^{2n}}{2^{2n+1}-1} \left[\frac{H_{2n} - \log\left(\frac{1}{4}\pi\right)}{(2n)!} + \sum_{k=1}^{n-1} \frac{(-1)^k \left(2^{2k+1}-1\right)}{(2n-2k)!} \frac{\zeta(2k+1)}{\pi^{2k}} - 4 \sum_{k=1}^{\infty} \frac{(2k-1)! \left(2^{2k-1}-1\right)}{(2n+2k)!} \frac{\zeta(2k)}{2^{4k}} \right] \qquad (n \in \mathbb{N}), \quad (29)$$

which is essentially the same as the *determinantal* expression for $\zeta(2n+1)$ derived by Ewell [13, p. 1010, Corollary 3] by employing an entirely different technique from ours.

A number of other similar combinations of the series representations (20) to (23) would yield some interesting companions of Ewell's result (29).

Next, by setting $t = \frac{1}{m}$ and differentiating both sides with respect to s, we find from the following obvious consequence of the series identity (11):

$$\sum_{k=0}^{\infty} \frac{(s)_{2k+1}}{(2k+1)!} \zeta(s+2k+1, a)\ t^{2k+1} = \frac{1}{2} \left[\zeta(s, a-t) - \zeta(s, a+t)\right] \qquad (|t| < |a|) \qquad (30)$$

that

$$\sum_{k=0}^{\infty} \frac{(s)_{2k+1}}{(2k+1)!\, m^{2k}} \left[\zeta'(s+2k+1, a) + \zeta(s+2k+1, a) \sum_{j=0}^{2k} \frac{1}{s+j} \right] = \frac{m}{2} \frac{\partial}{\partial s} \left\{ \zeta\left(s, a - \frac{1}{m}\right) - \zeta\left(s, a + \frac{1}{m}\right) \right\} \qquad (m \in \mathbb{N} \setminus \{1\}). \quad (31)$$

In the particular case when $m = 2$, (31) immediately yields

$$\sum_{k=0}^{\infty} \frac{(s)_{2k+1}}{(2k+1)!\, 2^{2k}} \left[\zeta'(s+2k+1, a) + \zeta(s+2k+1, a) \sum_{j=0}^{2k} \frac{1}{s+j} \right] = -\left(a - \frac{1}{2}\right)^{-s} \log\left(a - \frac{1}{2}\right). \qquad (32)$$

Upon letting $s \to -2n-1$ $(n \in \mathbb{N})$ in the *further* special of this last identity (32) when $a = 1$, Wilton [43, p. 92] deduced the following series representation for $\zeta(2n+1)$ (see also [20, p. 357, Entry (54.6.9)]):

$$\zeta(2n+1) = (-1)^{n-1}\ \pi^{2n}\left[\frac{H_{2n+1}-\log\pi}{(2n+1)!} + \sum_{k=1}^{n-1}\frac{(-1)^k}{(2n-2k+1)!}\ \frac{\zeta(2k+1)}{\pi^{2k}}\right.$$
$$\left. +2\sum_{k=1}^{\infty}\frac{(2k-1)!}{(2n+2k+1)!}\ \frac{\zeta(2k)}{2^{2k}}\right] \qquad (n \in \mathbb{N}), \tag{33}$$

which, in light of the elementary identity:

$$\frac{(2k)!}{(2n+2k)!} = \frac{(2k-1)!}{(2n+2k-1)!} - 2n\,\frac{(2k-1)!}{(2n+2k)!} \qquad (n \in \mathbb{N}), \tag{34}$$

would combine with the result (20) to yield the following series representation:

$$\zeta(2n+1) = (-1)^n\,\frac{(2\pi)^{2n}}{n\,(2^{2n+1}-1)}\left[\sum_{k=1}^{n-1}\frac{(-1)^{k-1}\,k}{(2n-2k)!}\ \frac{\zeta(2k+1)}{\pi^{2k}}\right.$$
$$\left. +\sum_{k=0}^{\infty}\frac{(2k)!}{(2n+2k)!}\ \frac{\zeta(2k)}{2^{2k}}\right] \qquad (n \in \mathbb{N}). \tag{35}$$

This last series representation (35) is precisely the aforementioned *main* result of Cvijović and Klinowski [8, p. 1265, Theorem A]. As a matter of fact, in view of a known derivative formula [40, p. 389, Eq. 2.8], the series representation (35) is essentially the same as a result given earlier by Zhang and Williams [60, p. 1590, Eq. 3.13] (see also Zhang and Williams [60, p. 1591, Eq. 3.16] where an obviously more complicated (*asymptotic*) version of (35) was proven similarly).

In light of another elementary identity:

$$\frac{(2k)!}{(2n+2k+1)!} = \frac{(2k-1)!}{(2n+2k)!} - (2n+1)\,\frac{(2k-1)!}{(2n+2k+1)!} \qquad (n, k \in \mathbb{N}), \tag{36}$$

we can obtain the following yet another series representation for $\zeta(2n+1)$ by applying (20) and (33):

$$\zeta(2n+1) = (-1)^n\,\frac{2\cdot(2\pi)^{2n}}{(2n-1)\,2^{2n}+1}\left[\sum_{k=1}^{n-1}\frac{(-1)^{k-1}\,k}{(2n-2k+1)!}\ \frac{\zeta(2k+1)}{\pi^{2k}}\right.$$
$$\left. +\sum_{k=0}^{\infty}\frac{(2k)!}{(2n+2k+1)!}\ \frac{\zeta(2k)}{2^{2k}}\right] \qquad (n \in \mathbb{N}), \tag{37}$$

which provides a *significantly* simpler (and *much more* rapidly convergent) version of the following other *main* result of Cvijović and Klinowski [8, p. 1265, Theorem B]:

$$\zeta\,(2n+1) = (-1)^n\,\frac{2\cdot(2\pi)^{2n}}{(2n)!}\sum_{k=0}^{\infty}\Omega_{n,k}\,\frac{\zeta\,(2k)}{2^{2k}} \qquad (n\in\mathbb{N}), \tag{38}$$

where the coefficients $\Omega_{n,k}$ $(n\in\mathbb{N};\ k\in\mathbb{N}_0)$ are given explicitly as a finite sum of Bernoulli numbers [8, p. 1265, Theorem B(i)] (see, for details, Srivastava [40, pp. 393–394]):

$$\Omega_{n,k} := \sum_{j=0}^{2n}\binom{2n}{j}\frac{B_{2n-j}}{(j+2k+1)\,(j+1)\,2^j} \qquad (n\in\mathbb{N};\ k\in\mathbb{N}_0). \tag{39}$$

3 Other Families of Series Representations for $\zeta\,(2n+1)$ $(n\in\mathbb{N})$

In this section, we start once again from the identity (11) with (of course) $a=1$, $t=\pm 1/m$, and s replaced by $s+1$. Thus, by applying (12), we find yet another class of series identities including, for example,

$$\sum_{k=1}^{\infty}\frac{(s+1)_{2k}}{(2k)!}\,\frac{\zeta\,(s+2k)}{2^{2k}} = (2^s-2)\;\zeta\,(s) \tag{40}$$

and

$$\begin{aligned}\sum_{k=1}^{\infty}&\frac{(s+1)_{2k}}{(2k)!}\,\frac{\zeta\,(s+2k)}{m^{2k}}\\ &= \frac{1}{2m}\Bigg[m\,(m^s-3)\,\zeta\,(s) + \left(m^{s+1}-1\right)\zeta\,(s+1) - 2\zeta\left(s+1,\frac{1}{m}\right)\\ &\qquad -\sum_{j=2}^{m-2}\left\{m\zeta\left(s,\frac{j}{m}\right)+\zeta\left(s+1,\frac{j}{m}\right)\right\}\Bigg] \qquad (m\in\mathbb{N}\setminus\{1,2\}).\end{aligned} \tag{41}$$

In fact, it is the series identity (40) which was first applied by Zhang and Williams [60] (and, subsequently, by Cvijović and Klinowski [8]) with a view to proving two (only seemingly different) versions of the series representation (35). Indeed, if we appeal to (41) *with* $m=4$, we can derive the following much more rapidly convergent series representation for $\zeta\,(2n+1)$ (see [39, p. 9, Eq. 41]):

$$\begin{aligned}\zeta(2n+1) = (-1)^n \, \frac{2\cdot(2\pi)^{2n}}{n\,(2^{4n+1}+2^{2n}-1)} \Bigg[& \frac{4^{n-1}-1}{(2n)!}\, B_{2n}\log 2 \\ & - \frac{2^{2n-1}-1}{2\,(2n-1)!}\,\zeta'(1-2n) - \frac{4^{2n-1}}{(2n-1)!}\,\zeta'\left(1-2n,\frac{1}{4}\right) \\ & + \sum_{k=1}^{n-1} \frac{(-1)^{k-1}\,k}{(2n-2k)!}\,\frac{\zeta(2k+1)}{\left(\frac{1}{2}\pi\right)^{2k}} + \sum_{k=0}^{\infty} \frac{(2k)!}{(2n+2k)!}\,\frac{\zeta(2k)}{4^{2k}} \Bigg] \qquad (42)\end{aligned}$$

$$(n \in \mathbb{N}),$$

where (*and in what follows*) a prime denotes the derivative of $\zeta(s)$ or $\zeta(s,a)$ with respect to s.

By virtue of the identities (34) and (36), the results (22) and (42) would lead us eventually to the following *additional* series representations for $\zeta(2n+1)$ $(n \in \mathbb{N})$ (see [39, p. 10, Eqs. 42 and 43]):

$$\begin{aligned}\zeta(2n+1) = (-1)^{n-1}\left(\frac{\pi}{2}\right)^{2n} \Bigg[& \frac{H_{2n+1}-\log\left(\frac{1}{2}\pi\right)}{(2n+1)!} + \frac{2\,(4^n-1)}{(2n+2)!} B_{2n+2}\log 2 \\ & - \frac{2^{2n+1}-1}{(2n+1)!}\,\zeta'(-2n-1) - \frac{2^{4n+3}}{(2n+1)!}\,\zeta'\left(-2n-1,\frac{1}{4}\right) \\ & + \sum_{k=1}^{n-1} \frac{(-1)^{k}}{(2n-2k+1)!}\,\frac{\zeta(2k+1)}{\left(\frac{1}{2}\pi\right)^{2k}} + 2\sum_{k=1}^{\infty} \frac{(2k-1)!}{(2n+2k+1)!}\,\frac{\zeta(2k)}{4^{2k}} \Bigg] \qquad (43)\end{aligned}$$

$$(n \in \mathbb{N})$$

and

$$\begin{aligned}\zeta(2n+1) = (-1)^n \, \frac{4\cdot(2\pi)^{2n}}{n\cdot 4^{2n+1}-2^{2n}+1} \Bigg[& \frac{2^{2n+1}-1}{2\cdot(2n)!}\,\zeta'(-2n-1) \\ & + \frac{4^{2n+1}}{(2n)!}\,\zeta'\left(-2n-1,\frac{1}{4}\right) - \frac{(2n+1)\,(4^n-1)}{(2n+2)!}\,B_{2n+2}\log 2 \\ & + \sum_{k=1}^{n-1} \frac{(-1)^{k-1}\,k}{(2n-2k+1)!}\,\frac{\zeta(2k+1)}{\left(\frac{1}{2}\pi\right)^{2k}} + \sum_{k=0}^{\infty} \frac{(2k)!}{(2n+2k+1)!}\,\frac{\zeta(2k)}{4^{2k}} \Bigg] \qquad (44)\end{aligned}$$

$$(n \in \mathbb{N}).$$

Explicit expressions for the derivatives $\zeta'(-2n \pm 1)$ and $\zeta'\left(-2n \pm 1, \frac{1}{4}\right)$, occurring in the series representations (42)–(44), can be found and substituted into these results in order to represent $\zeta(2n+1)$ in terms of Bernoulli numbers and polynomials and various rapidly convergent series of the ζ-functions (see, for details, the work by Srivastava [39, Sect. 3]).

Out of the four seemingly analogous results (22), (42), (43), and (44), the infinite series in (43) would obviously converge most rapidly, with its general term having the order estimate:

$$O\left(k^{-2n-2}\cdot 4^{-2k}\right) \qquad (k\to\infty;\ n\in\mathbb{N}).$$

From the work by Srivastava and Tsumura [45], we recall the following three *new* members of the class of the series representations (22) and (43):

$$\begin{aligned}\zeta(2n+1) = (-1)^{n-1}\left(\frac{2\pi}{3}\right)^{2n}&\left[\frac{H_{2n+1}-\log\left(\frac{2}{3}\pi\right)}{(2n+1)!}+\frac{\left(3^{2n+2}-1\right)\pi}{2\sqrt{3}\,(2n+2)!}\,B_{2n+2}\right.\\&+\frac{(-1)^{n-1}}{\sqrt{3}\,(2\pi)^{2n+1}}\,\zeta\left(2n+2,\frac{1}{3}\right)\\&\left.+\sum_{k=1}^{n-1}\frac{(-1)^k}{(2n-2k+1)!}\,\frac{\zeta(2k+1)}{\left(\frac{2}{3}\pi\right)^{2k}}+2\sum_{k=1}^{\infty}\frac{(2k-1)!}{(2n+2k+1)!}\,\frac{\zeta(2k)}{3^{2k}}\right] \qquad (45)\\&(n\in\mathbb{N}).\end{aligned}$$

$$\begin{aligned}\zeta(2n+1) = (-1)^{n-1}\left(\frac{\pi}{2}\right)^{2n}&\left[\frac{H_{2n+1}-\log\left(\frac{1}{2}\pi\right)}{(2n+1)!}+\frac{2^{2n}\left(2^{2n+2}-1\right)\pi}{(2n+2)!}\,B_{2n+2}\right.\\&+\frac{(-1)^{n-1}}{2\cdot(2\pi)^{2n+1}}\,\zeta\left(2n+2,\frac{1}{4}\right)\\&\left.+\sum_{k=1}^{n-1}\frac{(-1)^k}{(2n-2k+1)!}\,\frac{\zeta(2k+1)}{\left(\frac{1}{2}\pi\right)^{2k}}+2\sum_{k=1}^{\infty}\frac{(2k-1)!}{(2n+2k+1)!}\,\frac{\zeta(2k)}{4^{2k}}\right] \qquad (46)\\&(n\in\mathbb{N})\end{aligned}$$

and

$$\begin{aligned}\zeta(2n+1) = (-1)^{n-1}\left(\frac{\pi}{3}\right)^{2n}&\left[\frac{H_{2n+1}-\log\left(\frac{1}{3}\pi\right)}{(2n+1)!}+\frac{2^{2n}\left(3^{2n+2}-1\right)\pi}{\sqrt{3}\,(2n+2)!}\,B_{2n+2}\right.\\&+\frac{(-1)^{n-1}}{2\sqrt{3}\,(2\pi)^{2n+1}}\left\{\zeta\left(2n+2,\frac{1}{3}\right)+\zeta\left(2n+2,\frac{1}{6}\right)\right\}\\&\left.+\sum_{k=1}^{n-1}\frac{(-1)^k}{(2n-2k+1)!}\,\frac{\zeta(2k+1)}{\left(\frac{1}{3}\pi\right)^{2k}}+2\sum_{k=1}^{\infty}\frac{(2k-1)!}{(2n+2k+1)!}\,\frac{\zeta(2k)}{6^{2k}}\right] \qquad (47)\\&(n\in\mathbb{N}).\end{aligned}$$

The general terms of the infinite series occurring in these three members (45)–(47) have the order estimates:

$$O\left(k^{-2n-2}\cdot m^{-2k}\right) \qquad (k\to\infty;\ n\in\mathbb{N};\ m=3,4,6)\,, \tag{48}$$

which exhibit the fact that *each* of these last three series representations (45)–(47) converges more rapidly than Wilton's result (33) and two of them [cf. (46) and (47)] at least as rapidly as Srivastava's result (43).

We next recall that, in their aforementioned work on the Ray-Singer torsion and topological field theories, Nash and O'Connor [29, 30] obtained a number of remarkable integral expressions for $\zeta(3)$, including (for example) the following result [26, p. 1489 *et seq.*]:

$$\zeta(3)=\frac{2\pi^2}{7}\log 2-\frac{8}{7}\int_0^{\pi/2} z^2\cot z\,dz. \tag{49}$$

In fact, in view of the following series expansion [10, p. 51, Eq. 1.20(3)]:

$$z\cot z=-2\sum_{k=0}^{\infty}\zeta(2k)\left(\frac{z}{\pi}\right)^{2k} \qquad (|z|<\pi)\,, \tag{50}$$

the result (49) equivalent to the series representation (cf. the work by Dąbrowski [9, p. 202]; see also the paper by Chen and Srivastava [13, p. 191, Eq. 3.19]):

$$\zeta(3)=\frac{2\pi^2}{7}\left(\log 2+\sum_{k=0}^{\infty}\frac{\zeta(2k)}{(k+1)\,2^{2k}}\right). \tag{51}$$

Moreover, if we integrate by parts, we easily find that

$$\int_0^{\pi/2} z^2\cot z\,dz=-2\int_0^{\pi/2} z\log\sin z\,dz, \tag{52}$$

so that the result (49) is equivalent *also* to the following integral representation:

$$\zeta(3)=\frac{2\pi^2}{7}\log 2+\frac{16}{7}\int_0^{\pi/2} z\log\sin z\,dz, \tag{53}$$

which was proven in the aforementioned 1772 paper by Euler (cf., e.g., [6, p. 1084]).

Furthermore, since

$$i\cot iz=\coth z=\frac{2}{e^{2z}-1}+1 \qquad \left(i:=\sqrt{-1}\right), \tag{54}$$

by replacing z in the known expansion (50) by $\frac{1}{2}i\pi z$, it is easily seen that (cf., e.g., [13, p. 25]; see also [10, p. 51, Eq. 1.20(1)])

$$\frac{\pi z}{e^{\pi z}-1}+\frac{\pi z}{2}=\sum_{k=0}^{\infty}\frac{(-1)^{k+1}\,\zeta(2k)}{2^{2k-1}}z^{2k} \qquad (|z|<2)\,. \tag{55}$$

Upon setting $z = it$ in (55), multiplying both sides by t^{m-1} ($m \in \mathbb{N}$), and then integrating the resulting equation from $t = 0$ to $t = \tau$ $(0 < \tau < 2)$, Srivastava [27] derived the following series representations for $\zeta(2n+1)$ (see also the work by Srivastava et al. [49]):

$$\zeta(2n+1)=(-1)^{n-1}\frac{(2\pi)^{2n}}{(2n)!\,(2^{2n+1}-1)} \cdot\left[\log 2+\sum_{j=1}^{n-1}(-1)^j\binom{2n}{2j}\frac{(2j)!\,(2^{2j}-1)}{(2\pi)^{2j}}\,\zeta(2j+1)+\sum_{k=0}^{\infty}\frac{\zeta(2k)}{(k+n)\,2^{2k}}\right] \tag{56}$$

$$(n \in \mathbb{N})$$

and

$$\zeta(2n+1)=(-1)^{n-1}\frac{(2\pi)^{2n}}{(2n+1)!\,(2^{2n}-1)} \cdot\left[\log 2+\sum_{j=1}^{n-1}(-1)^j\binom{2n+1}{2j}\frac{(2j)!\,(2^{2j}-1)}{(2\pi)^{2j}}\,\zeta(2j+1) +\sum_{k=0}^{\infty}\frac{\zeta(2k)}{\left(k+n+\frac{1}{2}\right)2^{2k}}\right] \qquad (n\in\mathbb{N})\,. \tag{57}$$

Upon setting $n = 1$, (57) immediately reduces to the following series representation for $\zeta(3)$:

$$\zeta(3)=\frac{2\pi^2}{9}\left(\log 2+2\sum_{k=0}^{\infty}\frac{\zeta(2k)}{(2k+3)\,2^{2k}}\right), \tag{58}$$

which was proven *independently* by (among others) Glasser [16, p. 446, Eq. 12], Zhang and Williams [60, p. 1585, Eq. 2.13], and Dąbrowski [9, p. 206] (see also the work by Chen and Srivastava [13, p. 183, Eq. 2.15]). Furthermore, a special case of (56) when $n = 1$ yields (cf. Dąbrowski [9, p. 202]; see also Chen and Srivastava [13, 5, p. 191, Eq. 3.19])

$$\zeta(3) = \frac{2\pi^2}{7}\left(\log 2 + \sum_{k=0}^{\infty} \frac{\zeta(2k)}{(k+1)\,2^{2k}}\right). \tag{59}$$

In fact, in view of the following familiar sum:

$$\sum_{k=0}^{\infty} \frac{\zeta(2k)}{(2k+1)\,2^{2k}} = -\frac{1}{2}\log 2, \tag{60}$$

Euler's formula (6) is indeed a rather *simple* consequence of (59).

In passing, we find it worthwhile to remark that an integral representation for $\zeta(2n+1)$, which is easily seen to be equivalent to the series representation (56), was given by Dąbrowski [9, p. 203, Eq. 16], who [9, p. 206] mentioned the existence of (but did not fully state) the series representation (57) as well. The series representation (56) was derived also in a paper by Borwein et al. (cf. [11, p. 269, Eq. 57]).

If we suitably combine the series occurring in (51), (58), and (60), it is not difficult to deduce several other series representations for $\zeta(3)$, which are analogous to Euler's formula (6). More generally, since

$$\frac{\lambda k^2 + \mu k + \nu}{(2k+2n-1)(2k+2n)(2k+2n+1)} = \frac{\mathcal{A}}{2k+2n-1} + \frac{\mathcal{B}}{2k+2n} + \frac{\mathcal{C}}{2k+2n+1}, \tag{61}$$

where, for convenience,

$$\mathcal{A} = \mathcal{A}_n(\lambda,\mu,\nu) := \frac{1}{2}\left[\lambda n^2 - (\lambda+\mu)\,n + \frac{1}{4}(\lambda + 2\mu + 4\nu)\right], \tag{62}$$

$$\mathcal{B} = \mathcal{B}_n(\lambda,\mu,\nu) := -\left(\lambda n^2 - \mu n + \nu\right), \tag{63}$$

and

$$\mathcal{C} = \mathcal{C}_n(\lambda,\mu,\nu) := \frac{1}{2}\left[\lambda n^2 + (\lambda-\mu)\,n + \frac{1}{4}(\lambda - 2\mu + 4\nu)\right], \tag{64}$$

by applying (56), (57), and another result (proven by Srivastava [41, p. 341, Eq.3.17]):

$$\sum_{j=1}^{n} (-1)^{j-1}\binom{2n+1}{2j}\frac{(2j)!\,\left(2^{2j}-1\right)}{(2\pi)^{2j}}\,\zeta(2j+1) = \log 2 + \sum_{k=0}^{\infty}\frac{\zeta(2k)}{\left(k+n+\frac{1}{2}\right)2^{2k}} \qquad (n \in \mathbb{N}_0), \tag{65}$$

with n replaced by $n-1$, Srivastava [41] derived the following unification of a large number of known (or new) series representations for $\zeta\,(2n+1)$ $(n \in \mathbb{N})$, including (for example) Euler's formula (6):

$$\begin{aligned}\zeta\,(2n+1) = {} & \frac{(-1)^{n-1}\,(2\pi)^{2n}}{(2n)!\,\{(2^{2n+1}-1)\,\mathcal{B} + (2n+1)\,(2^{2n}-1)\,\mathcal{C}\}} \\ & \cdot \left[\frac{1}{4}\lambda \log 2 + \sum_{j=1}^{n-1} (-1)^j \binom{2n-1}{2j-2}\right. \\ & \qquad \cdot \left\{2j\,(2j-1)\,\mathcal{A} + [\lambda\,(4n-1) - 2\mu]\,nj + \lambda n \left(n+\frac{1}{2}\right)\right\} \\ & \qquad \cdot \frac{(2j-2)!\,\left(2^{2j}-1\right)}{(2\pi)^{2j}}\,\zeta\,(2j+1) \\ & \left. + \sum_{k=0}^{\infty} \frac{\left(\lambda k^2 + \mu k + \nu\right)\,\zeta\,(2k)}{(2k+2n-1)\,(k+n)\,(2k+2n+1)\,2^{2k}}\right] \end{aligned} \tag{66}$$

$$(n \in \mathbb{N};\ \lambda, \mu, \nu \in \mathbb{C}),$$

where $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ are given by (62)–(64), respectively.

Numerous other interesting series representations for $\zeta\,(2n+1)$, which are analogous to (56) and (57), were also given by Srivastava et al. [49].

4 Computationally Useful Deductions and Consequences

In this section, we suitably specialize the parameter λ, μ, and ν in (66) and then apply a rather elaborate scheme. We thus eventually arrive at the following remarkably rapidly convergent series representation for $\zeta\,(2n+1)$ $(n \in \mathbb{N})$, which was derived by Srivastava [41, pp. 348–349, Eq. 3.50]):

$$\begin{aligned}\zeta\,(2n+1) = {} & (-1)^{n-1}\,\frac{(2\pi)^{2n}}{(2n)!\varDelta_n}\left[\sum_{j=1}^{n-1}(-1)^j\right. \\ & \cdot \left(\left\{(2n-3)\,2^{2n+2} - 2n\right\}\left\{\binom{2n-1}{2j} - \binom{2n+2}{2j} + 6n\binom{2n-1}{2j-2}\right\} - \left(2^{2n+3}-1\right)\right. \\ & \cdot \left.\left\{\binom{2n}{2j} - \binom{2n+3}{2j} + 3\binom{2n+1}{2j-1}\right\}\right)\frac{(2j)!\,\left(2^{2j}-1\right)}{(2\pi)^{2j}}\,\zeta\,(2j+1)\end{aligned}$$

$$\left. +12\sum_{k=0}^{\infty}\frac{(\xi_n k+\eta_n)\ \zeta(2k)}{(2k+2n-1)(2k+2n)(2k+2n+1)(2k+2n+2)(2k+2n+3)\,2^{2k}}\right] \tag{67}$$

$$(n\in\mathbb{N}),$$

where, for convenience,

$$\begin{aligned}\Delta_n:=&\left(2^{2n+3}-1\right)\left\{\frac{1}{3}(2n+1)\left(2n^2-4n+3\right)\left(2^{2n}-1\right)-2^{2n+1}+1\right\}\\&-\left\{(2n-3)\,2^{2n+2}-2n\right\}\left\{2^{2n+2}+n(2n-3)\left(2^{2n}-1\right)-1\right\},\end{aligned} \tag{68}$$

$$\xi_n:=2\left\{(2n-5)\,2^{2n+2}-2n+1\right\}, \tag{69}$$

and

$$\eta_n:=\left(4n^2-4n-7\right)2^{2n+2}-(2n+1)^2. \tag{70}$$

In its special case when $n=1$, (67) yields the following (*rather curious*) series representation:

$$\zeta(3)=-\frac{6\pi^2}{23}\sum_{k=0}^{\infty}\frac{(98k+121)\ \zeta(2k)}{(2k+1)(2k+2)(2k+3)(2k+4)(2k+5)\,2^{2k}}, \tag{71}$$

where the series obviously converges much more rapidly than that in *each* of the *celebrated* results (6) and (7).

An interesting companion of (71) in the following form:

$$\begin{aligned}\zeta(3)=&-\frac{120}{1573}\pi^2\\&\cdot\sum_{k=0}^{\infty}\frac{8576k^2+24286k+17283}{(2k+1)(2k+2)(2k+3)(2k+4)(2k+5)(2k+6)(2k+7)}\frac{\zeta(2k)}{2^{2k}}.\end{aligned} \tag{72}$$

was deduced by Srivastava and Tsumura [47], who indeed presented an *inductive* construction of several general series representations for $\zeta(2n+1)$ $(n\in\mathbb{N})$ (see also [46]).

5 Numerical Verifications and Symbolic Computations Based Upon *Mathematica*

In this section, we first summarize the results of numerical verifications and symbolic computations with the series in (71) by using *Mathematica* (Version 4.0) for Linux:

$$\begin{aligned}
\text{In[1]} &:= (98k+121)\ \text{Zeta}\,[2k]\,/((2k+1)\,(2k+2)\,(2k+3)\\
&\quad \times (2k+4)\,(2k+5)\,2^{\urcorner}\,(2k))\\
\text{Out[1]} &= \frac{(121+98k)\,\text{Zeta}\,[2k]}{2^{2k}\,(1+2k)\,(2+2k)\,(3+2k)\,(4+2k)\,(5+2k)}\\
\text{In[2]} &:= \text{Sum}[\%, \{k, 1,\ \text{Infinity}\}]//\ \text{Simplify}\\
\text{Out[2]} &= \frac{121}{240}-\frac{23\ \text{Zeta[3]}}{6\text{Pi}^2}\\
\text{In[3]} &:= \text{N}[\%]\\
\text{Out[3]} &= 0.0372903\\
\text{In[4]} &:= \text{Sum}\,[\text{N}\,[\%1]\ //\ \text{Evaluate},\ \{k, 1, 50\}]\\
\text{Out[4]} &= 0.0372903\\
\text{In[5]} &:= \text{N Sum}\,[\%1\ //\ \text{Evaluate},\ \{k, 1,\ \text{Infinity}\}]\\
\text{Out[5]} &= 0.0372903
\end{aligned}$$

Since

$$\zeta\,(0)=-\frac{1}{2},$$

Out[2] evidently validates the series representation (71) *symbolically*. Furthermore, our *numerical* computations in Out[3], Out[4], and Out[5], together, exhibit the fact that only 50 terms ($k=1$ to $k=50$) of the series in (71) can produce an accuracy of as many as *seven* decimal places.

Our symbolic computations and numerical verifications with the series in (72) using *Mathematica* (Version 4.0) for Linux lead us to the following table:

Number of terms	Precision of computation
4	6
10	11
20	18
50	38
98	69

In fact, since the general term of the series in (72) has the following order estimate:

$$O\left(2^{-2k}\cdot k^{-5}\right) \qquad (k \longrightarrow \infty)\,,$$

for getting p *exact* digits, we must have

$$2^{-2k}\cdot k^{-5} < 10^{-p}.$$

Upon solving this inequality *symbolically*, we find that

$$k \cong \frac{5}{\log 4}\text{ProductLog}\left(\frac{10^{p/5}\log 4}{5}\right),$$

where the function ProductLog (also known as *Lambert's function*) is the solution of the equation:

$$xe^{x} = a.$$

Some relevant details about the symbolic computations and numerical verifications with the series in (72) using *Mathematica* (Version 4.0) for Linux are being summarized below.

$$\begin{aligned}
\text{In [1]} &:= \text{expr } = (8576k^{\urcorner}2 + 24286k + 17283)\ \text{Zeta}[2k]/\\
&\qquad ((2k+1)(2k+2)(2k+3)(2k+4)(2k+5)(2k+6)(2k+7)2^{\urcorner}(2k))\\
\text{Out [1]} &= \frac{(17283 + 24286k + 8576k^{2})\ \text{Zeta}[2k]}{2^{2k}(1+2k)(2+2k)(3+2k)(4+2k)(5+2k)(6+2k)(7+2k)}\\
\text{In [2]} &:= \text{Sum[expr}, \{k, 0, \text{infinity}\}]//\ \text{Simplify}\\
\text{Out [2]} &= -\frac{1573}{120\text{Pi}^{2}}\ \text{Zeta}[3]\\
\text{In [3]} &:= \text{N}[-1573/(120\text{Pi}^{\urcorner}2)\ \text{Zeta}[3], 50]\\
&\qquad -\text{Sum[expr}, \{k, 0, 50\}]\\
\text{Out [3]} &= 4.00751120011 \cdot 10^{-38}\\
\text{In [4]} &:= \text{N}[-1573/(120\text{Pi}^{\urcorner}2)\ \text{Zeta}[3], 100]\\
&\qquad -\text{Sum [expr}, \{k, 0, 50\}]\\
\text{Out [4]} &= 4.0075112001 \text{ <skip> } 3481 \cdot 10^{-38}
\end{aligned}$$

Thus, clearly, the result does not change appreciably when we increase the precision of computation of the symbolic result from 50 to 100. This is expected, because of the following numerical computation of the last term for $k = 50$:

$$\begin{aligned}
\text{In [5]} &:= \text{N [expr } /.k \to 50, 50]\\
\text{Out [5]} &= 1.3608530374922376861443887454551514233575702860179 \cdot 10^{-37}
\end{aligned}$$

6 The Hurwitz-Lerch Zeta Function $\Phi(z, s, a)$: Extensions and Generalizations

The potentially and computationally useful foregoing developments (which we have attempted to present here in a rather concise form) have essentially motivated a large number of further investigations on the subject, not only involving the Riemann Zeta function $\zeta(s)$ and the Hurwitz (or generalized) Zeta function $\zeta(s,a)$ (and their such relatives as the multiple Zeta functions and the multiple Gamma functions), but indeed also the substantially general Hurwitz-Lerch Zeta function $\Phi(z,s,a)$ defined by (cf., e.g., [10, p. 27. Eq. 1.11 (1)]; see also [43, p. 121, *et seq*.])

$$\Phi(z,s,a) := \sum_{n=0}^{\infty} \frac{z^n}{(n+a)^s} \tag{73}$$

$$\left(a \in \mathbb{C} \setminus \mathbb{Z}_0^-;\ s \in \mathbb{C} \quad \text{when} \quad |z| < 1;\ \Re(s) > 1 \quad \text{when} \quad |z| = 1\right).$$

Just as in the cases of the Riemann Zeta function $\zeta(s)$ and the Hurwitz (or generalized) Zeta function $\zeta(s,a)$, the Hurwitz-Lerch Zeta function $\Phi(z,s,a)$ can be continued *meromorphically* to the whole complex s-plane, except for a simple pole at $s = 1$ with its residue 1. It is also known that [10, p. 27, Eq. 1.11 (3)]

$$\Phi(z,s,a) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{t^{s-1}\, \mathrm{e}^{-at}}{1 - z\mathrm{e}^{-t}}\, dt = \frac{1}{\Gamma(s)} \int_0^\infty \frac{t^{s-1}\, \mathrm{e}^{-(a-1)t}}{\mathrm{e}^t - z}\, dt \tag{74}$$

$$(\Re(a) > 0;\ \Re(s) > 0 \text{ when } |z| \leqq 1 (z \neq 1); \Re(s) > 1 \quad \text{when} \quad z = 1).$$

The Hurwitz-Lerch Zeta function $\Phi(z,s,a)$ defined by (73) contains, as its *special* cases, not only the Riemann Zeta function $\zeta(s)$ and the Hurwitz (or generalized) Zeta function $\zeta(s,a)$ [cf. (1) and (2)]:

$$\zeta(s) = \Phi(1,s,1) \qquad \text{and} \qquad \zeta(s,a) = \Phi(1,s,a) \tag{75}$$

and the Lerch Zeta function $\ell_s(\xi)$ defined by (see, for details, [10, Chap. I] and [43, Chap. 2])

$$\ell_s(\xi) := \sum_{n=1}^{\infty} \frac{e^{2n\pi i\xi}}{n^s} = e^{2\pi i\xi}\, \Phi\left(e^{2\pi i\xi}, s, 1\right) \tag{76}$$

$$(\xi \in \mathbb{R};\ \Re(s) > 1)\,,$$

but also such other important functions of *Analytic Function Theory* as the Polylogarithmic function (or *Jonquère's function*) $\mathrm{Li}_s(z)$:

$$\mathrm{Li}_s(z) := \sum_{n=1}^{\infty} \frac{z^n}{n^s} = z\, \Phi(z,s,1) \tag{77}$$

$$\big(s \in \mathbb{C} \quad \text{when} \quad |z| < 1;\ \Re(s) > 1 \quad \text{when} \quad |z| = 1\big)$$

and the Lipschitz-Lerch Zeta function (cf. [43, p. 122, Eq. 2.5 (11)]):

$$\phi(\xi, a, s) := \sum_{n=0}^{\infty} \frac{e^{2n\pi i\xi}}{(n+a)^s} = \Phi\left(e^{2\pi i\xi}, s, a\right) =: L\,(\xi, s, a) \tag{78}$$

$$\big(a \in \mathbb{C} \setminus \mathbb{Z}_0^-;\ \Re(s) > 0 \quad \text{when} \quad \xi \in \mathbb{R} \setminus \mathbb{Z};\ \Re(s) > 1 \quad \text{when} \quad \xi \in \mathbb{Z}\big),$$

which was first studied by Rudolf Lipschitz (1832–1903) and Matyáš Lerch (1860–1922) in connection with Dirichlet's famous theorem on primes in arithmetic progressions. For details, the interested reader should be referred, in connection with some of these developments, to the recent works including (among others) [2, 6] to [5, 14, 22, 23, 26].

Yen et al. [59, p. 100, Theorem] derived the following sum-integral representation for the Hurwitz (or generalized) Zeta function $\zeta(s, a)$ defined by (2):

$$\zeta(s, a) = \frac{1}{\Gamma(s)} \sum_{j=0}^{k-1} \int_0^{\infty} \frac{t^{s-1}\, \mathrm{e}^{-(a+j)t}}{1 - \mathrm{e}^{-kt}}\, dt \tag{79}$$

$$\big(k \in \mathbb{N};\ \Re(s) > 1;\ \Re(a) > 0\big),$$

which, for $k = 2$, was given earlier by Nishimoto et al. [31, p. 94, Theorem 4]. A straightforward generalization of the sum-integral representation (79) was given subsequently by Lin and Srivastava [25, p. 727, Eq. 7] in the form:

$$\Phi(z, s, a) = \frac{1}{\Gamma(s)} \sum_{j=0}^{k-1} z^j \int_0^{\infty} \frac{t^{s-1}\, \mathrm{e}^{-(a+j)t}}{1 - z^k \mathrm{e}^{-kt}}\, dt \tag{80}$$

$$\big(k \in \mathbb{N};\ \Re(a) > 0;\ \Re(s) > 0 \ \text{ when } \ |z| \leqq 1\ (z \neq 1);\ \Re(s) > 1 \ \text{ when } \ z = 1\big).$$

Motivated essentially by the sum-integral representations (79) and (80), a generalization of the Hurwitz-Lerch Zeta function $\Phi(z, s, a)$ was introduced and investigated by Lin and Srivastava [25] in the following form [25, p. 727, Eq. 8]:

$$\Phi_{\mu,\nu}^{(\rho,\sigma)}(z, s, a) := \sum_{n=0}^{\infty} \frac{(\mu)_{\rho n}}{(\nu)_{\sigma n}} \frac{z^n}{(n+a)^s} \tag{81}$$

$$\begin{gathered}\big(\mu \in \mathbb{C};\ a, \nu \in \mathbb{C} \setminus \mathbb{Z}_0^-;\ \rho, \sigma \in \mathbb{R}^+;\ \rho < \sigma \quad \text{when} \quad s, z \in \mathbb{C}; \\ \rho = \sigma \quad \text{and} \quad s \in \mathbb{C} \quad \text{when} \quad |z| < \delta := \rho^{-\rho}\, \sigma^{\sigma}; \\ \rho = \sigma \quad \text{and} \quad \Re(s - \mu + \nu) > 1 \quad \text{when} \quad |z| = \delta\big),\end{gathered}$$

where $(\lambda)_\nu$ denotes the Pochhammer symbol defined in conjunction with (11) and (12). Clearly, we find from the definition (81) that

$$\Phi_{\nu,\nu}^{(\sigma,\sigma)}(z,s,a) = \Phi_{\mu,\nu}^{(0,0)}(z,s,a) = \Phi(z,s,a) \tag{82}$$

and

$$\Phi_{\mu,1}^{(1,1)}(z,s,a) = \Phi_{\mu}^{*}(z,s,a) := \sum_{n=0}^{\infty} \frac{(\mu)_n}{n!} \frac{z^n}{(n+a)^s} \tag{83}$$

$$\big(\mu \in \mathbb{C};\ a \in \mathbb{C} \setminus \mathbb{Z}_0^-;\ s \in \mathbb{C} \quad \text{when} \quad |z| < 1;\ \Re(s-\mu) > 1 \quad \text{when} \quad |z| = 1\big),$$

where, as already noted by Lin and Srivastava [25], $\Phi_{\mu}^{*}(z,s,a)$ is a generalization of the Hurwitz-Lerch Zeta function considered by Goyal and Laddha [18, p. 100, Eq. 1.5]. For further results involving these classes of generalized Hurwitz-Lerch Zeta functions, see the recent works by Garg et al. [14] and Lin et al. [26].

A generalization of the above-defined Hurwitz-Lerch Zeta functions $\Phi(z,s,a)$ and $\Phi_{\mu}^{*}(z,s,a)$ was studied by Garg et al. [15] in the following form [15, p. 313, Eq. 1.7]:

$$\Phi_{\lambda,\mu;\nu}(z,s,a) := \sum_{n=0}^{\infty} \frac{(\lambda)_n(\mu)_n}{(\nu)_n \cdot n!} \frac{z^n}{(n+a)^s} \tag{84}$$

$$\big(\lambda, \mu \in \mathbb{C};\ \nu, a \in \mathbb{C} \setminus \mathbb{Z}_0^-;\ s \in \mathbb{C} \quad \text{when} \quad |z| < 1;$$
$$\Re(s+\nu-\lambda-\mu) > 1 \quad \text{when} \quad |z| = 1\big).$$

By comparing the definitions (81) and (83), it is easily observed that the function $\Phi_{\lambda,\mu;\nu}(z,s,a)$ studied by Garg et al. [15] does *not* provide a generalization of the function $\Phi_{\mu,\nu}^{(\rho,\sigma)}(z,s,a)$ which was introduced earlier by Lin and Srivastava [25]. Indeed, for $\lambda = 1$, the function $\Phi_{\lambda,\mu;\nu}(z,s,a)$ coincides with a *special* case of the function $\Phi_{\mu,\nu}^{(\rho,\sigma)}(z,s,a)$ when $\rho = \sigma = 1$.

For the *Riemann-Liouville fractional derivative operator* $\mathcal{D}_z^{\mu}$ defined by (see, for example, [11, p. 181], [35] and [24, p. 70 *et seq.*])

$$\mathcal{D}_z^{\mu}\{f(z)\} := \begin{cases} \dfrac{1}{\Gamma(-\mu)} \displaystyle\int_0^z (z-t)^{-\mu-1} f(t)\,dt \ \big(\Re(\mu) < 0\big) \\[2ex] \dfrac{d^m}{dz^m}\Big\{\mathcal{D}_z^{\mu-m}\{f(z)\}\Big\} \qquad \big(m-1 \leqq \Re(\mu) < m \ \ (m \in \mathbb{N})\big), \end{cases} \tag{85}$$

it is known that

$$\mathcal{D}_z^{\mu}\left\{z^{\lambda}\right\} = \frac{\Gamma(\lambda+1)}{\Gamma(\lambda-\mu+1)} z^{\lambda-\mu} \qquad \big(\Re(\lambda) > -1\big), \tag{86}$$

which, in view of the definition (81), yields the following fractional derivative formula for the generalized Hurwitz-Lerch Zeta function $\Phi_{\mu,\nu}^{(\rho,\sigma)}(z,s,a)$ *with* $\rho = \sigma$ [25, p. 730, Eq. 24]:

$$\mathcal{D}_z^{\mu-\nu}\left\{z^{\mu-1}\,\Phi\left(z^{\sigma},s,a\right)\right\}=\frac{\Gamma\left(\mu\right)}{\Gamma\left(\nu\right)}\,z^{\nu-1}\,\Phi_{\mu,\nu}^{(\sigma,\sigma)}\left(z^{\sigma},s,a\right) \tag{87}$$

$$\left(\Re\left(\mu\right)>0;\ \sigma\in\mathbb{R}^{+}\right).$$

In particular, when $\nu=\sigma=1$, the fractional derivative formula (87) would reduce at once to the following form:

$$\Phi_{\mu}^{*}\left(z,s,a\right)=\frac{1}{\Gamma\left(\mu\right)}\,\mathcal{D}_z^{\mu-1}\left\{z^{\mu-1}\,\Phi\left(z,s,a\right)\right\}\qquad\left(\Re(\mu)>0\right), \tag{88}$$

which (as already remarked by Lin and Srivastava [25, p. 730]) exhibits the interesting (and useful) fact that $\Phi_{\mu}^{*}(z,s,a)$ is essentially a Riemann-Liouville fractional derivative of the classical Hurwitz-Lerch function $\Phi\left(z,s,a\right)$. Moreover, it is easily deduced from the fractional derivative formula (86) that

$$\begin{aligned}\Phi_{\lambda,\mu;\nu}(z,s,a)&=\frac{\Gamma(\nu)}{\Gamma(\lambda)}\,z^{1-\lambda}\,\mathcal{D}_z^{\lambda-\nu}\left\{z^{\lambda-1}\,\Phi_{\mu}^{*}(z,s,a)\right\}\\&=\frac{\Gamma(\nu)}{\Gamma(\lambda)\Gamma(\mu)}\,z^{1-\lambda}\\&\qquad\cdot\mathcal{D}_z^{\lambda-\nu}\left\{z^{\lambda-1}\,\mathcal{D}_z^{\mu-1}\left\{z^{\mu-1}\,\Phi_{\mu}(z,s,a)\right\}\right\},\end{aligned} \tag{89}$$

which exhibits the *hitherto unnoticed* fact that the function $\Phi_{\lambda,\mu;\nu}(z,s,a)$ studied by Garg et al. [15] is essentially a consequence of the classical Hurwitz-Lerch Zeta function $\Phi(z,s,a)$ when we apply the Riemann-Liouville fractional derivative operator $\mathcal{D}_z^{\mu}$ *two times* as indicated above (see also [53]). Many other explicit representations for $\Phi_{\mu}^{*}(z,s,a)$ and $\Phi_{\mu,\nu}^{(\rho,\sigma)}(z,s,a)$, including a potentially useful Eulerian integral representation of the first kind [25, p. 731, Eq. 28], were proven by Lin and Srivastava [25].

A multiple (or, simply, n-dimentional) Hurwitz-Lerch Zeta function $\Phi_n(z,s,a)$ was studied recently by Choi et al. [7, p. 66, Eq. 6]. Răducanu and Srivastava (see [33] and the references cited therein), on the other hand, made use of the Hurwitz-Lerch Zeta function $\Phi(z,s,a)$ in defining a certain linear convolution operator in their systematic investigation of various analytic function classes in *Geometric Function Theory* in *Complex Analysis*. Furthermore, Gupta et al. [19] revisited the study of the familiar Hurwitz-Lerch Zeta distribution by investigating its structural properties, reliability properties and statistical inference. These investigations by Gupta et al. [19] and others (see, for example, [42, 43, 51, 52]), fruitfully using the Hurwitz-Lerch Zeta function $\Phi(z,s,a)$ and some of its above-mentioned generalizations, motivated Srivastava et al. [53] to present a further generalization and analogous investigation of a new family of Hurwitz-Lerch Zeta functions defined in the following form [53, p. 491, Eq. 1.20]:

$$\Phi_{\lambda,\mu;\nu}^{(\rho,\sigma,\kappa)}(z,s,a) := \sum_{n=0}^{\infty} \frac{(\lambda)_{\rho n}(\mu)_{\sigma n}}{(\nu)_{\kappa n} \cdot n!} \frac{z^n}{(n+a)^s} \tag{90}$$

$$\big(\lambda, \mu \in \mathbb{C};\; a, \nu \in \mathbb{C} \setminus \mathbb{Z}_0^-;\; \rho, \sigma, \kappa \in \mathbb{R}^+;\; \kappa - \rho - \sigma > -1 \quad \text{when} \quad s, z \in \mathbb{C};$$
$$\kappa - \rho - \sigma = -1 \quad \text{and} \quad s \in \mathbb{C} \quad \text{when} \quad |z| < \delta^* := \rho^{-\rho}\, \sigma^{-\sigma}\, \kappa^{\kappa};$$
$$\kappa - \rho - \sigma = -1 \quad \text{and} \quad \Re(s + \nu - \lambda - \mu) > 1 \quad \text{when} \quad |z| = \delta^*\big).$$

For the above-defined function in (90), Srivastava et al. [53] established various integral representations, relationships with the $\overline{H}$-function which is defined by means of a Mellin-Barnes type contour integral (see, for example, [50, 53]), fractional derivative and analytic continuation formulas, as well as an extension of the generalized Hurwitz-Lerch Zeta function $\Phi_{\lambda,\mu;\nu}^{(\rho,\sigma,\kappa)}(z,s,a)$ in (90). This natural further extension and generalization of the function $\Phi_{\lambda,\mu;\nu}^{(\rho,\sigma,\kappa)}(z,s,a)$ was indeed accomplished by introducing essentially arbirary numbers of numerator and denominator parameters in the definition (90). For this purpose, in addition to the symbol ∇^* defined by

$$\nabla^* := \left(\prod_{j=1}^{p} \rho_j^{-\rho_j}\right) \cdot \left(\prod_{j=1}^{q} \sigma_j^{\sigma_j}\right), \tag{91}$$

the following notations will be employed:

$$\Delta := \sum_{j=1}^{q} \sigma_j - \sum_{j=1}^{p} \rho_j \qquad \text{and} \qquad \Xi := s + \sum_{j=1}^{q} \mu_j - \sum_{j=1}^{p} \lambda_j + \frac{p-q}{2}. \tag{92}$$

Then the extended Hurwitz-Lerch Zeta function

$$\Phi_{\lambda_1,\cdots,\lambda_p;\mu_1,\cdots,\mu_q}^{(\rho_1,\cdots,\rho_p,\sigma_1,\cdots,\sigma_q)}(z,s,a)$$

is defined by [53, p. 503, Eq. 6.2]

$$\Phi_{\lambda_1,\cdots,\lambda_p;\mu_1,\cdots,\mu_q}^{(\rho_1,\cdots,\rho_p,\sigma_1,\cdots,\sigma_q)}(z,s,a) := \sum_{n=0}^{\infty} \frac{\prod_{j=1}^{p} (\lambda_j)_{n\rho_j}}{n! \prod_{j=1}^{q} (\mu_j)_{n\sigma_j}} \frac{z^n}{(n+a)^s} \tag{93}$$

$$\Bigg(p, q \in \mathbb{N}_0;\;\; \lambda_j \in \mathbb{C} \;\; (j = 1, \cdots, p);\;\; a, \mu_j \in \mathbb{C} \setminus Z_0^- \;\; (j = 1, \cdots, q);$$
$$\rho_j, \sigma_k \in \mathbb{R}^+ \;\; (j = 1, \cdots, p;\; k = 1, \cdots, q);$$
$$\Delta > -1 \quad \text{when} \quad s, z \in \mathbb{C};\;\; \Delta = -1 \quad \text{and} \quad s \in \mathbb{C} \quad \text{when} \quad |z| < \nabla^*;$$
$$\Delta = -1 \quad \text{and} \quad \Re(\Xi) > \frac{1}{2} \quad \text{when} \quad |z| = \nabla^*\Bigg).$$

The special case of the definition (93) when $p-1=q=1$ would obviously correspond to the above-investigated generalized Hurwitz-Lerch Zeta function $\Phi_{\lambda,\mu;\nu}^{(\rho,\sigma,\kappa)}(z,s,a)$ defined by (90).

If we set

$$p \mapsto p+1 \qquad \big(\rho_1=\cdots=\rho_p=1;\ \ \lambda_{p+1}=\rho_{p+1}=1\big)$$

and

$$q \mapsto q+1 \qquad \big(\sigma_1=\cdots=\sigma_q=1;\ \ \mu_{q+1}=\beta;\ \ \sigma_{q+1}=\alpha\big),$$

then (93) reduces to the following generalized M-series which was recently introduced by Sharma and Jain [36] as follows:

$$\begin{aligned}
&{}_p\overset{\alpha,\beta}{M}_q(a_1,\cdots,a_p;b_1,\cdots,b_q;z)\\
&\qquad=\sum_{k=0}^{\infty}\frac{(a_1)_k\cdots(a_p)_k}{(b_1)_k\cdots(b_q)_k}\,\frac{z^k}{\Gamma(\alpha k+\beta)}\\
&\qquad=\frac{1}{\Gamma(\beta)}\ {}_{p+1}\Psi_{q+1}^{*}\left[\begin{matrix}(a_1,1),\ \cdots,\big(a_p,1\big),(1,1);\\ \\(b_1,1),\ \cdots,\big(b_q,1\big),(\beta,\alpha);\end{matrix}\ z\right].
\end{aligned}\tag{94}$$

This last relationship (94) exhibits the fact that the so-called generalized M-series is, in fact, an *obvious* variant of the Fox-Wright function ${}_p\Psi_q^*$ or ${}_p\Psi_q^*$ $(p,q\in\mathbb{N}_0)$, which is a generalization of the familiar generalized hypergeometric function ${}_pF_q(p,q\in\mathbb{N}_0)$, with p numerator parameters $a_1,\ \cdots,a_p$ and q denominator parameters $b_1,\ \cdots,b_q$ such that

$$a_j\in\mathbb{C}\quad(j=1,\ \cdots,p)\qquad\text{and}\qquad b_j\in\mathbb{C}\setminus\mathbb{Z}_0^-\quad(j=1,\ \cdots,q),$$

defined by (see, for details, [10, p. 183] and [44, p. 21]; see also [24, p. 56], [28, p. 30] and [48, p. 19])

$$\begin{aligned}
&{}_p\Psi_q^{*}\left[\begin{matrix}(a_1,A_1),\ \cdots,\big(a_p,A_p\big);\\ \\(b_1,B_1),\ \cdots,\big(b_q,B_q\big);\end{matrix}\ z\right]\\
&\qquad:=\sum_{n=0}^{\infty}\frac{(a_1)_{A_1n}\ \cdots\big(a_p\big)_{A_pn}}{(b_1)_{B_1n}\ \cdots\big(b_q\big)_{B_qn}}\,\frac{z^n}{n!}\\
&\qquad=\frac{\Gamma\,(b_1)\ \cdots\Gamma\,\big(b_q\big)}{\Gamma\,(a_1)\ \cdots\Gamma\,\big(a_p\big)}\ {}_p\Psi_q\left[\begin{matrix}(a_1,A_1),\ \cdots,\big(a_p,A_p\big);\\ \\(b_1,B_1),\ \cdots,\big(b_q,B_q\big);\end{matrix}\ z\right]
\end{aligned}\tag{95}$$

$$\left(A_j > 0 \quad (j = 1, \cdots, p);\; B_j > 0 \quad (j = 1, \cdots, q);\; 1 + \sum_{j=1}^{q} B_j - \sum_{j=1}^{p} A_j \geqq 0\right),$$

where the equality in the convergence condition holds true for suitably bounded values of $|z|$ given by

$$|z| < \nabla := \left(\prod_{j=1}^{p} A_j^{-A_j}\right) \cdot \left(\prod_{j=1}^{q} B_j^{B_j}\right).$$

In the particular case when

$$A_j = B_k = 1 \qquad (j = 1, \cdots, p;\;\; k = 1, \cdots, q),$$

we have the following relationship (see, for details, [44, p. 21]):

$$\begin{aligned} &{}_p\Psi_q^* \left[\begin{matrix} (a_1, 1), \cdots, (a_p, 1); \\ \\ (b_1, 1), \cdots, (b_q, 1); \end{matrix}\; z\right] \\ &\quad = {}_pF_q \left[\begin{matrix} a_1, \cdots, a_p; \\ \\ b_1, \cdots, b_q; \end{matrix}\; z\right] \\ &\quad = \frac{\Gamma(b_1) \cdots \Gamma(b_q)}{\Gamma(a_1) \cdots \Gamma(a_p)}\; {}_p\Psi_q \left[\begin{matrix} (a_1, 1), \cdots, (a_p, 1); \\ \\ (b_1, 1), \cdots, (b_q, 1); \end{matrix}\; z\right], \end{aligned} \tag{96}$$

in terms of the generalized hypergeometric function ${}_pF_q \quad (p, q \in \mathbb{N}_0)$.

Each of the following results involving the extended Hurwitz-Lerch Zeta function

$$\Phi^{(\rho_1, \cdots, \rho_p, \sigma_1, \cdots, \sigma_q)}_{\lambda_1, \cdots, \lambda_p; \mu_1, \cdots, \mu_q}(z, s, a)$$

can be proven by applying the definition (93) in precisely the same manner as for the corresponding result involving the general Hurwitz-Lerch Zeta function $\Phi^{(\rho,\sigma,\kappa)}_{\lambda,\mu;\nu}(z, s, a)$ (see, for details, [53, Sect. 6]).

$$\begin{aligned} &\Phi^{(\rho_1, \cdots, \rho_p, \sigma_1, \cdots, \sigma_q)}_{\lambda_1, \cdots, \lambda_p; \mu_1, \cdots, \mu_q}(z, s, a) \\ &\quad = \frac{1}{\Gamma(s)} \int_0^\infty t^{s-1}\, \mathrm{e}^{-at}\; {}_p\Psi_q^* \left[\begin{matrix} (\lambda_1, \rho_1), \cdots, (\lambda_p, \rho_p); \\ \\ (\mu_1, \sigma_1), \cdots, (\mu_q, \sigma_q); \end{matrix}\; z\mathrm{e}^{-t}\right] dt \end{aligned} \tag{97}$$

$$\left(\min\{\Re(a), \Re(s)\} > 0\right),$$

$$\Phi_{\lambda_1,\cdots,\lambda_p;\mu_1,\cdots,\mu_q}^{(\rho_1,\cdots,\rho_p,\sigma_1,\cdots,\sigma_q)}(z,s,a) = \frac{\prod\limits_{j=1}^{q} \Gamma\left(\mu_j\right)}{\prod\limits_{j=1}^{p} \Gamma\left(\lambda_j\right)}$$

$$\cdot \frac{1}{2\pi i} \int_{\mathfrak{L}} \frac{\Gamma(-\xi)\left\{\Gamma(\xi+a)\right\}^s \prod\limits_{j=1}^{p} \Gamma\left(\lambda_j+\rho_j\xi\right)}{\left\{\Gamma(\xi+a+1)\right\}^s \prod\limits_{j=1}^{q} \Gamma\left(\mu_j+\sigma_j\xi\right)} (-z)^{\xi}\, d\xi \tag{98}$$

$$\left(|\arg(-z)| < \pi\right)$$

or, equivalently,

$$\Phi_{\lambda_1,\cdots,\lambda_p;\mu_1,\cdots,\mu_q}^{(\rho_1,\cdots,\rho_p,\sigma_1,\cdots,\sigma_q)}(z,s,a) = \frac{\prod\limits_{j=1}^{q} \Gamma\left(\mu_j\right)}{\prod\limits_{j=1}^{p} \Gamma\left(\lambda_j\right)}$$

$$\cdot\, \overline{H}_{p+1,q+2}^{1,p+1}\left[-z \,\middle|\, \begin{matrix} (1-\lambda_1,\rho_1;1),\cdots,(1-\lambda_p,\rho_p;1),(1-a,1;s) \\ \\ (0,1),(1-\mu_1,\sigma_1;1),\cdots,(1-\mu_q,\sigma_q;1),(-a,1;s) \end{matrix}\right], \tag{99}$$

provided that both sides of the assertions (97)–(99) exist, the path of integration $\mathfrak{L}$ in (99) being a Mellin-Barnes type contour in the complex ξ-plane, which starts at the point $-i\infty$ and terminates at the point $i\infty$ with indentations, if necessary, in such a manner as to separate the poles of $\Gamma(-\xi)$ from the poles of $\Gamma\left(\lambda_j+\rho_j\xi\right)$ $(j=1,\cdots,p)$.

The $\overline{H}$-function representation given by (99) can be applied in order to derive various properties of the extended Hurwitz-Lerch Zeta function

$$\Phi_{\lambda_1,\cdots,\lambda_p;\mu_1,\cdots,\mu_q}^{(\rho_1,\cdots,\rho_p,\sigma_1,\cdots,\sigma_q)}(z,s,a)$$

from those of the $\overline{H}$-function. Thus, for example, by making use of the following fractional-calculus result due to Srivastava et al. [50, p. 97, Eq. 2.4]:

$$\mathcal{D}_z^{\nu}\left\{z^{\lambda-1}\, \overline{H}_{p,q}^{m,n}(\omega z^{\kappa})\right\} = z^{\lambda-\nu-1}$$

$$\cdot\, \overline{H}_{p+1,q+1}^{m,n+1}\left[\omega z^{\kappa} \,\middle|\, \begin{matrix} (1-\lambda,\kappa;1),\left(a_j,A_j;\alpha_j\right)_{j=1}^{n},\left(a_j,A_j\right)_{j=n+1}^{p} \\ \\ \left(b_j,B_j\right)_{j=1}^{m},\left(b_j,B_j;\beta_j\right)_{j=m+1}^{q},(1-\lambda+\nu,\kappa;1) \end{matrix}\right] \tag{100}$$

$$\left(\Re(\lambda) > 0;\ \kappa > 0\right),$$

we readily obtain an extension of such fractional derivative formulas as (for example) (87) given by

$$\mathcal{D}_z^{\nu-\tau}\left\{z^{\nu-1}\,\Phi^{(\rho_1,\cdots,\rho_p,\sigma_1,\cdots,\sigma_q)}_{\lambda_1,\cdots,\lambda_p;\mu_1,\cdots,\mu_q}(z^\kappa,s,a)\right\}=\frac{\prod\limits_{j=1}^{q}\Gamma(\mu_j)}{\prod\limits_{j=1}^{p}\Gamma(\lambda_j)}\,z^{\tau-1}$$

$$\cdot\,\overline{H}^{1,p+2}_{p+2,q+3}\left[-z^\kappa\;\middle|\;\begin{matrix}(1-\lambda_1,\rho_1;1),\cdots,(1-\lambda_p,\rho_p;1),(1-\nu,\kappa;1),(1-a,1;s)\\ \\ (0,1),(1-\mu_1,\sigma_1;1),\cdots,(1-\mu_q,\sigma_q;1),(1-\tau,\kappa;1),(-a,1;s)\end{matrix}\right]$$

$$=\frac{\Gamma(\nu)}{\Gamma(\tau)}\,z^{\tau-1}\,\Phi^{(\rho_1,\cdots,\rho_p,\kappa,\sigma_1,\cdots,\sigma_q,\kappa)}_{\lambda_1,\cdots,\lambda_p,\nu;\mu_1,\cdots,\mu_q,\tau}(z^\kappa,s,a)\qquad\big(\Re(\nu)>0;\;\kappa>0\big).\qquad(101)$$

Finally, we present the following extension of a known result [53, p. 496, Theorem 3] (see also [53, p. 505, Theorem 9].

Theorem 1. *Let $\big(\alpha_n\big)_{n\in\mathbb{N}_0}$ be a positive sequence such that the following infinite series:*

$$\sum_{n=0}^{\infty}e^{-\alpha_n t}$$

converges for any $t\in\mathbb{R}^+$. Then

$$\Phi^{(\rho_1,\cdots,\rho_p,\sigma_1,\cdots,\sigma_q)}_{\lambda_1,\cdots,\lambda_p;\mu_1,\cdots,\mu_q}(z,s,a)=\frac{1}{\Gamma(s)}\sum_{n=0}^{\infty}\int_0^\infty t^{s-1}\,e^{-(a-\alpha_0+\alpha_n)t}\,\left(1-e^{-(\alpha_{n+1}-\alpha_n)t}\right)$$

$$\cdot\,{}_p\Psi_q^*\left[\begin{matrix}(\lambda_1,\rho_1),\cdots,(\lambda_p,\rho_p);\\ \\ (\mu_1,\sigma_1),\cdots,(\mu_q,\sigma_q);\end{matrix}\;ze^{-t}\right]dt\qquad(102)$$

$$\big(\min\{\Re(a),\Re(s)\}>0\big),$$

provided that each member of (102) exists.

It would be nice and worthwhile to be able to extend the results presented in Sects. 2–5 of this article to hold true for the Hurwitz-Lerch Zeta function $\Phi\,(z,s,a)$ and for some of its generalizations given by the Lin-Srivastava Zeta function $\Phi^{(\rho,\sigma)}_{\mu,\nu}\,(z,s,a)$ and the extended Hurwitz-Lerch Zeta function

$$\Phi^{(\rho_1,\cdots,\rho_p,\sigma_1,\cdots,\sigma_q)}_{\lambda_1,\cdots,\lambda_p;\mu_1,\cdots,\mu_q}(z,s,a)$$

defined by (93) for special values of the varous parameters involved in the definitions (81) and (93).

Acknowledgements It is a great pleasure for me to dedicate this presentation to Prof. Dr. Stephen Smale on the occasion of his eightieth birthday. The present investigation was supported, in part, by the *Natural Sciences and Engineering Research Council of Canada* under Grant OGP0007353.

References

1. M. Abramowitz, I.A. Stegun (eds.), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Applied Mathematics Series, vol. 55 (National Bureau of Standards, Washington, 1964); Reprinted by Dover Publications, New York, 1965 (see also [32])
2. H. Alzer, D. Karayannakis, H.M. Srivastava, Series representations for some mathematical constants. J. Math. Anal. Appl. **320**, 145–162 (2006)
3. R. Apéry, Irrationalité de $\zeta(2)$ et $\zeta(3)$, in *Journées Arithmétiques de Luminy* (Colloq. Internat. CNRS, Centre University Luminy, Luminy, 1978), pp. 11–13, *Astérisque* **61**, Society Mathematical, France, Paris, 1979.
4. M.-P. Chen, H.M. Srivastava, Some families of series representations for the Riemann $\zeta(3)$. Result. Math. **33**, 179–197 (1998)
5. J. Choi, H.M. Srivastava, V.S. Adamchik, Multiple Gamma and related functions. Appl. Math. Comput. **134**, 515–533 (2003)
6. J. Choi, Y.J. Cho, H.M. Srivastava, Series involving the Zeta and multiple Gamma functions. Appl. Math. Comput. **159**, 509–537 (2004)
7. J. Choi, D.S. Jang, H.M. Srivastava, A generalization of the Hurwitz-Lerch Zeta function. Integral Transform. Spec. Funct. **19**, 65–79 (2008)
8. D. Cvijović, J. Klinowski, New rapidly convergent series representations for $\zeta(2n+1)$. Proc. Am. Math. Soc. **125**, 1263–1271 (1997)
9. A. Dąbrowski, A note on the values of the Riemann Zeta function at positive odd integers. Nieuw Arch. Wiskd. **14**(4), 199–207 (1996)
10. A. Erdélyi, W. Magnus, F. Oberhettinger, F. G. Tricomi, *Higher Transcendental Functions*, vol. I (McGraw-Hill Book Company, New York/Toronto/London, 1953)
11. A. Erdélyi, W. Magnus, F. Oberhettinger, F.G. Tricomi, *Tables of Integral Transforms*, vol. II (McGraw-Hill Book Company, New York/Toronto/London, 1954)
12. J.A. Ewell, A new series representation for $\zeta(3)$, Am. Math. Mon. **97**, 219–220 (1990)
13. J.A. Ewell, On the Zeta function values $\zeta(2k+1),\ k = 1, 2, \cdots$. Rocky Mt. J. Math. **25**, 1003–1012 (1995)
14. M. Garg, K. Jain, H.M. Srivastava, Some relationships between the generalized Apostol-Bernoulli polynomials and Hurwitz-Lerch Zeta functions. Integral Transform Spec. Funct. **17**, 803–815 (2006)
15. M. Garg, K. Jain, S.L. Kalla, A further study of general Hurwitz-Lerch zeta function. Algebras Groups Geom. **25**, 311–319 (2008)
16. M.L. Glasser, Some integrals of the arctangent function. Math. Comput. **22**, 445–447 (1968)
17. R.W. Gosper Jr., A calculus of series rearrangements, in *Algorithms and Complexity: New Directions and Recent Results*, ed. by J.F. Traub (Academic, New York/London/Toronto, 1976), pp. 121–151
18. S.P. Goyal, R.K. Laddha, On the generalized Zeta function and the generalized Lambert function. Gadṇita Sandesh **11**, 99–108 (1997)
19. P.L. Gupta, R.C. Gupta, S.-H. Ong, H.M. Srivastava, A class of Hurwitz-Lerch Zeta distributions and their applications in reliability. Appl. Math. Comput. **196**, 521–531 (2008)
20. E.R. Hansen, *A Table of Series and Products* (Prentice-Hall, Englewood Cliffs, 1975)
21. M.M. Hjortnaes, Overføring av rekken $\sum_{k=1}^{\infty}\left(1/k^3\right)$ til et bestemt integral, in *Proceedings of the Twelfth Scandanavian Mathematical Congress*, Lund, 10–15 Aug 1953 (Scandanavian Mathematical Society, Lund, 1954), pp. 211–213

22. S. Kanemitsu, H. Kumagai, M. Yoshimoto, Sums involving the Hurwitz Zeta functions. Ramanujan J. **5**, 5–19 (2001)
23. S. Kanemitsu, H. Kumagai, H.M. Srivastava, M. Yoshimoto, Some integral and asymptotic formulas associated with the Hurwitz Zeta function. Appl. Math. Comput. **154**, 641–664 (2004)
24. A.A. Kilbas, H.M. Srivastava, J.J. Trujillo, *Theory and Applications of Fractional Differential Equations*. North-Holland Mathematical Studies, vol. 204 (Elsevier (North-Holland) Science Publishers, Amsterdam/London/New York, 2006)
25. S.-D. Lin, H.M. Srivastava, Some families of the Hurwitz–Lerch Zeta functions and associated fractional derivative and other integral representations. Appl. Math. Comput. **154**, 725–733 (2004)
26. S.-D. Lin, H.M. Srivastava, P.-Y. Wang, Some expansion formulas for a class of generalized Hurwitz-Lerch Zeta functions. Integral Transform Spec. Funct. **17**, 817–827 (2006)
27. W. Magnus, F. Oberhettinger, R.P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*, Third Enlarged Edition, Die Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtingung der Anwendungsgebiete, Bd. **52** (Springer, New York, 1966)
28. A.M. Mathai, R.K. Saxena, H.J. Haubold, *The H-Function*: *Theory and Applications* (Springer, New York/Dordrecht/Heidelberg/London, 2010)
29. C. Nash, D. O'Connor, Ray-Singer torsion, topological field theories and the Riemann Zeta function at $s = 3$, in *Low-Dimensional Topology and Quantum Field Theory* (Proceedings of a NATO Advanced Research Workshop held at the Isaac Newton Institute at Cambridge, September 6–12, 1992), ed. by H. Osborn (Plenum Press, New York/London, 1993), pp. 279–288
30. C. Nash, D. O'Connor, Determinants of Laplacians, the Ray-Singer torsion on lens spaces and the Riemann Zeta function. J. Math. Phys. **36**, 1462–1505 (1995)
31. K. Nishimoto, C.-E Yen, M.-L. Lin, Some integral forms for a generalized Zeta function. J. Fract. Calc. **22**, 91–97 (2002)
32. F.W.J. Olver, D.W. Lozier, R.F. Boisvert, C.W. Clark (eds.), *NIST Handbook of Mathematical Functions* [With 1 CD-ROM (Windows, Macintosh and UNIX)] (U.S. Department of Commerce, National Institute of Standards and Technology, Washington, DC., 2010; Cambridge University Press, Cambridge/London/New York, 2010) (see also [1])
33. D. Răducanu, H.M. Srivastava, A new class of analytic functions defined by means of a convolution operator involving the Hurwitz-Lerch Zeta function. Integral Transform Spec. Funct. **18**, 933–943 (2007)
34. V. Ramaswami, Notes on Riemann's ζ-function. J. Lond. Math. Soc. **9**, 165–169 (1934)
35. S.G. Samko, A.A. Kilbas, O.I. Marichev, *Fractional Integrals and Derivatives*: *Theory and Applications*, Translated from the Russian: *Integrals and Derivatives of Fractional Order and Some of Their Applications* ("Nauka i Tekhnika", Minsk, 1987) (Gordon and Breach Science Publishers, Reading/Tokyo/Paris/Berlin/Langhorne (Pennsylvania), 1993)
36. M. Sharma, R. Jain, A note on a generalzed M-series as a special function of fractional calculus. Fract. Calc. Appl. Anal. **12**, 449–452 (2009)
37. H.M. Srivastava, A unified presentation of certain classes of series of the Riemann Zeta function. Riv. Mat. Univ. Parma (Ser. 4) **14**, 1–23 (1988)
38. H.M. Srivastava, Sums of certain series of the Riemann Zeta function. J. Math. Anal. Appl. **134**, 129–140 (1988)
39. H.M. Srivastava, Further series representations for $\zeta(2n + 1)$, Appl. Math. Comput. **97**, 1–15 (1998)
40. H.M. Srivastava, Some rapidly converging series for $\zeta(2n + 1)$. Proc. Am. Math. Soc. **127**, 385–396 (1999)
41. H.M. Srivastava, Some simple algorithms for the evaluations and representations of the Riemann Zeta function at positive integer arguments. J. Math. Anal. Appl. **246**, 331–351 (2000)
42. H.M. Srivastava, Some formulas for the Bernoulli and Euler polynomials at rational arguments. Math. Proc. Camb. Philos. Soc. **129**, 77–84 (2000)

43. H.M. Srivastava, J. Choi, *Series Associated with the Zeta and Related Functions* (Kluwer, Dordrecht/Boston/London, 2001)
44. H.M. Srivastava, P.W. Karlsson, *Multiple Gaussian Hypergeometric Series* (Halsted Press (Ellis Horwood Limited, Chichester), Wiley, New York/Chichester/Brisbane/Toronto, 1985)
45. H.M. Srivastava, H. Tsumura, A certain class of rapidly convergent series representations for $\zeta(2n+1)$. J. Comput. Appl. Math. **118**, 323–335 (2000)
46. H.M. Srivastava, H. Tsumura, New rapidly convergent series representations for $\zeta(2n+1)$, $L(2n,\chi)$ and $L(2n+1,\chi)$. Math. Sci. Res. Hot-Line **4**(7), 17–24 (2000) (Research Announcement)
47. H.M. Srivastava, H. Tsumura, Inductive construction of rapidly convergent series representations for $\zeta(2n+1)$. Int. J. Comput. Math. **80**, 1161–1173 (2003)
48. H.M. Srivastava, K.C. Gupta, S.P. Goyal, *The H-Functions of One and Two Variables with Applications* (South Asian Publishers, New Delhi/Madras, 1982)
49. H.M. Srivastava, M.L. Glasser, V.S. Adamchik, Some definite integrals associated with the Riemann Zeta function. Z. Anal. Anwend. **19**, 831–846 (2000)
50. H.M. Srivastava, S.-D. Lin, P.-Y. Wang, Some fractional-calculus results for the $\overline{H}$-function associated with a class of Feynman integrals. Russ. J. Math. Phys. **13**, 94–100 (2006)
51. H.M. Srivastava, M. Garg, S. Choudhary, A new generalization of the Bernoulli and related polynomials. Russ. J. Math. Phys. **17**, 251–261 (2010)
52. H.M. Srivastava, M. Garg, S. Choudhary, Some new families of generalized Euler and Genocchi polynomials. Taiwan. J. Math. **15**, 283–305 (2011)
53. H.M. Srivastava, R.K. Saxena, T.K. Pogány, R. Saxena, Integral and computational representations of the extended Hurwitz–Lerch Zeta function. Integral Transform Spec. Funct. **22**, 487–506 (2011)
54. E.C. Titchmarsh, *The Theory of the Riemann Zeta-Function* (Oxford University (Clarendon) Press, Oxford/London, 1951); Second Edition (Revised by D. R. Heath-Brown), 1986
55. F.G. Tricomi, Sulla somma delle inverse delle terze e quinte potenze dei numeri naturali. Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (Ser. 8) **47**, 16–18 (1969)
56. H. Tsumura, On evaluation of the Dirichlet series at positive integers by q-calculation. J. Number Theory **48**, 383–391 (1994)
57. E.T. Whittaker, G.N. Watson, *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, Cambridge/London/New York, 1927)
58. E. Witten, On quantum gauge theories in two dimensions. Commun. Math. Phys. **141**, 153–209 (1991)
59. C.-E Yen, M.-L. Lin, K. Nishimoto, An integral form for a generalized Zeta function. J. Fract. Calc. **22**, 99–102 (2002)
60. N.-Y. Zhang, K.S. Williams, Some series representations of $\zeta(2n+1)$. Rocky Mt. J. Math. **23**, 1581–1592 (1993)
61. N.-Y. Zhang, K.S. Williams, Some infinite series involving the Riemann Zeta function, in *Analysis, Geometry and Groups: A Riemann Legacy Volume, Parts I and II* ed. by H.M. Srivastava, Th. M. Rassias, Part II, (Hadronic Press, Palm Harbor, 1993), pp. 691–712
62. N.-Y. Zhang, K.S. Williams, Values of the Riemann Zeta function and integrals involving $\log(2\sinh\frac{\theta}{2})$ and $\log(2\sin\frac{\theta}{2})$. Pac. J. Math. **168**, 271–289 (1995)

Gyrations: The Missing Link Between Classical Mechanics with Its Underlying Euclidean Geometry and Relativistic Mechanics with Its Underlying Hyperbolic Geometry

Abraham Albert Ungar

Dedicated to the 80th Anniversary of Professor Stephen Smale

Abstract The present article on the hyperbolic geometric interpretation of the relativistic mechanical effect known as Thomas precession is dedicated to the 80th Anniversary of Steve Smale for his leadership and commitment to excellence in the field of geometric mechanics. A study of Thomas precession in terms of its underlying hyperbolic geometry and elegant algebra is presented here in order to clarify the concept of Thomas precession. We review the studies of both Thomas precession and its abstract version, gyration. Based on the review we derive the correct Thomas precession angular velocity. We demonstrate here convincingly that the Thomas precession angle ϵ and its generating angle θ have opposite signs. We present the path from Einstein velocity addition to the gyroalgebra of gyrogroups and gyrations, and to the gyrogeometry that coincides with the hyperbolic geometry of Bolyai and Lobachevsky. We, then, demonstrate that the concept of Thomas precession in Einstein's special theory of relativity is a concrete realization of the abstract concept of gyration in gyroalgebra.

1 Introduction

It has been the lifetime desire of Steve Smale to improve our understanding of geometric mechanics [36], a theory that fits extraordinarily well into dynamical systems framework, as he explains in his 1967 survey article [35]. An overview of his involvement with geometric mechanics, without entering into technical details, is presented by Marsden in [25]. The impact of Einstein's work [29] led

A.A. Ungar (✉)
Department of Mathematics, North Dakota State University, Fargo, ND 58108, USA
e-mail: Ungar.Abraham@ndsu.edu

P.M. Pardalos and T.M. Rassias (eds.), *Essays in Mathematics and its Applications*,
DOI 10.1007/978-3-642-28821-0_18, © Springer-Verlag Berlin Heidelberg 2012

Smale to remark in [36, p. 365] that relativity theory respects classical mechanics since "Einstein worked from a very deep understanding of Newtonian theory". The present article on the hyperbolic geometric interpretation of the relativistic mechanical effect known as Thomas precession is therefore dedicated to the 80th Anniversary of Steve Smale for his leadership and commitment to excellence in the field of geometric mechanics.

Thomas precession of Einstein's special theory of relativity is a physical realization of the abstract gyration. The latter, in turn, is an automorphism (defined in Definition 4) that provides the missing link between Einstein's special theory of relativity and the hyperbolic geometry of Bolyai and Lobachevsky. Named after Llewellyn Hilleth Thomas (1902–1992) who discovered its physical significance in 1926 [1, 41, 42], Thomas precession is a special relativistic kinematic effect that regulates Einstein velocity addition both algebraically and geometrically [52]. In an exhaustive review of the vast literature on Thomas precession [24], G.B. Malykin emphasizes the importance of the frequency of the precession, pinpointing related erroneous results that are common in the literature.

Accordingly, a study of Thomas precession in terms of its underlying hyperbolic geometry and elegant algebra is presented here in order to clarify the concept of Thomas precession. Thomas precession is an important special relativistic rotation that results from the nonassociativity of Einstein velocity addition and, hence, does not exist classically. Indeed, it was discovered in 1988 [45] that Thomas precession regulates Einstein velocity addition, endowing it with a rich algebraic structure. As such, Thomas precession admits extension by abstraction, in which *precession* becomes *gyration*. The latter, in turn, gives rise to two new algebraic structures called a gyrogroup and a gyrovector space, thus introducing new realms to explore. The basic importance of gyrations is emphasized in gyrolanguage, where we prefix a gyro to any term that describes a concept in Euclidean geometry and in associative algebra to mean the analogous concept in hyperbolic geometry and nonassociative algebra.

Gyrovector spaces turn out to form the algebraic setting for the hyperbolic geometry of Bolyai and Lobachevsky just as vector spaces form the algebraic setting for Euclidean geometry [51, 52, 66]. This discovery resulted in the extension by abstraction of Thomas precession into gyration, and in subsequent studies presented in several books [52, 54, 56, 58–60] reviewed, for instance, in [66] and [28]. The hyperbolic geometric character of Thomas precession is emphasized here by the observation that it can be interpreted as the defect of a related hyperbolic triangle. Other novel relationships between the special relativity theory of Einstein and the hyperbolic geometry of Bolyai and Lobachevsky are presented in [57, 61].

Following Malykin's observations in [24], there is a need to demonstrate that our study of Thomas precession does lead to the correct Thomas precession angular velocity. Accordingly, in this article we review the studies in [52, 54, 56, 58–60] of both Thomas precession and its abstract version, gyration. Based on the review we derive the correct Thomas precession angular velocity, illustrated in Fig. 4, p. 484.

A boost is a Lorentz transformation without rotation [48]. The Thomas precession angle ϵ is generated by the application of two successive boosts with velocity parameters, say, **u** and **v**. The angle θ between **u** and **v** is the generating angle of the resulting Thomas precession angle ϵ, shown in Fig. 4.

An important question about the Thomas precession angle ϵ and its generating angle θ is whether or not ϵ and θ have equal signs. According to Malykin [24], some explorers claim that ϵ and θ have equal signs while some other explorers claim that ϵ and θ have opposite signs. In particular, Malykin claims that these angles have equal signs while, in contrast, we demonstrate here convincingly that these angles have opposite signs. Our demonstration is convincing since it accompanies a focal identity, (118), that interested explorers can test (both theoretically and) numerically in order to corroborate our claim that ϵ and θ have opposite signs.

In order to pave the way to the study Thomas precession we present the path from Einstein velocity addition to the gyroalgebra of gyrogroups and gyrations, and to the gyrogeometry that coincides with the hyperbolic geometry of Bolyai and Lobachevsky. We, then, demonstrate that the concept of Thomas precession in Einstein's special theory of relativity is a concrete realization of the abstract concept of gyration in gyroalgebra.

A signed angle θ, $-\pi < \theta < \pi$, between two non-parallel vectors **u** and **v** in the Euclidean 3-space $\mathbb{R}^3$ is positive (negative) if the angle θ drawn from **u** to **v** is drawn counterclockwise (clockwise). The relationship between the Thomas precession signed angle of rotation, ϵ, and its generating signed angle, θ, shown in Fig. 4, is important. Hence, finally, we pay special attention to the relationship between the Thomas precession signed angle of rotation ϵ and its generating signed angle θ, demonstrating that these have opposite signs.

2 Einstein Velocity Addition and Scalar Multiplication

Let $(\mathbb{R}^3, +, \cdot)$ be the Euclidean 3-space with its common vector addition, $+$, and inner product, $\cdot$, and let

$$\mathbb{R}^3_c = \{\mathbf{v} \in \mathbb{R}^3 : \|\mathbf{v}\| < c\} \tag{1}$$

be the c-ball of all relativistically admissible velocities of material particles, where c is the vacuum speed of light.

Einstein velocity addition is a binary operation, $\oplus$, in the c-ball $\mathbb{R}^3_c$ of all relativistically admissible velocities, given by the equation, [52], [33, Eq. 2.9.2],[27, p. 55] and [11],

$$\begin{aligned}\mathbf{u}\oplus\mathbf{v} &= \frac{1}{1+\frac{\mathbf{u}\cdot\mathbf{v}}{c^2}}\left\{\mathbf{u}+\frac{1}{\gamma_{\mathbf{u}}}\mathbf{v}+\frac{1}{c^2}\frac{\gamma_{\mathbf{u}}}{1+\gamma_{\mathbf{u}}}(\mathbf{u}\cdot\mathbf{v})\mathbf{u}\right\}\\ &= \frac{1}{1+\frac{\mathbf{u}\cdot\mathbf{v}}{c^2}}\left\{\mathbf{u}+\mathbf{v}+\frac{1}{c^2}\frac{\gamma_{\mathbf{u}}}{1+\gamma_{\mathbf{u}}}\left(\mathbf{u}\times(\mathbf{u}\times\mathbf{v})\right)\right\}\end{aligned} \tag{2}$$

for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3_c$, where $\gamma_{\mathbf{u}}$ is the gamma factor given by the equation

$$\gamma_{\mathbf{v}} = \frac{1}{\sqrt{1 - \dfrac{\|\mathbf{v}\|^2}{c^2}}} \tag{3}$$

Here $\mathbf{u}\cdot\mathbf{v}$ and $\|\mathbf{v}\|$ are the inner product and the norm in the ball, which the ball $\mathbb{R}^3_c$ inherits from its space $\mathbb{R}^3$, $\|\mathbf{v}\|^2 = \mathbf{v}\cdot\mathbf{v} = \mathbf{v}^2$. Recalling that a nonempty set with a binary operation is called a *groupoid*, the Einstein groupoid $(\mathbb{R}^3_c, \oplus)$ is called an *Einstein gyrogroup*. A formal definition of the abstract gyrogroup will be presented in Definition 4.

Einstein addition admits scalar multiplication $\otimes$, giving rise to the Einstein gyrovector space $(\mathbb{R}^3_c, \oplus, \otimes)$. Remarkably, the resulting Einstein gyrovector spaces $(\mathbb{R}^n_c, \oplus, \otimes)$ form the setting for the Cartesian-Beltrami-Klein ball model of hyperbolic geometry, just as vector spaces form the setting for the standard Cartesian model of Euclidean geometry, as we will see in the sequel.

Let $k\otimes\mathbf{v}$ be the Einstein addition of k copies of $\mathbf{v} \in \mathbb{R}^n_c$, that is $k\otimes\mathbf{v} = \mathbf{v}\oplus\mathbf{v}\ldots\oplus\mathbf{v}$ (k terms). Then,

$$k\otimes\mathbf{v} = c\,\frac{\left(1+\dfrac{\|\mathbf{v}\|}{c}\right)^k - \left(1-\dfrac{\|\mathbf{v}\|}{c}\right)^k}{\left(1+\dfrac{\|\mathbf{v}\|}{c}\right)^k + \left(1-\dfrac{\|\mathbf{v}\|}{c}\right)^k}\,\frac{\mathbf{v}}{\|\mathbf{v}\|} \tag{4}$$

The definition of scalar multiplication in an Einstein gyrovector space requires analytically continuing k off the positive integers, thus obtaining the following definition:

Definition 1. (Einstein Scalar Multiplication; Einstein Gyrovector Spaces). An Einstein gyrovector space $(\mathbb{R}^n_s, \oplus, \otimes)$ is an Einstein gyrogroup $(\mathbb{R}^n_s, \oplus)$ with scalar multiplication $\otimes$ given by

$$r\otimes\mathbf{v} = s\,\frac{\left(1+\dfrac{\|\mathbf{v}\|}{s}\right)^r - \left(1-\dfrac{\|\mathbf{v}\|}{s}\right)^r}{\left(1+\dfrac{\|\mathbf{v}\|}{s}\right)^r + \left(1-\dfrac{\|\mathbf{v}\|}{s}\right)^r}\,\frac{\mathbf{v}}{\|\mathbf{v}\|} = s\tanh(r\ \tanh^{-1}\frac{\|\mathbf{v}\|}{s})\frac{\mathbf{v}}{\|\mathbf{v}\|} \tag{5}$$

where r is any real number, $r \in \mathbb{R}$, $\mathbf{v} \in \mathbb{R}^n_s$, $\mathbf{v} \neq \mathbf{0}$, and $r\otimes\mathbf{0} = \mathbf{0}$, and with which we use the notation $\mathbf{v}\otimes r = r\otimes\mathbf{v}$.

In the Newtonian limit of large c, $c \to \infty$, the ball $\mathbb{R}^3_c$ expands to the whole of its space $\mathbb{R}^3$, as we see from (1), and Einstein addition $\oplus$ in $\mathbb{R}^3_c$ reduces to the ordinary vector addition $+$ in $\mathbb{R}^3$, as we see from (2) and (3).

Einstein addition (2) of relativistically admissible velocities was introduced by Einstein in his 1905 paper [7] and [8, p. 141] that founded the special theory of relativity. One has to remember here that the Euclidean 3-vector algebra was not so widely known in 1905 and, consequently, was not used by Einstein. Einstein calculated in [7] the behavior of the velocity components parallel and orthogonal to the relative velocity between inertial systems, which is as close as one can get without vectors to the vectorial version (2) of Einstein addition.

We naturally use the abbreviation $\mathbf{u}\ominus\mathbf{v} = \mathbf{u}\oplus(-\mathbf{v})$ for Einstein subtraction, so that, for instance, $\mathbf{v}\ominus\mathbf{v} = \mathbf{0}$, $\ominus\mathbf{v} = \mathbf{0}\ominus\mathbf{v} = -\mathbf{v}$ and, in particular,

$$\ominus(\mathbf{u}\oplus\mathbf{v}) = \ominus\mathbf{u}\ominus\mathbf{v} \tag{6}$$

and

$$\ominus\mathbf{u}\oplus(\mathbf{u}\oplus\mathbf{v}) = \mathbf{v} \tag{7}$$

for all $\mathbf{u}, \mathbf{v}$ in the ball $\mathbb{R}^3_c$, in full analogy with vector addition and subtraction in $\mathbb{R}^3$. Identity (6) is known as the *automorphic inverse property*, and Identity (7) is known as the *left cancellation law* of Einstein addition [56]. We may note that Einstein addition does not obey the naive right counterpart of the left cancellation law (7) since, in general,

$$(\mathbf{u}\oplus\mathbf{v})\ominus\mathbf{v} \neq \mathbf{u} \tag{8}$$

The seemingly lack of a right cancellation law for Einstein addition is repaired in (83) by the introduction of a second binary operation, called *Einstein coaddition*, (81), which captures important analogies with classical results.

Einstein addition and the gamma factor are related by the *gamma identity*,

$$\gamma_{\mathbf{u}\oplus\mathbf{v}} = \gamma_{\mathbf{u}}\gamma_{\mathbf{v}}\left(1 + \frac{\mathbf{u}\cdot\mathbf{v}}{c^2}\right) \tag{9a}$$

which can be written, equivalently, as

$$\gamma_{\ominus\mathbf{u}\oplus\mathbf{v}} = \gamma_{\mathbf{u}}\gamma_{\mathbf{v}}\left(1 - \frac{\mathbf{u}\cdot\mathbf{v}}{c^2}\right) \tag{9b}$$

for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3_c$. Here, (9b) is obtained from (9a) by replacing $\mathbf{u}$ by $\ominus\mathbf{u} = -\mathbf{u}$.

A frequently used identity that follows immediately from (3) is

$$\frac{\mathbf{v}^2}{c^2} = \frac{\|\mathbf{v}\|^2}{c^2} = \frac{\gamma_{\mathbf{v}}^2 - 1}{\gamma_{\mathbf{v}}^2} \tag{10}$$

and, similarly, useful identities that follow immediately from (9) are

$$\frac{\mathbf{u}\cdot\mathbf{v}}{c^2} = -1 + \frac{\gamma_{\mathbf{u}\oplus\mathbf{v}}}{\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}} \tag{11a}$$

and

$$\frac{\mathbf{u}\cdot\mathbf{v}}{c^2} = 1 - \frac{\gamma_{\ominus\mathbf{u}\oplus\mathbf{v}}}{\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}} \tag{11b}$$

implying

$$\gamma_{\mathbf{u}\oplus\mathbf{v}} - \gamma_{\mathbf{u}}\gamma_{\mathbf{v}} = -\gamma_{\ominus\mathbf{u}\oplus\mathbf{v}} + \gamma_{\mathbf{u}}\gamma_{\mathbf{v}} \tag{11c}$$

In Identity (11c) the left-hand side seems to be more elegant, in form, than the right-hand side. Geometrically, however, the right-hand side of this identity is advantageous over its left-hand side because the gamma factor $\gamma_{\ominus\mathbf{u}\oplus\mathbf{v}}$ that appears on the right-hand side possesses a geometric interpretation. It has a geometric interpretation in hyperbolic triangles, called *gyrotriangles*, as explained in [60, Sect. 2.10]. To be more specific, we recall in Sect. 4 relevant results from [60].

Einstein addition is noncommutative. Indeed, in general,

$$\mathbf{u}\oplus\mathbf{v} \neq \mathbf{v}\oplus\mathbf{u} \tag{12}$$

$\mathbf{u}, \mathbf{v} \in \mathbb{R}^3_c$. Moreover, Einstein addition is also nonassociative since, in general,

$$(\mathbf{u}\oplus\mathbf{v})\oplus\mathbf{w} \neq \mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w}) \tag{13}$$

$\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3_c$.

It seems that following the breakdown of commutativity and associativity in Einstein addition some mathematical regularity has been lost in the transition from Newton's velocity vector addition in $\mathbb{R}^3$ to Einstein's velocity addition (2) in $\mathbb{R}^3_c$. This is, however, not the case since gyrations come to the rescue, as we will see in Sect. 5. Owing to the presence of gyrations, the Einstein groupoid $(\mathbb{R}^3_c, \oplus)$ has a grouplike structure [49] that we naturally call an *Einstein gyrogroup* [52]. The formal definition of the resulting abstract gyrogroup will be presented in Definition 4 and Sect. 7.

3 Linking Einstein Addition to Hyperbolic Geometry

The Einstein gyrodistance function, $d(\mathbf{u}, \mathbf{v})$, in an Einstein gyrovector space $(\mathbb{R}^n_c, \oplus, \otimes)$ is given by the equation

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u}\ominus\mathbf{v}\| \tag{14}$$

$\mathbf{u}, \mathbf{v} \in \mathbb{R}^n_c$. We call it a *gyrodistance function* in order to emphasize the analogies it shares with its Euclidean counterpart, the distance function $\|\mathbf{u} - \mathbf{v}\|$ in $\mathbb{R}^n$. Among these analogies is the gyrotriangle inequality according to which

$$\|\mathbf{u}\oplus\mathbf{v}\| \leq \|\mathbf{u}\|\oplus\|\mathbf{v}\| \tag{15}$$

for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n_c$. For this and other analogies that distance and gyrodistance functions share see [54, 56].

In a two dimensional Einstein gyrovector space $(\mathbb{R}^2_s, \oplus, \otimes)$ the squared gyrodistance between a point $\mathbf{x} \in \mathbb{R}^2_s$ and an infinitesimally nearby point $\mathbf{x} + d\mathbf{x} \in \mathbb{R}^2_s$, $d\mathbf{x} = (dx_1,\ dx_2)$, is given by the equation [56, Sect. 7.5] and [54, Sect. 7.5]

$$\begin{aligned} ds^2 &= \|(\mathbf{x} + d\mathbf{x}) \ominus \mathbf{x}\|^2 \\ &= E dx_1^2 + 2F dx_1 dx_2 + G dx_2^2 + \dots \end{aligned} \tag{16}$$

where, if we use the notation $r^2 = x_1^2 + x_2^2$, we have

$$\begin{aligned} E &= c^2 \frac{c^2 - x_2^2}{(c^2 - r^2)^2} \\ F &= c^2 \frac{x_1 x_2}{(c^2 - r^2)^2} \\ G &= c^2 \frac{c^2 - x_1^2}{(c^2 - r^2)^2} \end{aligned} \tag{17}$$

The triple $(g_{11}, g_{12}, g_{22}) = (E, F, G)$ along with $g_{21} = g_{12}$ is known in differential geometry as the metric tensor g_{ij} [21]. It turns out to be the metric tensor of the Beltrami-Klein disc model of hyperbolic geometry [26, p. 220]. Hence, ds^2 in (16) and (17) is the Riemannian line element of the Beltrami-Klein disc model of hyperbolic geometry, linked to Einstein velocity addition (2) and to Einstein gyrodistance function (14) [55].

The link between Einstein gyrovector spaces and the Beltrami-Klein ball model of hyperbolic geometry, already noted by Fock [11, p. 39], has thus been established in (14)–(17) in two dimensions. The extension of the link to higher dimensions is presented in [52, Sect. 9, Chap. 3], [56, Sect. 7.5] [54, Sect. 7.5] and [55]. For a brief account of the history of linking Einstein's velocity addition law with hyperbolic geometry see [30, p. 943].

4 Gyrotriangle, the Hyperbolic Triangle

In this inspirational section we present recollections from [60] that intend to motivate Thomas precession explorers to study the gyrostructure of Einstein addition and its underlying hyperbolic geometry.

Let U, V and W be the three vertices of a gyrotriangle UVW in an Einstein gyrovector space $(\mathbb{R}_c^3, \oplus, \otimes)$, shown in Fig. 1. Then, in full analogy with Euclidean geometry, the three sides of the gyrotriangle form the three *gyrovectors*

$$\begin{aligned} \mathbf{u} &= \ominus W \oplus V \\ \mathbf{v} &= \ominus W \oplus U \\ \mathbf{w} &= \ominus U \oplus V \end{aligned} \tag{18}$$

and the corresponding three side-gyrolengths of the gyrotriangle are

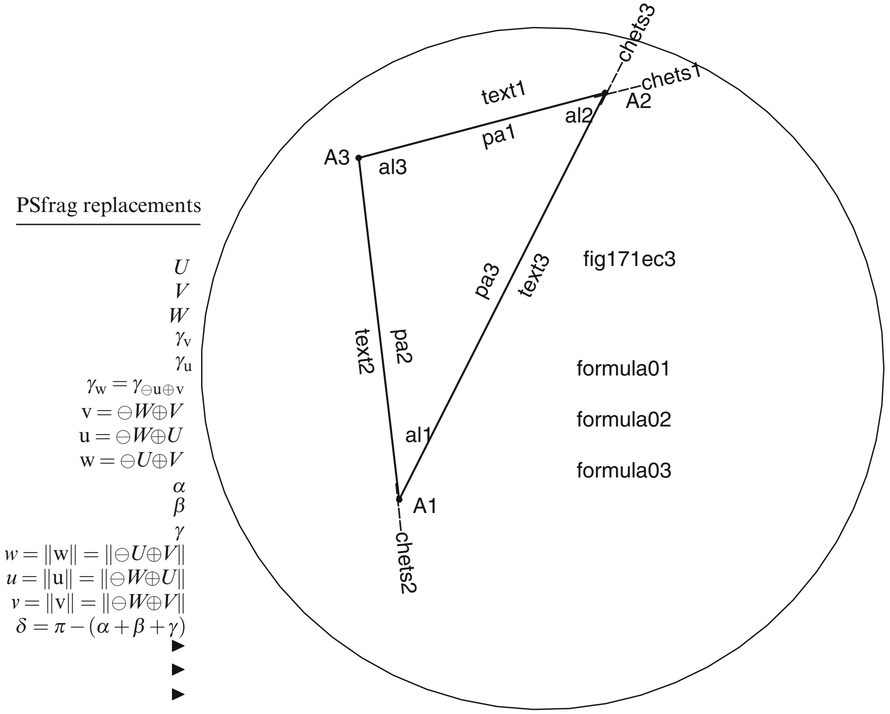

PSfrag replacements
U
V
W
$\gamma_{\rm v}$
$\gamma_{\rm u}$
$\gamma_{\rm w} = \gamma_{\ominus{\rm u}\oplus{\rm v}}$
${\rm v} = \ominus W \oplus V$
${\rm u} = \ominus W \oplus U$
${\rm w} = \ominus U \oplus V$
α
β
γ
$w = \|{\rm w}\| = \|\ominus U \oplus V\|$
$u = \|{\rm u}\| = \|\ominus W \oplus U\|$
$v = \|{\rm v}\| = \|\ominus W \oplus V\|$
$\delta = \pi - (\alpha + \beta + \gamma)$
▶
▶
▶
chets3
chets1
text1
A2
al2
pa1
A3
al3
fig171ec3
pa3
text3
text2
pa2
formula01
formula02
al1
formula03
A1
chets2

Fig. 1 The gyrotriangle UVW in an Einstein gyrovector space $(\mathbb{R}^n_s, \oplus, \otimes)$ is shown for $n = 2$. Its sides are presented graphically as gyrosegments that join the vertices. They form the gyrovectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$, side-gyrolengths, u, v, w, and gyroangles, α, β, γ. The gyrotriangle gyroangle sum is less than π, the difference, $\delta = \pi - (\alpha + \beta + \gamma)$, being the gyrotriangular defect. For similar figures, which illustrate Einsteinian gyrotriangle gyroangles vs. Euclidean triangle angles(see, for instance, [60, Fig. 7.1, p. 150] and [60, Fig. 7.2, p. 155])

$$\begin{aligned} u &= \|\mathbf{u}\| = \|\ominus W \oplus V\| \\ v &= \|\mathbf{v}\| = \|\ominus W \oplus U\| \\ w &= \|\mathbf{w}\| = \|\ominus U \oplus V\| = \|\ominus\mathbf{u}\oplus\mathbf{v}\| \end{aligned} \tag{19}$$

and the side gamma factors of the gyrotriangle are, accordingly, $\gamma_{\mathbf{u}}$, $\gamma_{\mathbf{v}}$ and

$$\gamma_{\mathbf{w}} = \gamma_{\ominus\mathbf{u}\oplus\mathbf{v}} \tag{20}$$

Hence, by (20), the gamma factors $\gamma_{\mathbf{u}}$, $\gamma_{\mathbf{v}}$ and $\gamma_{\ominus\mathbf{u}\oplus\mathbf{v}}$ in (11c) can be interpreted geometrically as the gamma factors of the three sides of a gyrotriangle UVW.

For later reference we recall here that the gyrotriangular defect δ of the gyrotriangle UVW is given in terms of the gyrotriangle side gamma factors by the equation [59, Theorem 2.32] [60, Theorem 6.11]

$$\tan\tfrac{\delta}{2} = \frac{\sqrt{1 + 2\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}\gamma_{\mathbf{w}} - \gamma_{\mathbf{u}}^2 - \gamma_{\mathbf{v}}^2 - \gamma_{\mathbf{w}}^2}}{1 + \gamma_{\mathbf{u}} + \gamma_{\mathbf{v}} + \gamma_{\mathbf{w}}} \tag{21}$$

where $\gamma_{\mathbf{w}} = \gamma_{\ominus\mathbf{u}\oplus\mathbf{v}}$.

It is the gamma identity (9a) that signaled the emergence of hyperbolic geometry in special relativity when it was first studied by Sommerfeld [38] and Varičak [62, 63]. Historically, it formed the first link between special relativity and the hyperbolic geometry of Bolyai and Lobachevsky, recently leading to the novel trigonometry in hyperbolic geometry that became known as *gyrotrigonometry*, studied in [52,54,56, 58–60].

If, unlike what we see in Fig. 1, $U, V, W \in \mathbb{R}_c^3 \subset \mathbb{R}^3$ are viewed as points of the Euclidean 3-space $\mathbb{R}^3$, then they give rise to the two vectors

$$\begin{aligned}\mathbf{u} &= -W + U\\ \mathbf{v} &= -W + V\end{aligned} \tag{22}$$

that emanate from the point W. The point W, in turn, forms the vertex of the included angle $\gamma = \angle UWV$ that satisfies the equation

$$\cos\gamma = \frac{-W + U}{\|-W + U\|}\cdot\frac{-W + U}{\|-W + U\|} \tag{23}$$

In full analogy, if $U, V, W \in \mathbb{R}_c^3 \subset \mathbb{R}^3$ are viewed as points of the Einsteinian 3-gyrospace $\mathbb{R}_c^3$, as we see in Fig. 1, then they give rise to the two *gyrovectors*

$$\begin{aligned}\mathbf{u} &= \ominus W \oplus U\\ \mathbf{v} &= \ominus W \oplus V\end{aligned} \tag{24}$$

that emanate from the point W. The point W, in turn, forms the vertex of the included gyroangle $\gamma = \angle UWV$ that satisfies the equation

$$\cos\gamma = \frac{\ominus W \oplus U}{\|\ominus W \oplus U\|}\cdot\frac{\ominus W \oplus U}{\|\ominus W \oplus U\|} \tag{25}$$

Accordingly, as an example, if the velocities **u** and **v** in Fig. 4, p. 484, are viewed as Newtonian, classical velocities, then these velocities are vectors such that their included angle is the angle θ in Fig. 4. The measure of angle θ, in turn, is given by an equation like (23).

If, however, the velocities **u** and **v** in Fig. 4 are viewed as Einsteinian, relativistic velocities, then these velocities are gyrovectors such that their included gyroangle is the gyroangle θ in Fig. 4. The measure of gyroangle θ, in turn, is given by an equation like (25).

Remarkably, the cosine functions in (23) and (25) are identically the same functions with different arguments: the argument of the cosine function in (23) is an angle, while the argument of the cosine function in (25) is a gyroangle. Yet, it is useful to give them different names, calling the cosine function that is applied to gyroangles the *gyrocosine function* of gyrotrigonometry, as opposed to the cosine function of trigonometry. Accordingly, (23) expresses the cosine function of trigonometry, while (25) expresses the cosine function of gyrotrigonometry.

5 The Gyrostructure of Einstein Addition

Vector addition, $+$, in $\mathbb{R}^3$ is both commutative and associative, satisfying

$$\begin{aligned} \mathbf{u}+\mathbf{v} &= \mathbf{v}+\mathbf{u} && \text{Commutative Law} \\ \mathbf{u}+(\mathbf{v}+\mathbf{w}) &= (\mathbf{u}+\mathbf{v})+\mathbf{w} && \text{Associative Law} \end{aligned} \tag{26}$$

for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3$. In contrast, Einstein addition, $\oplus$, in $\mathbb{R}^3_c$ is neither commutative nor associative.

In order to measure the extent to which Einstein addition deviates from associativity we introduce *gyrations*, which are maps that are *trivial* in the special cases when the application of $\oplus$ is associative. For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3_c$ the gyration $\mathrm{gyr}[\mathbf{u}, \mathbf{v}]$ is an automorphism of the Einstein groupoid $(\mathbb{R}^3_c, \oplus)$ onto itself, given in terms of Einstein addition by the equation

$$\mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{w} = \ominus(\mathbf{u}\oplus\mathbf{v})\oplus\{\mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w})\} \tag{27}$$

for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3_c$.

We recall that an automorphism of a groupoid $(S, \oplus)$ is a one-to-one map f of S onto itself that respects the binary operation, that is, $f(a\oplus b) = f(a)\oplus f(b)$ for all $a, b \in S$. The set of all automorphisms of a groupoid $(S, \oplus)$ forms a group, denoted $\mathrm{Aut}(S, \oplus)$, where the group operation is given by automorphism composition. To emphasize that the gyrations of an Einstein gyrogroup $(\mathbb{R}^3_c, \oplus)$ are automorphisms of the gyrogroup, gyrations are also called *gyroautomorphisms*.

A gyration $\mathrm{gyr}[\mathbf{u}, \mathbf{v}]$, $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3_c$, is *trivial* if $\mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{w} = \mathbf{w}$ for all $\mathbf{w} \in \mathbb{R}^3_c$. Thus, for instance, the gyrations $\mathrm{gyr}[\mathbf{0}, \mathbf{v}]$, $\mathrm{gyr}[\mathbf{v}, \mathbf{v}]$ and $\mathrm{gyr}[\mathbf{v}, \ominus\mathbf{v}]$ are trivial for all $\mathbf{v} \in \mathbb{R}^3_c$, as we see from (27) and (7).

Einstein gyrations, which possess their own rich structure, measure the extent to which Einstein addition deviates from both commutativity and associativity as we see from the gyrocommutative and the gyroassociative laws of Einstein addition in the following list of identities [52, 54, 56, 58–60], each of which has a name:

$$
\begin{aligned}
\mathbf{u}\oplus\mathbf{v} &= \mathrm{gyr}[\mathbf{u},\mathbf{v}](\mathbf{v}\oplus\mathbf{u}) && \text{Gyrocommutative Law}\\
\mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w}) &= (\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w} && \text{Left Gyroassociative Law}\\
(\mathbf{u}\oplus\mathbf{v})\oplus\mathbf{w} &= \mathbf{u}\oplus(\mathbf{v}\oplus\mathrm{gyr}[\mathbf{v},\mathbf{u}]\mathbf{w}) && \text{Right Gyroassociative Law}\\
\mathrm{gyr}[\mathbf{u}\oplus\mathbf{v},\mathbf{v}] &= \mathrm{gyr}[\mathbf{u},\mathbf{v}] && \text{Gyration Left Loop Property}\\
\mathrm{gyr}[\mathbf{u},\mathbf{v}\oplus\mathbf{u}] &= \mathrm{gyr}[\mathbf{u},\mathbf{v}] && \text{Gyration Right Loop Property}\\
\mathrm{gyr}[\ominus\mathbf{u},\ominus\mathbf{v}] &= \mathrm{gyr}[\mathbf{u},\mathbf{v}] && \text{Gyration Even Property}\\
(\mathrm{gyr}[\mathbf{u},\mathbf{v}])^{-1} &= \mathrm{gyr}[\mathbf{v},\mathbf{u}] && \text{Gyration Inversion Law}
\end{aligned}
\tag{28}
$$

for all $\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbb{R}^3_c$.

Einstein addition is thus regulated by gyrations to which it gives rise owing to its nonassociativity, so that Einstein addition and its gyrations are inextricably linked. The resulting gyrocommutative gyrogroup structure of Einstein addition was discovered in 1988 [45]. Interestingly, gyrations are the mathematical abstraction of the relativistic mechanical effect known as *Thomas precession* [56, Sect. 10.3], as we will see in Sect. 9.

6 Gyrations

Owing to its nonassociativity, Einstein addition gives rise in (27) to gyrations

$$\mathrm{gyr}[\mathbf{u},\mathbf{v}] : \mathbb{R}^3_c \to \mathbb{R}^3_c \tag{29}$$

for any $\mathbf{u},\mathbf{v}\in\mathbb{R}^3_c$. Gyrations, in turn, regulate Einstein addition, endowing it with the rich structure of a gyrocommutative gyrogroup that will be formalize in Sect. 7.

The gyration equation is expressed in (27) in terms of Einstein addition. Expressing it explicitly, in terms of vector addition and vector scalar product rather than Einstein addition, we obtain the equation

$$\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w} = \mathbf{w} + \frac{A\mathbf{u}+B\mathbf{v}}{D} \tag{30}$$

where

$$
\begin{aligned}
A = &-\frac{1}{c^2}\frac{\gamma_{\mathbf{u}}^2}{(\gamma_{\mathbf{u}}+1)}(\gamma_{\mathbf{v}}-1)(\mathbf{u}\cdot\mathbf{w}) + \frac{1}{c^2}\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}(\mathbf{v}\cdot\mathbf{w})\\
&+\frac{2}{c^4}\frac{\gamma_{\mathbf{u}}^2\gamma_{\mathbf{v}}^2}{(\gamma_{\mathbf{u}}+1)(\gamma_{\mathbf{v}}+1)}(\mathbf{u}\cdot\mathbf{v})(\mathbf{v}\cdot\mathbf{w})
\end{aligned}
$$

$$B = -\frac{1}{c^2}\frac{\gamma_{\mathbf{v}}}{\gamma_{\mathbf{v}}+1}\{\gamma_{\mathbf{u}}(\gamma_{\mathbf{v}}+1)(\mathbf{u}\cdot\mathbf{w}) + (\gamma_{\mathbf{u}}-1)\gamma_{\mathbf{v}}(\mathbf{v}\cdot\mathbf{w})\}$$

$$D = \gamma_{\mathbf{u}}\gamma_{\mathbf{v}}(1+\frac{\mathbf{u}\cdot\mathbf{v}}{c^2}) + 1 = \gamma_{\mathbf{u}\oplus\mathbf{v}} + 1 > 1 \tag{31}$$

for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}_c^3$. Clearly, the domain of $\mathbf{u}$ and $\mathbf{v}$ in (31) must be restricted to the open ball $\mathbb{R}_c^3$ to insure the reality of the Lorentz factors $\gamma_{\mathbf{u}}$ and $\gamma_{\mathbf{v}}$. In contrast, however, the domain of $\mathbf{w}$ need not be restricted to the ball.

Allowing $\mathbf{w} \in \mathbb{R}^3 \supset \mathbb{R}_c^3$ in (30) and (31), that is, extending the domain of $\mathbf{w}$ from $\mathbb{R}_c^3$ to $\mathbb{R}^3$, gyrations gyr$[\mathbf{u}, \mathbf{v}]$ are expendable from self-maps of $\mathbb{R}_c^3$ to linear self-maps of $\mathbb{R}^3$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}_c^3$. Indeed,

$$\mathrm{gyr}[\mathbf{u}, \mathbf{v}](r_1\mathbf{w}_1 + r_2\mathbf{w}_2) = r_1\mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{w}_1 + r_2\mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{w}_2 \tag{32}$$

for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}_c^3$, $\mathbf{w} \in \mathbb{R}^3$ and $r_1, r_2 \in \mathbb{R}$.

In each of the three special cases when (1) $\mathbf{u} = \mathbf{0}$, or (2) $\mathbf{v} = \mathbf{0}$, or (3) $\mathbf{u}$ and $\mathbf{v}$ are parallel in $\mathbb{R}^3$, $\mathbf{u}\|\mathbf{v}$, we have $A\mathbf{u} + B\mathbf{v} = \mathbf{0}$, so that in these cases gyr$[\mathbf{u}, \mathbf{v}]$ is trivial. Thus, we have

$$\begin{aligned} \mathrm{gyr}[\mathbf{0}, \mathbf{v}]\mathbf{w} &= \mathbf{w} \\ \mathrm{gyr}[\mathbf{u}, \mathbf{0}]\mathbf{w} &= \mathbf{w} \\ \mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{w} &= \mathbf{w}, \qquad \mathbf{u}\|\mathbf{v} \end{aligned} \tag{33}$$

for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}_c^3 \subset \mathbb{R}^3$ and all $\mathbf{w} \in \mathbb{R}^3$.

It follows from (30) and (31) that

$$\mathrm{gyr}[\mathbf{v}, \mathbf{u}](\mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{w}) = \mathbf{w} \tag{34}$$

for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}_c^3$, $\mathbf{w} \in \mathbb{R}^3$, so that gyrations are invertible linear maps of $\mathbb{R}^3$, the inverse, gyr$^{-1}[\mathbf{u}, \mathbf{v}]$, (28), of gyr$[\mathbf{u}, \mathbf{v}]$ being gyr$[\mathbf{v}, \mathbf{u}]$. We thus obtain from (34) the gyration inversion property in (28),

$$\mathrm{gyr}^{-1}[\mathbf{u}, \mathbf{v}] = \mathrm{gyr}[\mathbf{v}, \mathbf{u}] \tag{35}$$

for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}_c^3$.

Gyrations keep the inner product of elements of the ball $\mathbb{R}_c^3$ invariant, that is,

$$\mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{a}\cdot\mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{b} = \mathbf{a}\cdot\mathbf{b} \tag{36}$$

for all $\mathbf{a}, \mathbf{b}, \mathbf{u}, \mathbf{v} \in \mathbb{R}_c^3$. Hence, in particular, gyr$[\mathbf{u}, \mathbf{v}]$ is an *isometry* of $\mathbb{R}_c^3$, keeping the norm of elements of the ball $\mathbb{R}_c^3$ invariant,

$$\|\mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{w}\| = \|\mathbf{w}\| \tag{37}$$

Accordingly, for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3_c$, $\mathrm{gyr}[\mathbf{u}, \mathbf{v}]$ represents a rotation of the ball $\mathbb{R}^3_c$ about its origin.

The invertible self-map $\mathrm{gyr}[\mathbf{u}, \mathbf{v}]$ of $\mathbb{R}^3_c$ respects Einstein addition in $\mathbb{R}^3_c$,

$$\mathrm{gyr}[\mathbf{u}, \mathbf{v}](\mathbf{a} \oplus \mathbf{b}) = \mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{a} \oplus \mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{b} \tag{38}$$

for all $\mathbf{a}, \mathbf{b}, \mathbf{u}, \mathbf{v} \in \mathbb{R}^3_c$, so that, by (35) and (38), $\mathrm{gyr}[\mathbf{u}, \mathbf{v}]$ is an automorphism of the Einstein groupoid $(\mathbb{R}^3_c, \oplus)$.

7 Gyrogroups

Taking the key features of the Einstein groupoid $(\mathbb{R}^3_c, \oplus)$ as axioms, and guided by analogies with groups, we are led to the formal gyrogroup definition in which gyrogroups turn out to form a most natural generalization of groups. Definitions related to groups and gyrogroups thus follow.

Definition 2. (Groups). A groupoid $(G, +)$ is a group if its binary operation satisfies the following axioms. In G there is at least one element, 0, called a left identity, satisfying
(G1) $\quad 0 + a = a$
for all $a \in G$. There is an element $0 \in G$ satisfying axiom $(G1)$ such that for each $a \in G$ there is an element $-a \in G$, called a left inverse of a, satisfying
(G2) $\quad -a + a = 0$
Moreover, the binary operation obeys the associative law
(G3) $\quad (a + b) + c = a + (b + c)$
for all $a, b, c \in G$.

Groups are classified into commutative and noncommutative groups.

Definition 3. (Commutative Groups). A group $(G, +)$ is commutative if its binary operation obeys the commutative law
(G6) $\quad a + b = b + a$
for all $a, b \in G$.

Definition 4. (Gyrogroups). A groupoid $(G, \oplus)$ is a gyrogroup if its binary operation satisfies the following axioms. In G there is at least one element, 0, called a left identity, satisfying
(G1) $\quad 0 \oplus a = a$
for all $a \in G$. There is an element $0 \in G$ satisfying axiom $(G1)$ such that for each $a \in G$ there is an element $\ominus a \in G$, called a left inverse of a, satisfying
(G2) $\quad \ominus a \oplus a = 0\,.$
Moreover, for any $a, b, c \in G$ there exists a unique element $\mathrm{gyr}[a, b]c \in G$ such that the binary operation obeys the left gyroassociative law
(G3) $\quad a \oplus (b \oplus c) = (a \oplus b) \oplus \mathrm{gyr}[a, b]c\,.$

The map $\mathrm{gyr}[a,b] : G \to G$ given by $c \mapsto \mathrm{gyr}[a,b]c$ is an automorphism of the groupoid $(G,\oplus)$, that is,

(G4) $\qquad \mathrm{gyr}[a,b] \in \mathrm{Aut}(G,\oplus)\,,$

and the automorphism $\mathrm{gyr}[a,b]$ of G is called the gyroautomorphism, or the gyration, of G generated by $a,b \in G$. The operator $\mathrm{gyr} : G \times G \to \mathrm{Aut}(G,\oplus)$ is called the gyrator of G. Finally, the gyroautomorphism $\mathrm{gyr}[a,b]$ generated by any $a,b \in G$ possesses the left loop property

(G5) $\qquad \mathrm{gyr}[a,b] = \mathrm{gyr}[a\oplus b, b]\,.$

The gyrogroup axioms $(G1)$–$(G5)$ in Definition 4 are classified into three classes:

1. The first pair of axioms, $(G1)$ and $(G2)$ in Definition 4, is a reminiscent of the group axioms $(G1)$ and $(G2)$ in Definition 2.
2. The last pair of axioms, $(G4)$ and $(G5)$ in Definition 4, presents the gyrator axioms.
3. The middle axiom, $(G3)$ in Definition 4, is a hybrid axiom linking the two pairs of axioms in Items (1) and (2).

As in group theory, we use the notation $a\ominus b = a\oplus(\ominus b)$ in gyrogroup theory as well.

In full analogy with groups, gyrogroups are classified into gyrocommutative and non-gyrocommutative gyrogroups.

Definition 5. (Gyrocommutative Gyrogroups). A gyrogroup $(G,\oplus)$ is gyrocommutative if its binary operation obeys the gyrocommutative law

(G6) $\qquad a \oplus b = \mathrm{gyr}[a,b](b \oplus a)$

for all $a,b \in G$.

Gyrogroups, finite and infinite, gyrocommutative and non-gyrocommutative, [32, 37, 53], abound in group theory, as demonstrated in [9, 10, 12, 13], forming fertile areas for research published in six books over the last 10 years [52, 54, 56, 58–60]. Some first gyrogroup theorems, some of which are analogous to group theorems, are presented, for instance, in [54, Chap. 2].

While it is clear how to define a right identity and a right inverse in a gyrogroup, the existence of such elements is not presumed. Indeed, the existence of a unique identity and a unique inverse, both left and right, is a consequence of the gyrogroup axioms, as the following theorem shows, along with other immediate results.

Theorem 1 (First Gyrogroup Properties). *Let $(G,\,\oplus)$ be a gyrogroup. For any elements $a,b,c,x \in G$ we have the following results:*

1. *If $a\oplus b = a\oplus c$, then $b = c$ (general left cancellation law; see Item (9) below).*
2. $\mathrm{gyr}[0,a] = I$ *for any left identity* 0 *in* G.
3. $\mathrm{gyr}[x,a] = I$ *for any left inverse x of a in G.*
4. $\mathrm{gyr}[a,a] = I$
5. *There is a left identity which is a right identity.*
6. *There is only one left identity.*

7. *Every left inverse is a right inverse.*
8. *There is only one left inverse, $\ominus a$, of a, and $\ominus(\ominus a) = a$.*
9. *The Left Cancellation Law:*

$$\ominus a \oplus (a \oplus b) = b \tag{39}$$

10. *The Gyrator Identity:*

$$\mathrm{gyr}[a, b]x = \ominus(a \oplus b) \oplus \{a \oplus (b \oplus x)\} \tag{40}$$

11. $\mathrm{gyr}[a, b]0 = 0$.
12. $\mathrm{gyr}[a, b](\ominus x) = \ominus \mathrm{gyr}[a, b]x$.
13. $\mathrm{gyr}[a, 0] = I$.

Proof.

1. Let x be a left inverse of a corresponding to a left identity, 0, in G. We have $x \oplus (a \oplus b) = x \oplus (a \oplus c)$, implying $(x \oplus a) \oplus \mathrm{gyr}[x, a]b = (x \oplus a) \oplus \mathrm{gyr}[x, a]c$ by left gyroassociativity. Since 0 is a left identity, $\mathrm{gyr}[x, a]b = \mathrm{gyr}[x, a]c$. Since automorphisms are bijective, $b = c$.
2. By left gyroassociativity we have for any left identity 0 of G, $a \oplus x = 0 \oplus (a \oplus x) = (0 \oplus a) \oplus \mathrm{gyr}[0, a]x = a \oplus \mathrm{gyr}[0, a]x$. Hence, by Item 1 above we have $x = \mathrm{gyr}[0, a]x$ for all $x \in G$ so that $\mathrm{gyr}[0, a] = I$.
3. By the left loop property and by Item 2 above we have $\mathrm{gyr}[x, a] = \mathrm{gyr}[x \oplus a, a] = \mathrm{gyr}[0, a] = I$.
4. Follows from an application of the left loop property and Item 2 above.
5. Let x be a left inverse of a corresponding to a left identity, 0, of G. Then by left gyroassociativity and Item 3 above, $x \oplus (a \oplus 0) = (x \oplus a) \oplus \mathrm{gyr}[x, a]0 = 0 \oplus 0 = 0 = x \oplus a$. Hence, by (1), $a \oplus 0 = a$ for all $a \in G$ so that 0 is a right identity.
6. Suppose 0 and 0^* are two left identities, one of which, say 0, is also a right identity. Then $0 = 0^* \oplus 0 = 0^*$.
7. Let x be a left inverse of a. Then $x \oplus (a \oplus x) = (x \oplus a) \oplus \mathrm{gyr}[x, a]x = 0 \oplus x = x = x \oplus 0$, by left gyroassociativity, (G2) of Definition 4 and Items 3, 5, 6 above. By Item 1 we have $a \oplus x = 0$ so that x is a right inverse of a.
8. Suppose x and y are left inverses of a. By Item 7 above, they are also right inverses, so $a \oplus x = 0 = a \oplus y$. By Item 1, $x = y$. Let $\ominus a$ be the resulting unique inverse of a. Then $\ominus a \oplus a = 0$ so that the inverse $\ominus(\ominus a)$ of $\ominus a$ is a.
9. By left gyroassociativity and by 3 we have

$$\ominus a \oplus (a \oplus b) = (\ominus a \oplus a) \oplus \mathrm{gyr}[\ominus a, a]b = b \tag{41}$$

10. By an application of the left cancellation law in Item 9 to the left gyroassociative law (G3) in Definition 4 we obtain the result in Item 10.
11. We obtain Item 11 from 10 with $x = 0$.
12. Since $\mathrm{gyr}[a, b]$ is an automorphism of $(G, \oplus)$ we have from 11

$$\begin{aligned}\mathrm{gyr}[a,b](\ominus x)\oplus\mathrm{gyr}[a,b]x &= \mathrm{gyr}[a,b](\ominus x\oplus x)\\ &= \mathrm{gyr}[a,b]0 = 0\end{aligned} \tag{42}$$

and hence the result.

13. We obtain Item 13 from 10 with $b = 0$, and a left cancellation, Item 9. □

Einstein addition admits scalar multiplication, giving rise to gyrovector spaces. Studies of gyrogroup theory and gyrovector space theory along with applications in hyperbolic geometry and in Einstein's special theory of relativity are presented in [52, 54, 56, 58–60].

We thus see in this section that Einsteinian velocity addition, $\oplus$, in $\mathbb{R}^3_c$ of relativistically admissible velocities gives rise to the gyrocommutative gyrogroup $(\mathbb{R}^3_c, \oplus)$, just as Newtonian velocity addition, $+$, in $\mathbb{R}^3$ of classical, Newtonian velocities gives rise to the commutative group $(\mathbb{R}^3, +)$. Newtonian velocity addition, in turn, is given by the common vector addition in $\mathbb{R}^3$. Clearly, it is owing to the presence of nontrivial gyrations that gyrogroups are generalized groups. Accordingly, gyrations provide the missing link between the common vector addition and Einstein velocity addition.

8 The Euclidean and Hyperbolic Lines

As shown in Figs. 2 and 3, we introduce Cartesian coordinates into $\mathbb{R}^n$ in the usual way in order to specify uniquely each point P of the Euclidean n-space $\mathbb{R}^n$ by an n-tuple of real numbers, called the coordinates, or components, of P. Cartesian coordinates provide a method of indicating the position of points and rendering graphs on a two-dimensional Euclidean plane $\mathbb{R}^2$ and in a three-dimensional Euclidean space $\mathbb{R}^3$.

As an example, Fig. 2 presents a Euclidean plane $\mathbb{R}^2$ equipped with a Cartesian coordinate system Σ. The position of points A and B and their midpoint M_{AB} with respect to Σ are shown.

The set of all points

$$A + (-A + B)t \tag{43}$$

$t \in \mathbb{R}$, forms a Euclidean line. The segment of this line, corresponding to $1 \le t \le 1$, and a generic point P on the segment, are shown in Fig. 2. Being collinear, the points A, P and B obey the triangle equality $d(A,P) + d(P,B) = d(A,B)$, where $d(A,B) = \| -A + B\|$ is the Euclidean distance function in $\mathbb{R}^n$.

Figure 2 demonstrates the use of the standard Cartesian model of Euclidean geometry for graphical presentations. In a fully analogous way, Fig. 3 demonstrates the use of the Cartesian-Beltrami-Klein model of hyperbolic geometry; see also [60, Figs. 2.3 and 2.4].

Now, let $A, B \in \mathbb{R}^n_s$ be two distinct points of the Einstein gyrovector space $(\mathbb{R}^n_s, \oplus, \otimes)$, and let $t \in \mathbb{R}$ be a real parameter. Then, in full analogy with the

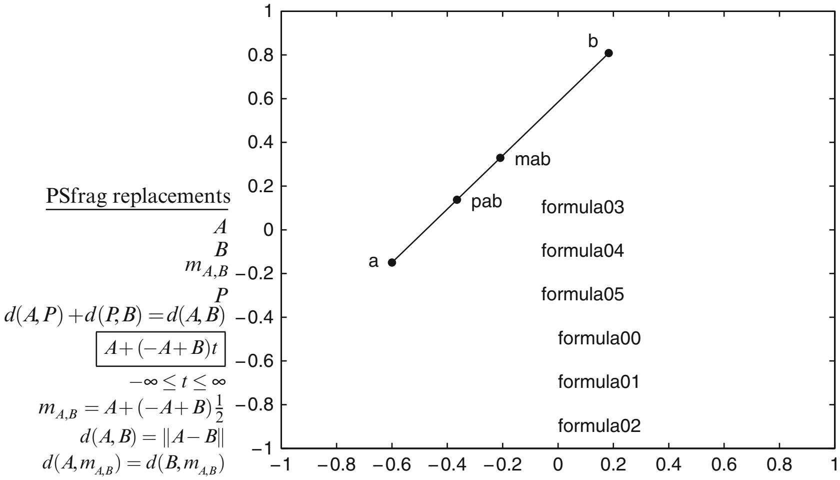


Fig. 2 Cartesian coordinates for the Euclidean plane $\mathbb{R}^2$, (x_1, x_2), $x_1^2 + x_2^2 < \infty$, are shown. The points A and B in the Euclidean plane $\mathbb{R}^2$ are given, with respect to these Cartesian coordinates, by $A = (-0.60, -0.15)$ and $B = (0.18, 0.80)$. The distance function, $d(A, B) = \|A - B\|$, and the equation of the line through A and B are shown along with the triangle equality for a generic point P on the segment AB. The midpoint $m_{A,B}$ of the segment AB is reached when the line parameter is $t = 1/2$

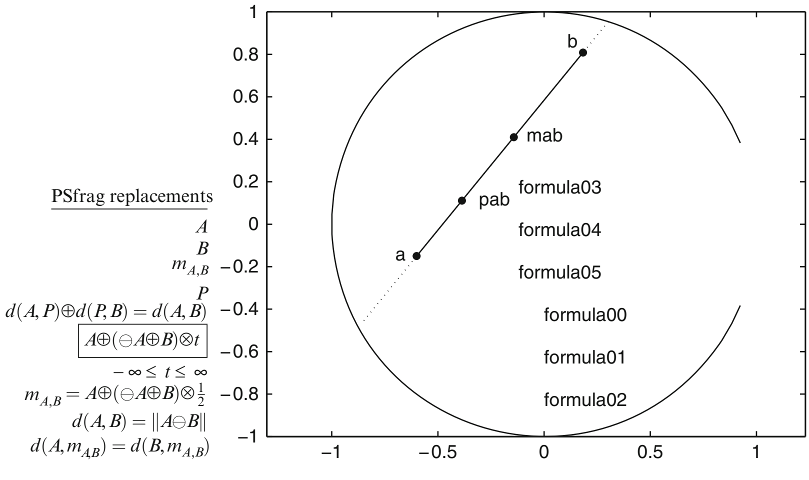


Fig. 3 Cartesian coordinates for the unit disc in the Euclidean plane $\mathbb{R}^2$, (x_1, x_2), $x_1^2 + x_2^2 < 1$, are shown. These, inside the disc, are viewed as Cartesian coordinates for the Einstein gyrovector plane $(\mathbb{R}^2, \oplus, \otimes)$. The points A and B in the Einstein gyrovector plane $(\mathbb{R}^2, \oplus, \otimes)$ are given, with respect to these Cartesian coordinates, by $A = (-0.60, -0.15)$ and $B = (0.18, 0.80)$. The gyrodistance function, $d(A, B) = \|A \ominus B\|$, and the equation of the gyroline through A and B are shown along with the gyrotriangle equality for a generic point P on the gyrosegment AB. The gyromidpoint $m_{A,B}$ of the gyrosegment AB, reached when the gyroline parameter is $t = 1/2$, shares obvious analogies with its Euclidean counterpart, the midpoint in Fig. 2.

Euclidean line (43), the graph of the set of all points, Fig. 3,

$$A \oplus (\ominus A \oplus B) \otimes t \qquad (44)$$

$t \in \mathbb{R}$, in the Einstein gyrovector space $(\mathbb{R}^n_s, \oplus, \otimes)$ is a chord of the ball $\mathbb{R}^n_s$. As such, it is a geodesic line of the Cartesian-Beltrami-Klein ball model of hyperbolic geometry.

The geodesic line (44) is the unique geodesic passing through the points A and B. It passes through the point A when $t = 0$ and, owing to the left cancellation law, (39), it passes through the point B when $t = 1$. Furthermore, it passes through the midpoint $M_{A,B}$ of A and B when $t = 1/2$. Accordingly, the *gyrosegment* that joins the points A and B in Fig. 3 is obtained from gyroline (44) with $0 \le t \le 1$.

9 Thomas Precession

It is owing to the gyrocommutative law, (28), of Einstein addition that Thomas precession of Einstein's special theory of relativity is recognized as a concrete example of the abstract gyrogroup gyration in Definition 4. Accordingly, the gyrogroup gyration is an extension by abstraction of the relativistic mechanical effect known as Thomas precession.

The gyrocommutative law of Einstein velocity addition was already known to Silberstein in 1914 [34] in the following sense: According to his 1914 book, Silberstein knew that the Thomas precession generated by $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3_c$ is the unique rotation that takes $\mathbf{v} \oplus \mathbf{u}$ into $\mathbf{u} \oplus \mathbf{v}$ about an axis perpendicular to the plane of $\mathbf{u}$ and $\mathbf{v}$ through an angle $< \pi$ in $\mathbb{R}^3$, thus giving rise to the gyrocommutative law. However, obviously, Silberstein did not use the terms "Thomas precession" and "gyrocommutative law". These terms have been coined later, respectively, (1) following Thomas' 1926 paper [41], and (2) in 1991 [49, 50], following the discovery of the accompanying gyroassociative law of Einstein addition in 1988 [45].

A description of the 3-space rotation, which since 1926 is named after Thomas, is found in Silberstein's 1914 book [34]. In 1914 Thomas precession did not have a name, and Silberstein called it in his 1914 book a "certain space-rotation" [34, p. 169]. An early study of Thomas precession, made by the famous mathematician Émile Borel in 1913, is described in his 1914 book [2] and, more recently, in [39]. According to Belloni and Reina [1], Sommerfeld's route to Thomas precession dates back to 1909. However, prior to Thomas discovery the relativistic peculiar 3-space rotation had a most uncertain physical status [65, p. 119]. The only knowledge Thomas had in 1925 about the peculiar relativistic gyroscopic precession [19], however, came from De Sitter's formula describing the relativistic corrections for the motion of the moon, found in Eddington's book [5], which was just published at that time [52, Sect. 1, Chap. 1].

The physical significance of the peculiar rotation in special relativity emerged in 1925 when Thomas relativistically re-computed the precessional frequency of the doublet separation in the fine structure of the atom, and thus rectified a missing factor of 1/2. This correction has come to be known as the *Thomas half* [4], presented in (101). Thomas' discovery of the relativistic precession of the electron spin on Christmas 1925 thus led to the understanding of the significance of the relativistic effect which became known as *Thomas precession*. Llewellyn Hilleth Thomas died in Raleigh, NC, on April 20, 1992. A paper [3] dedicated to the centenary of the birth of Llewellyn H. Thomas (1902–1992) describes the Bloch gyrovector of quantum information and computation.

Once recognized as gyration, it is clear that Thomas precession owes its existence solely to the nonassociativity of Einstein addition of Einsteinian velocities. Accordingly, Thomas precession has no classical counterpart since the addition of classical, Newtonian velocities is associative.

It is widely believed that special relativistic effects are negligible when the velocities involved are much less than the vacuum speed of light c. Yet, Thomas precession effect in the orbital motion of spinning electrons in atoms is clearly observed in resulting spectral lines despite the speed of electrons in atoms being small compared with the speed of light. One may, therefore, ask whether it is possible to furnish a classical background to Thomas precession [23]. Hence, it is important to realize that Thomas precession stems from the nonassociativity of Einsteinian velocity addition, so that it has no echo in Newtonian velocities.

In 1966, Ehlers, Rindler and Robinson [6] proposed a new formalism for dealing with the Lorentz group. Their formalism, however, did not find its way to the mainstream literature. Therefore, 33 years later, two of them suggested considering the "notorious Thomas precession formula" (in their words [31, p. 431]) as an indicator of the quality of a formalism for dealing with the Lorentz group. The idea of Rindler and Robinson to use the "notorious Thomas precession formula" as an indicator works fine in our analytic hyperbolic geometric viewpoint of special relativity [56], where the ugly duckling of special relativity, the "notorious Thomas precession formula", becomes the beautiful swan of special relativity and its underlying analytic hyperbolic geometry. The abstract Thomas precession, called gyration, is now recognized as the missing link between classical mechanics with its underlying Euclidean geometry and relativistic mechanics with its underlying hyperbolic geometry.

10 Thomas Precession Matrix

For any two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$, $\mathbf{a} = (a_1, a_2, a_3)$, *etc.*, determined by their components with respect to a given Cartesian coordinate system, we define the square 3×3 matrix $\Omega(\mathbf{a}, \mathbf{b})$ by the equation

$$\Omega(\mathbf{a},\mathbf{b}) = -\begin{pmatrix} a_1b_1 & a_1b_2 & a_1b_3 \\ a_2b_1 & a_2b_2 & a_2b_3 \\ a_3b_1 & a_3b_2 & a_3b_3 \end{pmatrix} + \begin{pmatrix} a_1b_1 & a_2b_1 & a_3b_1 \\ a_1b_2 & a_2b_2 & a_3b_2 \\ a_1b_3 & a_2b_3 & a_3b_3 \end{pmatrix} \tag{45}$$

or, equivalently,

$$\Omega(\mathbf{a},\mathbf{b}) = -\begin{pmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{pmatrix} \tag{46}$$

where

$$\mathbf{w} = (\omega_1,\, \omega_2,\, \omega_3,\,) = \mathbf{a}\times\mathbf{b} \tag{47}$$

Accordingly,

$$\Omega(\mathbf{a},\mathbf{b})\mathbf{x} = (\mathbf{a}\times\mathbf{b})\times\mathbf{x} = -\mathbf{a}(\mathbf{b}\cdot\mathbf{x}) + \mathbf{b}(\mathbf{a}\cdot\mathbf{x}) \tag{48}$$

for any $\mathbf{x}\in\mathbb{R}^3$. Hence,

1. $\Omega(\mathbf{a},\mathbf{b}) = 0$ if and only if $\mathbf{a}\times\mathbf{b} = \mathbf{0}$;
2. and
$$\Omega(\mathbf{a},\mathbf{b})(\mathbf{a}\times\mathbf{b}) = \mathbf{0} \tag{49}$$
3. and, for $\Omega = \Omega(\mathbf{a},\mathbf{b})$,
$$\Omega^3 = -(\mathbf{a}\times\mathbf{b})^2\Omega \tag{50}$$

The matrix $\Omega = \Omega(\mathbf{u},\mathbf{v})$ can be used to simplify the presentation of both Einstein addition $\mathbf{u}\oplus\mathbf{v}$ and its associated gyration gyr[$\mathbf{u},\mathbf{v}$],

$$\mathbf{u}\oplus\mathbf{v} = \frac{1}{1+\frac{\mathbf{u}\cdot\mathbf{v}}{c^2}}\left\{\mathbf{u}+\mathbf{v}-\frac{1}{c^2}\frac{\gamma_{\mathbf{u}}}{1+\gamma_{\mathbf{u}}}\Omega\mathbf{u}\right\} \tag{51}$$

$$\mathrm{gyr}[\mathbf{u},\mathbf{v}] = I + \alpha\Omega + \beta\Omega^2 \tag{52}$$

where I is the 3×3 identity matrix, and where

$$\begin{aligned} \alpha = \alpha(\mathbf{u},\mathbf{v}) &= -\frac{1}{c^2}\frac{\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}(1+\gamma_{\mathbf{u}}+\gamma_{\mathbf{v}}+\gamma_{\mathbf{u}\oplus\mathbf{v}})}{(1+\gamma_{\mathbf{u}})(1+\gamma_{\mathbf{v}})(1+\gamma_{\mathbf{u}\oplus\mathbf{v}})} \\ \beta = \beta(\mathbf{u},\mathbf{v}) &= \quad \frac{1}{c^4}\frac{\gamma_{\mathbf{u}}^2\gamma_{\mathbf{v}}^2}{(1+\gamma_{\mathbf{u}})(1+\gamma_{\mathbf{v}})(1+\gamma_{\mathbf{u}\oplus\mathbf{v}})} \end{aligned} \tag{53}$$

satisfying $\alpha < 0$, $\beta > 0$, and

$$\alpha^2 + [\mathbf{u}^2\mathbf{v}^2 - (\mathbf{u}\cdot\mathbf{v})^2]\beta^2 - 2\beta = 0 \tag{54}$$

for all $\mathbf{u},\mathbf{v}\in\mathbb{R}^3_c$.

The gyration matrix gyr[**u**, **v**] in (52) satisfies the cubic equation

$$\begin{aligned}\mathrm{gyr}^3[\mathbf{u},\mathbf{v}] - \mathrm{trace}(\mathrm{gyr}[\mathbf{u},\mathbf{v}])\mathrm{gyr}^2[\mathbf{u},\mathbf{v}]&\\ + \mathrm{trace}(\mathrm{gyr}[\mathbf{u},\mathbf{v}])\mathrm{gyr}[\mathbf{u},\mathbf{v}] - I &= 0\end{aligned} \tag{55}$$

called the *trace identity*.

The trace identity (55) characterizes 3×3 matrices that represent proper rotations of the Euclidean 3-space $\mathbb{R}^3$ about its origin.

The matrix representation of gyr[**u**, **v**] in $\mathbb{R}^3$ relative to an orthonormal basis is thus an orthogonal 3×3 matrix with determinant 1. It follows from (49) and (52) that

$$\mathrm{gyr}[\mathbf{u},\mathbf{v}](\mathbf{u}\times\mathbf{v}) = \mathbf{u}\times\mathbf{v} \tag{56}$$

so that the vector $\mathbf{u}\times\mathbf{v}$ lies on the rotation axis of the gyration gyr[**u**, **v**]. Accordingly, the gyrocommutative law in (28) implies that the gyration gyr[**u**, **v**] generated by two non-parallel, non-zero vectors $\mathbf{u},\mathbf{v}\in\mathbb{R}^3_c$ is the rotation that rotates the vector $\mathbf{v}\oplus\mathbf{u}$ into the vector $\mathbf{u}\oplus\mathbf{v}$ through a rotation axis perpendicular to the plane spanned by **u** and **v** by an angle $< \pi$.

Interesting studies of the trace identity, using analysis, algebra and geometry is found in an elementary form in [20] and in a more advanced form in [14–17].

11 Thomas Precession Graphical Presentation

Let Σ'', Σ' and Σ be three inertial frames in the Euclidean 3-space $\mathbb{R}^3$ with respective spatial coordinates (x'', y''), (x', y') and (x, y). The third spatial coordinate of each frame is omitted for simplicity. Accordingly, these are shown in Fig. 4 in $\mathbb{R}^2$ rather than $\mathbb{R}^3$. Frame Σ'' moves with velocity $\mathbf{v}\in\mathbb{R}^3_c$, without rotation, relative to frame Σ' which, in turn, moves with velocity $\mathbf{u}\in\mathbb{R}^3_c$, without rotation, relative to frame Σ. The angle between **u** and **v** is θ, shown in Fig. 4, satisfying

$$\cos\theta = \frac{\mathbf{u}}{\|\mathbf{u}\|}\cdot\frac{\mathbf{v}}{\|\mathbf{v}\|} \tag{57}$$

so that, by (10),

$$\frac{1}{c^2}\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}\mathbf{u}\cdot\mathbf{v} = \sqrt{\gamma_{\mathbf{u}}^2 - 1}\sqrt{\gamma_{\mathbf{v}}^2 - 1}\cos\theta \tag{58}$$

Observers at rest relative to Σ and observers at rest relative to Σ' agree that their coordinates (x, y) and (x', y') are parallel. Similarly, observers at rest relative to Σ' and observers at rest relative to Σ'' agree that their coordinates (x', y') and (x'', y'') are parallel, as shown in the left part of Fig. 4.

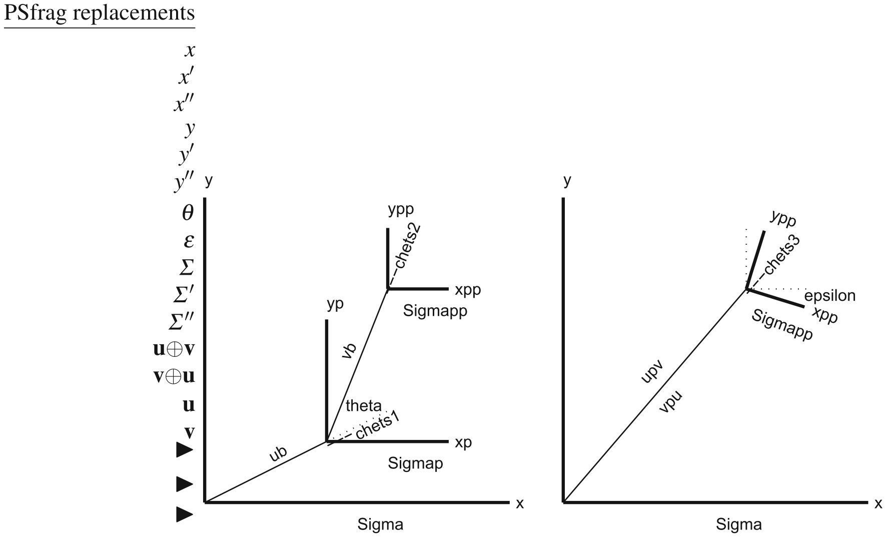


Fig. 4 In the Euclidean plane $\mathbb{R}^2$ an inertial frame Σ'' moves uniformly, without rotation, with velocity $\mathbf{v} \in \mathbb{R}_s^2$ relative to inertial frame Σ'. The latter, in turn, moves uniformly, without rotation, with velocity $\mathbf{u} \in \mathbb{R}_s^2$ relative to inertial frame Σ. Owing to the presence of Thomas precession, the inertial frame Σ'' moves uniformly, with rotation angle ϵ, with a composite velocity relative to the inertial frame Σ. Is the composite velocity of Σ'' relative to Σ $\mathbf{u}\oplus\mathbf{v}$ or $\mathbf{v}\oplus\mathbf{u}$? The answer is: neither; see (81). The Thomas precession signed angle ϵ, $-\pi < \epsilon < \pi$, turns out to be the unique rotation angle with rotation axis parallel to $\mathbf{u} \times \mathbf{v}$ in $\mathbb{R}_c^3$ that takes $\mathbf{v}\oplus\mathbf{u}$ into $\mathbf{u}\oplus\mathbf{v}$ according to the gyrocommutative law $\mathbf{u}\oplus\mathbf{v} = \mathrm{gyr}[\mathbf{u}, \mathbf{v}](\mathbf{v}\oplus\mathbf{u})$. Being related by (61), the Thomas precession signed angle ϵ and its generating signed angle θ from $\mathbf{u}$ and $\mathbf{v}$ have opposite signs, illustrated graphically in Figs. 5–6

Counterintuitively, if $\theta \neq 0$ and $\theta \neq \pi$, observers at rest relative to Σ and observers at rest relative to Σ'' agree that their coordinates are not parallel. Rather, they find that their coordinates are oriented relative to each other by a Thomas precession angle ϵ, $0 < \epsilon < \pi$, as shown in the right part of Fig. 4.

Let $\mathbf{u}$ and $\mathbf{v}$ be two nonzero vectors in the ball $\mathbb{R}_c^3$. By the gyrocommutative law in (28), the gyration $\mathrm{gyr}[\mathbf{u}, \mathbf{v}]$ takes the composite velocity $\mathbf{v}\oplus\mathbf{u}$ into $\mathbf{u}\oplus\mathbf{v}$. Indeed, $\mathrm{gyr}[\mathbf{u}, \mathbf{v}]$ is the unique rotation with rotation axis parallel to $\mathbf{u} \times \mathbf{v}$ that takes $\mathbf{v}\oplus\mathbf{u}$ into $\mathbf{u}\oplus\mathbf{v}$ through the gyration angle ϵ, $0 \leq \epsilon < \pi$. We call ϵ the Thomas precession (or, rotation) angle of the gyration $\mathrm{gyr}[\mathbf{u}, \mathbf{v}]$, and use the notation

$$\epsilon = \angle\mathrm{gyr}[\mathbf{u}, \mathbf{v}] \tag{59}$$

Accordingly, the Thomas precession angle $\epsilon = \angle\mathrm{gyr}[\mathbf{u}, \mathbf{v}]$ generated by $\mathbf{u}, \mathbf{v} \in \mathbb{R}_c^3$, shown in the right part of Fig. 4, satisfies the equations

$$\begin{aligned} \cos\epsilon &= \frac{(\mathbf{u}\oplus\mathbf{v}) \cdot (\mathbf{v}\oplus\mathbf{u})}{\|\mathbf{u}\oplus\mathbf{v}\|^2} \\ \sin\epsilon &= \pm\frac{\|(\mathbf{u}\oplus\mathbf{v}) \times (\mathbf{v}\oplus\mathbf{u})\|}{\|\mathbf{u}\oplus\mathbf{v}\|^2} \end{aligned} \tag{60}$$

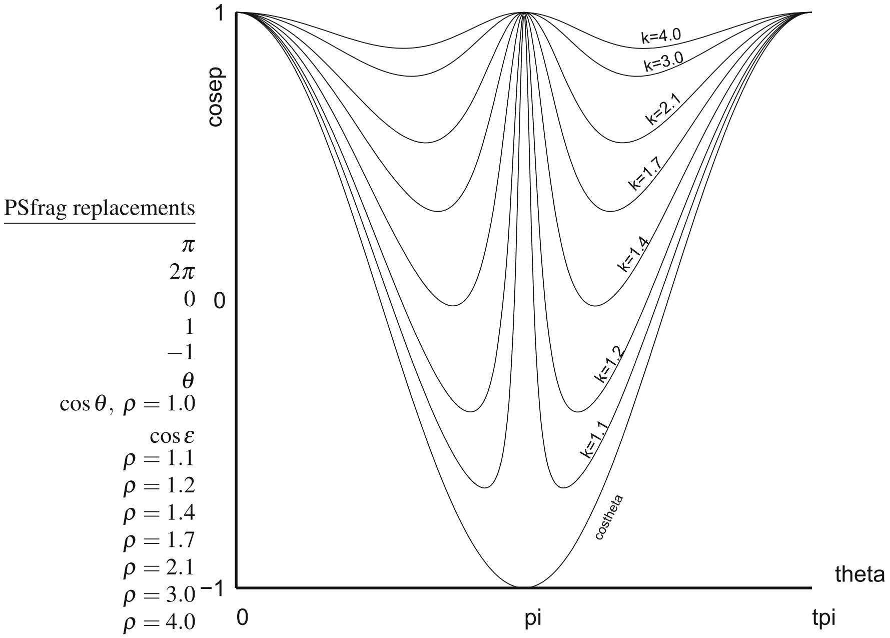


Fig. 5 A graphical presentation of the cosine of the Thomas precession angle ϵ, $\cos\epsilon$, (61), as a function of the angle θ between its two generating relativistically admissible velocities $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ for several values of ρ, ρ being a function, (62), of $\gamma_{\mathbf{u}}$ and $\gamma_{\mathbf{v}}$

where the ambigious sign of $\sin\epsilon$ is determined in (61). Indeed, the sign of $\sin\epsilon$ is selected to be opposite to that of $\sin\theta$, as shown in (61) below.

The Herculean task of simplifying (60) was accomplished in [45–47, 49], obtaining

$$\begin{aligned} \cos\epsilon &= \frac{(\rho + \cos\theta)^2 - \sin^2\theta}{(\rho + \cos\theta)^2 + \sin^2\theta} \\ \sin\epsilon &= \frac{-2(\rho + \cos\theta)\,\sin\theta}{(\rho + \cos\theta)^2 + \sin^2\theta} \end{aligned} \tag{61}$$

where θ, $0 \le \theta < 2\pi$, is the angle between the vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$, forming the horizontal axes in Figs. 5–6, and where ρ, $\rho > 1$, is a velocity parameter given by the equation

$$\rho^2 = \frac{\gamma_{\mathbf{u}} + 1}{\gamma_{\mathbf{u}} - 1}\frac{\gamma_{\mathbf{v}} + 1}{\gamma_{\mathbf{v}} - 1} \tag{62}$$

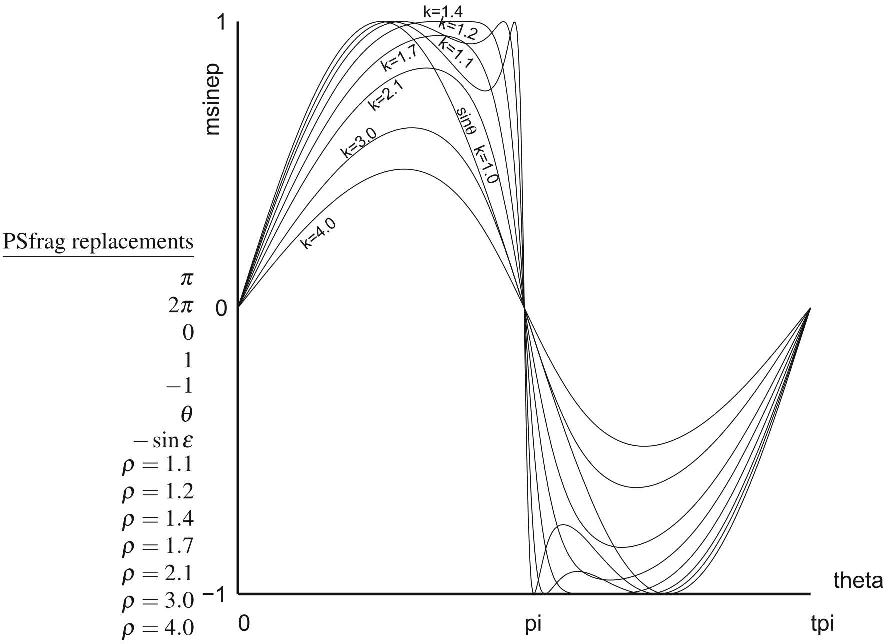


Fig. 6 A graphical presentation of the negative sine of the Thomas precession angle ϵ, $-\sin\epsilon$, (61), as a function of the angle θ between its two generating relativistically admissible velocities $\mathbf{u}, \mathbf{v} \in \mathbb{R}^3$ for several values of ρ, ρ being a function, (62), of $\gamma_{\mathbf{u}}$ and $\gamma_{\mathbf{v}}$

The parameter ρ approaches 1 when both $\|\mathbf{u}\|$ and $\|\mathbf{v}\|$ approach c. We clearly have the limits

$$\begin{aligned} \lim_{\rho\to 1} \cos\epsilon &= \cos\theta \\ \lim_{\rho\to 1} \sin\epsilon &= -\sin\theta \end{aligned} \tag{63}$$

for $0 \le \theta \le 2\pi$, $\theta \ne \pi$, seen in Figs. 5 and 6.

Figures 5 and 6 present graphically $\cos\epsilon$ and $-\sin\epsilon$ as functions of θ for several values of ρ. As expected, the graphs in these figures show that for all values of the parameter ρ, $\rho > 1$, Thomas precession angle ϵ vanishes when $\theta = 0$, when $\theta = \pi$, and again, when $\theta = 2\pi$. In the limit of high relativistic speeds approaching the vacuum speed of light c, $\|\mathbf{u}\|, \|\mathbf{v}\| \to c$, the parameter ρ approaches unity, $\rho \to 1$, and $\epsilon \to -\theta$ for all θ in the punctured interval $[0, \pi) \bigcup (\pi, 2\pi]$. The punctured interval is the union of the two connected intervals $[0, \pi)$ and $(\pi, 2\pi]$ which is the closed connected interval $[0, 2\pi]$ from which the point π has been removed. Thus, there is no Thomas precession angle π, that is, $\epsilon \ne \pi$; see also (75).

The extension by abstraction of Thomas precession into gyration enables the development of techniques that explain the non-existence of a gyration whose rotation angle is π; see the Gyration Exclusion Theorem in [56, Theorem 3.36].

As we see from Figs. 5 and 6, the variation of ϵ for $0 \leq \theta \leq 2\pi$ is over the interval $[0,\ 2\pi]$ punctured by a ρ-dependent subinterval centered at $\epsilon = \pi$.

It is interesting to derive $\cos\frac{\epsilon}{2}$ and $\sin\frac{\epsilon}{2}$ from (61):

$$\begin{aligned}\cos\frac{\epsilon}{2} &= \pm\sqrt{\frac{1+\cos\epsilon}{2}} = \frac{\rho+\cos\theta}{\sqrt{(\rho+\cos\theta)^2+\sin^2\theta}} \\ \sin\frac{\epsilon}{2} &= \pm\sqrt{\frac{1-\cos\epsilon}{2}} = -\frac{\sin\theta}{\sqrt{(\rho+\cos\theta)^2+\sin^2\theta}}\end{aligned} \tag{64}$$

so that

$$\tan\frac{\epsilon}{2} = \frac{-\sin\theta}{\rho+\cos\theta} \tag{65}$$

Equation 65 modulo -1 is found, for instance, in [44, Eq. 6.37, p. 171]. The validity and the importance of the negative sign in (65) is explained in detail in Sect. 15.

Following Fig. 4, the ambiguous signs in (64) are selected such that $\cos\frac{\epsilon}{2} > 0$ while $\sin\frac{\epsilon}{2}$ and $\sin\theta$ have opposite signs.

12 Thomas Precession Angle

Thomas precession gyr[**u**, **v**] in (52) can be recast into a form familiar as the representation of a rotation about an axis by an angle ϵ,

$$\mathrm{gyr}[\mathbf{u},\mathbf{v}] = \begin{cases} I + \sin\epsilon\,\frac{\Omega(\mathbf{u},\mathbf{v})}{\omega_\theta} + (1-\cos\epsilon)\frac{\Omega^2(\mathbf{u},\mathbf{v})}{\omega_\theta^2}, & \omega_\theta \neq 0 \\ I, & \omega_\theta = 0 \end{cases} \tag{66}$$

where $\mathbf{u},\mathbf{v} \in \mathbb{R}_c^3$, and where ϵ is the Thomas precession angle shown in Fig. 4.

Comparing (66) with (52), we see that

$$\begin{aligned}\sin\epsilon &= \alpha(\mathbf{u},\mathbf{v})\omega_\theta \\ 1-\cos\epsilon &= \beta(\mathbf{u},\mathbf{v})\omega_\theta\end{aligned} \tag{67}$$

and

$$\begin{aligned}\omega_\theta &= \pm\|\mathbf{u}\times\mathbf{v}\| \\ &= \|\mathbf{u}\|\,\|\mathbf{v}\|\sin\theta \\ &= c^2\frac{\sqrt{\gamma_{\mathbf{u}}^2-1}\sqrt{\gamma_{\mathbf{v}}^2-1}}{\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}}\sin\theta\end{aligned} \tag{68}$$

where the ambiguous sign is selected such that ω_θ and $\sin\theta$ have equal signs.

It follows from (67) and (68), and from the definition of $\alpha(\mathbf{u},\mathbf{v})$ and $\beta(\mathbf{u},\mathbf{v})$ in (53) that

$$\begin{aligned}\cos\epsilon &= 1-\frac{(\gamma_{\mathbf{u}}-1)(\gamma_{\mathbf{v}}-1)}{\gamma_{\mathbf{u}\oplus\mathbf{v}}+1}\sin^2\theta\\ \sin\epsilon &= -\frac{\sqrt{\gamma_{\mathbf{u}}^2-1}\sqrt{\gamma_{\mathbf{v}}^2-1}+(\gamma_{\mathbf{u}}-1)(\gamma_{\mathbf{v}}-1)\cos\theta}{\gamma_{\mathbf{u}\oplus\mathbf{v}}+1}\sin\theta\end{aligned}\tag{69}$$

Following (9)–(11) and (58) we have

$$\gamma_{\mathbf{u}\oplus\mathbf{v}}=\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}+\sqrt{\gamma_{\mathbf{u}}^2-1}\sqrt{\gamma_{\mathbf{v}}^2-1}\;\cos\theta\tag{70a}$$

and

$$\gamma_{\ominus\mathbf{u}\oplus\mathbf{v}}=\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}-\sqrt{\gamma_{\mathbf{u}}^2-1}\sqrt{\gamma_{\mathbf{v}}^2-1}\;\cos\theta\tag{70b}$$

so that, by (69) and (70a)

$$\begin{aligned}\cos\epsilon &= 1-\frac{(\gamma_{\mathbf{u}}-1)(\gamma_{\mathbf{v}}-1)}{1+\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}+\sqrt{\gamma_{\mathbf{u}}^2-1}\sqrt{\gamma_{\mathbf{v}}^2-1}\cos\theta}\sin^2\theta\\ \sin\epsilon &= -\frac{(\gamma_{\mathbf{u}}-1)(\gamma_{\mathbf{v}}-1)(\rho+\cos\theta)}{1+\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}+\sqrt{\gamma_{\mathbf{u}}^2-1}\sqrt{\gamma_{\mathbf{v}}^2-1}\cos\theta}\sin\theta\end{aligned}\tag{71}$$

where $\rho > 1$ is given by (62).

The special case when $\mathbf{u}$ and $\mathbf{v}$ have equal magnitudes is required for later reference related to Fig. 4. In this special case $\gamma_{\mathbf{u}}=\gamma_{\mathbf{v}}$, so that ϵ in (71) reduces to ϵ_s given by

$$\begin{aligned}\cos\epsilon_s &= 1-\frac{(\gamma_{\mathbf{v}}-1)^2\sin^2\theta}{1+\gamma_{\mathbf{v}}^2+(\gamma_{\mathbf{v}}^2-1)\cos\theta}\\ \sin\epsilon_s &= -\frac{(\gamma_{\mathbf{v}}^2-1)+(\gamma_{\mathbf{v}}-1)^2\cos\theta}{1+\gamma_{\mathbf{v}}^2+(\gamma_{\mathbf{v}}^2-1)\cos\theta}\sin\theta\end{aligned}\tag{72}$$

Solving (70) for $\cos\theta$ we obtain the equations

$$\begin{aligned}\cos\theta &= \frac{\gamma_{\mathbf{u}\oplus\mathbf{v}}-\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}}{\sqrt{\gamma_{\mathbf{u}}^2-1}\sqrt{\gamma_{\mathbf{v}}^2-1}}=\frac{-\gamma_{\ominus\mathbf{u}\oplus\mathbf{v}}+\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}}{\sqrt{\gamma_{\mathbf{u}}^2-1}\sqrt{\gamma_{\mathbf{v}}^2-1}}\\ \sin^2\theta &= 1-\cos^2\theta\\ &= \frac{1-\gamma_{\mathbf{u}}^2-\gamma_{\mathbf{v}}^2-\gamma_{\mathbf{u}\oplus\mathbf{v}}^2+2\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}\gamma_{\mathbf{u}\oplus\mathbf{v}}}{(\gamma_{\mathbf{u}}^2-1)(\gamma_{\mathbf{v}}^2-1)}\end{aligned}\tag{73}$$

The substitution of (73) into (69) gives

$$\begin{aligned}\cos\epsilon &= \frac{1}{(\gamma_{\mathbf{u}}+1)(\gamma_{\mathbf{v}}+1)(\gamma_{\mathbf{u}\oplus\mathbf{v}}+1)}\\ &\quad\times\{-\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}\gamma_{\mathbf{u}\oplus\mathbf{v}}+\gamma_{\mathbf{u}}^2+\gamma_{\mathbf{v}}^2+\gamma_{\mathbf{u}\oplus\mathbf{v}}^2\\ &\quad+\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}+\gamma_{\mathbf{u}}\gamma_{\mathbf{u}\oplus\mathbf{v}}+\gamma_{\mathbf{v}}\gamma_{\mathbf{u}\oplus\mathbf{v}}+\gamma_{\mathbf{u}}+\gamma_{\mathbf{v}}+\gamma_{\mathbf{u}\oplus\mathbf{v}}\}\end{aligned} \tag{74}$$

so that, finally, we obtain the elegant expression

$$1+\cos\epsilon = \frac{(1+\gamma_{\mathbf{u}}+\gamma_{\mathbf{v}}+\gamma_{\mathbf{u}\oplus\mathbf{v}})^2}{(1+\gamma_{\mathbf{u}})(1+\gamma_{\mathbf{v}})(1+\gamma_{\mathbf{u}\oplus\mathbf{v}})} > 0 \tag{75}$$

which agrees with McFarlane's result, cited in [33, Eq. 2.10.7]. It implies that $\epsilon \neq \pi$ for all $\mathbf{u},\mathbf{v}\in\mathbb{R}_c^3$; and that

$$\cos\frac{\epsilon}{2} = \sqrt{\frac{1+\cos\epsilon}{2}} = \frac{1+\gamma_{\mathbf{u}}+\gamma_{\mathbf{v}}+\gamma_{\mathbf{u}\oplus\mathbf{v}}}{\sqrt{2}\sqrt{1+\gamma_{\mathbf{u}}}\sqrt{1+\gamma_{\mathbf{v}}}\sqrt{1+\gamma_{\mathbf{u}\oplus\mathbf{v}}}} \tag{76}$$

Consequently, we also have the elegant identity

$$\tan^2\frac{\epsilon}{2} = \left(\frac{\sin\epsilon}{1+\cos\epsilon}\right)^2 = \frac{1+2\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}\gamma_{\mathbf{u}\oplus\mathbf{v}}-\gamma_{\mathbf{u}}^2-\gamma_{\mathbf{v}}^2-\gamma_{\mathbf{u}\oplus\mathbf{v}}^2}{(1+\gamma_{\mathbf{u}}+\gamma_{\mathbf{v}}+\gamma_{\mathbf{u}\oplus\mathbf{v}})^2} \tag{77}$$

Hence, (59), the Thomas precession angle $\epsilon = \angle\mathrm{gyr}[\mathbf{u},\mathbf{v}]$ in Fig. 4 is given by the equation

$$\tan^2\frac{\angle\mathrm{gyr}[\mathbf{u},\mathbf{v}]}{2} = \frac{1+2\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}\gamma_{\mathbf{u}\oplus\mathbf{v}}-\gamma_{\mathbf{u}}^2-\gamma_{\mathbf{v}}^2-\gamma_{\mathbf{u}\oplus\mathbf{v}}^2}{(1+\gamma_{\mathbf{u}}+\gamma_{\mathbf{v}}+\gamma_{\mathbf{u}\oplus\mathbf{v}})^2} \tag{78}$$

Noting the gyration even property in (28) and replacing $\mathbf{u}$ by $\ominus\mathbf{u}$ in (78), we obtain the equations

$$\begin{aligned}\tan^2\frac{\angle\mathrm{gyr}[\mathbf{u},\ominus\mathbf{v}]}{2} &= \tan^2\frac{\angle\mathrm{gyr}[\ominus\mathbf{u},\mathbf{v}]}{2}\\ &= \frac{1+2\gamma_{\mathbf{u}}\gamma_{\mathbf{v}}\gamma_{\ominus\mathbf{u}\oplus\mathbf{v}}-\gamma_{\mathbf{u}}^2-\gamma_{\mathbf{v}}^2-\gamma_{\ominus\mathbf{u}\oplus\mathbf{v}}^2}{(1+\gamma_{\mathbf{u}}+\gamma_{\mathbf{v}}+\gamma_{\ominus\mathbf{u}\oplus\mathbf{v}})^2}\\ &= \tan^2\tfrac{\delta}{2}\end{aligned} \tag{79}$$

so that

$$\angle\text{gyr}[\mathbf{u}, \ominus\mathbf{v}] = \delta \tag{80}$$

The extreme right-hand side of (79) follows from (20) and (21).

Interestingly, it follows from (80) that the Thomas precession angle generated by $\mathbf{u}$ and $\ominus\mathbf{v}$, that is, $\angle\text{gyr}[\mathbf{u}, \ominus\mathbf{v}]$, possesses an important hyperbolic geometric property [56, 64]. It equals the defect δ of the gyrotriangle generated by $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^3_c$, shown in Fig. 1; see also [52, pp. 236–237] and the Gyration–Defect Theorem in [56, Theorem 8.55, p. 317].

The gyration $\text{gyr}[\mathbf{u}, \ominus\mathbf{v}]$ possesses an important gyroalgebraic property as well. It gives rise to a second binary operation $\boxplus$, called *Einstein coaddition*, given by the equation

$$\mathbf{u} \boxplus \mathbf{v} = \mathbf{u} \oplus \text{gyr}[\mathbf{u}, \ominus\mathbf{v}]\mathbf{v} \tag{81}$$

which can be dualized into the equation [56, Theorem 2.14]

$$\mathbf{u} \oplus \mathbf{v} = \mathbf{u} \boxplus \text{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{v} \tag{82}$$

Unlike Einstein addition, which is gyrocommutative, Einstein coaddition is commutative. Furthermore, it possesses a geometric interpretation as a *gyroparallelogram addition law*, and it gives rise to the two mutually dual right cancellation laws [56]

$$\begin{aligned} (\mathbf{v} \oplus \mathbf{u}) \boxminus \mathbf{u} &= \mathbf{v} \\ (\mathbf{v} \boxplus \mathbf{u}) \ominus \mathbf{u} &= \mathbf{v} \end{aligned} \tag{83}$$

Einstein coaddition $\boxplus$ captures useful analogies with classical results, two of which are the right cancellation laws in (83). Another important analogy that Einstein coaddition captures is associated with the gyromidpoint. Indeed, the midpoint $M_{A,B}$ shown in Fig. 2 is expressed in terms of the points A and B by the equation

$$M_{A,B} = A + (-A + B)\tfrac{1}{2} = \tfrac{1}{2}(A + B) \tag{84}$$

while, in full analogy, the gyromidpoint $M_{A,B}$ shown in Fig. 3 is expressed in terms of the points A and B by the equation [56]

$$M_{A,B} = A \oplus (\ominus A \oplus B) \otimes \tfrac{1}{2} = \tfrac{1}{2} \otimes (A \boxplus B) \tag{85}$$

The right part of Fig. 4 raises the question as to whether the composite velocity of frame Σ'' relative to frame Σ is $\mathbf{u} \oplus \mathbf{v}$ or $\mathbf{v} \oplus \mathbf{u}$. The answer is that the composite velocity of frame Σ'' relative to frame Σ is neither $\mathbf{u} \oplus \mathbf{v}$ nor $\mathbf{v} \oplus \mathbf{u}$. Rather, it is given by the commutative composite velocity $\mathbf{u} \boxplus \mathbf{v}$. Indeed, it is demonstrated in [60, Chap. 10–Epilogue], and in more details in [56, Chap. 13], that looking at the relativistic velocity addition law and its underlying hyperbolic geometry through the lens of the cosmological stellar aberration effect leads to a startling conclusion: relativistic velocities are gyrovectors that add in the cosmos according to the gyroparallelogram addition law of hyperbolic geometry, that is, according to the commutative addition $\mathbf{u} \boxplus \mathbf{v}$, rather than either Einstein addition $\mathbf{u}\oplus\mathbf{v}$ or $\mathbf{v}\oplus\mathbf{u}$.

13 Thomas Precession Frequency

Let us consider a spinning spherical object moving with velocity $\mathbf{v}$ of uniform magnitude $v = \|\mathbf{v}\|$ along a circular path in some inertial frame Σ. We assume that the spin axis lies in the plane containing the circular orbit, as shown in Fig. 7. The spinning object acts like a gyroscope, maintaining the direction of its spin axis in the transition from one inertial frame into another one, as seen by inertial observers moving instantaneously with the accelerated object. Following Taylor and Wheeler, we approximate the circular path by a regular polygon of n sides [40], as shown in Fig. 7 for $n = 8$. In moving once around this orbit the object moves with uniform velocity $\mathbf{v}$ in straight-line paths interrupted by n sudden changes of direction, each through an angle $\theta_n = 2\pi/n$.

An observer at rest relative to the laboratory frame Σ views the motion of the object along the polygonal path as the result of successive boosts (A boost being a Lorentz transformation without rotation [48]); see Sect. 14. He therefore measures a Thomas precession angle ϵ_n by which the object spin axis is precessed when the object rounds a corner. By (72), this Thomas precession angle ϵ_n is determined by the equations

$$\begin{aligned} \cos\epsilon_n &= 1 - \frac{(\gamma_{\mathbf{v}} - 1)^2 \sin^2 \frac{2\pi}{n}}{(\gamma_{\mathbf{v}}^2 + 1) + (\gamma_{\mathbf{v}}^2 - 1)\cos\frac{2\pi}{n}} \\ \sin\epsilon_n &= -\frac{(\gamma_{\mathbf{v}}^2 - 1) + (\gamma_{\mathbf{v}} - 1)^2 \cos\frac{2\pi}{n}}{(\gamma_{\mathbf{v}}^2 + 1) + (\gamma_{\mathbf{v}}^2 - 1)\cos\frac{2\pi}{n}} \sin\frac{2\pi}{n} \end{aligned} \tag{86}$$

By Euler's equation we have

$$\begin{aligned} e^{i\epsilon_n} &= \cos\epsilon_n + i\sin\epsilon_n \\ &= 1 + f(\frac{2\pi}{n}) \end{aligned} \tag{87}$$

where $i = \sqrt{-1}$ and

$$f(\phi) = -\frac{(\gamma_{\mathbf{v}} - 1)^2 \sin\phi + i\{(\gamma_{\mathbf{v}}^2 - 1) + (\gamma_{\mathbf{v}} - 1)^2 \cos\phi\}}{2 + (\gamma_{\mathbf{v}}^2 - 1)(1 + \cos\phi)} \sin\phi \tag{88}$$

$\phi \in \mathbb{R}$.

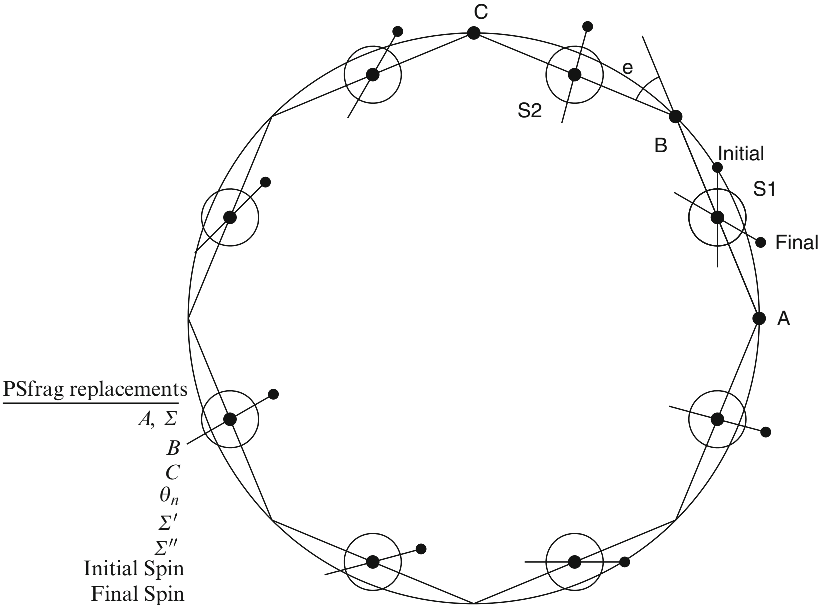


Fig. 7 A regular polygonal path in $\mathbb{R}^3$ as an approximation to the Newtonian circular path of a spinning spherical object. The change of direction at each vertex of the polygon is $\theta_n = 2\pi/n$, where n is the number of the polygon sides. Here, $n = 8$. In the limit $n \to \infty$, the polygonal path tends to the circular path. A spinning spherical object is moving with velocity of uniform magnitude along the polygonal path. The points $A, B, C \in \mathbb{R}^3$ are three adjacent vertices of the polygon in the rest (laboratory) frame Σ. When the object moves from A to B it is at rest relative to the frame Σ', and when the object moves from B to C it is at rest relative to the frame Σ''. The relationship between the three inertial frames Σ, Σ' and Σ'' is thus the one shown in Fig. 4 with $\theta = \theta_n$. Accordingly, since the object moves in the counterclockwise direction, it precesses in the clockwise direction.
Initially, the spin of the object is vertical when the object moves uniformly from A to B. After completing its first closed orbit in the counterclockwise direction, the object returns to is original position, now moving from A to B with a spin that is precessed in the clockwise direction. The initial spin and the final spin for the first closed orbit starting at A are shown

As the spinning object moves around its polygonal orbit, its spin axis, as observed in Σ, precesses by the Thomas precession angle ϵ_n when it rounds each of the n corners of the polygon as shown in Fig. 7. The total angle of precession is thus $n\epsilon_n$, represented by the unimodular complex number

$$e^{in\epsilon_n} = \left\{ 1 + f(\frac{2\pi}{n}) \right\}^n \tag{89}$$

In the limit $n \to \infty$ the polygonal path becomes a circular path, and the frame of reference in which the center of momentum of the spinning object is momentarily at rest is being changed continually. The total Thomas precession is thus the angle ϵ_t given by the equation

$$e^{i\epsilon_t} = \lim_{n\to\infty} e^{in\epsilon_n} = \lim_{n\to\infty} \left\{1 + f(\frac{2\pi}{n})\right\}^n \tag{90}$$

Let $g(x)$, $x \in \mathbb{R}$, be the function

$$g(x) = \lim_{n\to\infty} \left\{1 + f(\frac{x}{n})\right\}^n \tag{91}$$

The function $f(\phi)$ is continuous on $\mathbb{R}$, satisfying $f(0) = 0$. Hence,

$$\lim_{n\to\infty} f(\frac{x}{n}) = 0 \tag{92}$$

for any $x \in \mathbb{R}$.

Interchanging the limit in (91) with a differentiation with respect to x we find that the function $g(x)$ satisfies the initial value problem

$$\begin{aligned} g'(x) &= f'(0)g(x) \\ g(0) &= 1 \end{aligned} \tag{93}$$

for $x \in \mathbb{R}$.

The unique solution of the initial value problem (93) is

$$g(x) = e^{f'(0)x} \tag{94}$$

Hence, in particular for $x = 2\pi$, it follows from (90), (91) and (94) that

$$e^{i\epsilon_t} = g(2\pi) = e^{2\pi f'(0)} \tag{95}$$

But,

$$f'(0) = -i\frac{\gamma_{\mathbf{v}} - 1}{\gamma_{\mathbf{v}}} \tag{96}$$

Hence, by (95) and (96), the Thomas precession angle ϵ_t is given by the equation

$$\epsilon_t = -2\pi\frac{\gamma_{\mathbf{v}} - 1}{\gamma_{\mathbf{v}}} \tag{97}$$

The Thomas precession angle ϵ_t is the angle through which the spin axis precesses in one complete circular orbit. It requires, therefore, $2\pi/\epsilon_t$ orbits for the object to precess to its original orientation through 2π radians. Hence, if the angular velocity of the circular motion of the object is ω, then the angular velocity ω_t of the Thomas precession angle of the object is given by the equation

$$\omega_t = \frac{\epsilon_t}{2\pi}\omega = -\frac{\gamma_{\mathbf{v}} - 1}{\gamma_{\mathbf{v}}}\omega \tag{98}$$

The quantity ω_t in (98) is the angular velocity of the Thomas precession angle ϵ_t of a particle that moves in a circular orbit with angular velocity ω.

Equation (98) relates the angular velocity ω_t of the Thomas precession angle ϵ_t to its generating angular velocity ω. It demonstrates that the angular velocities ω_t and ω are oppositely directed, as shown graphically in Fig. 4.

If the magnitude of the velocity $\mathbf{v}$ and the acceleration $\mathbf{a}$ of the spinning object are v and a then its angular velocity is given by the equation $\omega = a/v$. Hence, the angular velocity ω_t of the Thomas precession angle ϵ_t is given by the equation

$$\omega_t = -\frac{\gamma_{\mathbf{v}} - 1}{\gamma_{\mathbf{v}}} \frac{a}{v} \tag{99}$$

Taking into account the direction of the Thomas precession axis and the velocity and the acceleration of the spinning object, and noting (10), the Thomas precession angular velocity ω_t in (99) can be written as a vector equation,

$$\boldsymbol{\omega}_t = \frac{\gamma_{\mathbf{v}} - 1}{\gamma_{\mathbf{v}}} \frac{\mathbf{a} \times \mathbf{v}}{v^2} = \frac{\gamma_{\mathbf{v}}}{1 + \gamma_{\mathbf{v}}} \frac{\mathbf{a} \times \mathbf{v}}{c^2} \tag{100}$$

The coordinate axes in the rest frame of any body in torque-free, accelerated motion precesses with respect to the laboratory axes with an angular velocity $\boldsymbol{\omega}_t$ is given by (100). Since $\gamma_{\mathbf{v}}/(1+\gamma_{\mathbf{v}}) = 1/2 + (1/8)(v^2/c^2) + \ldots$, the angular velocity $\boldsymbol{\omega}_t$ of the resulting Thomas precession, for the case when $v = \|\mathbf{v}\| \ll c$, is given approximately by the equation

$$\boldsymbol{\omega}_t = \frac{1}{2} \frac{\mathbf{a} \times \mathbf{v}}{c^2} \tag{101}$$

As noted by Herbert Goldstein [18, p. 288], $\boldsymbol{\omega}_t$ in (101) is known as the *Thomas precession frequency*.

Thomas precession frequency (101) involves the famous factor 1/2, known as *Thomas half*. The experimental significance of this factor is well known. The spinning electron of the Goudsmit-Uhlenbeck model gives twice the observed precession effect, which is reduced to the observed one by means of the Thomas half [43].

14 Thomas Precession and Boost Composition

Einstein addition underlies the Lorentz transformation group of special relativity. A Lorentz transformation is a linear transformation of spacetime coordinates that fixes the spacetime origin. A Lorentz boost, $B(\mathbf{v})$, is a Lorentz transformation without rotation, parametrized by a velocity parameter $\mathbf{v} \in \mathbb{R}^3_c$. The velocity parameter is given by its components, $\mathbf{v} = (v_1, v_2, v_3)$, with respect to a given Cartesian coordinate system of $\mathbb{R}^3_c$. Being linear, the Lorentz boost has a matrix representation, which turns out to be [27],

$$B(\mathbf{v}) =$$

$$\begin{pmatrix} \gamma_{\mathbf{v}} & c^{-2}\gamma_{\mathbf{v}}v_1 & c^{-2}\gamma_{\mathbf{v}}v_2 & c^{-2}\gamma_{\mathbf{v}}v_3 \\ \gamma_{\mathbf{v}}v_1 & 1+c^{-2}\frac{\gamma_{\mathbf{v}}^2}{\gamma_{\mathbf{v}}+1}v_1^2 & c^{-2}\frac{\gamma_{\mathbf{v}}^2}{\gamma_{\mathbf{v}}+1}v_1v_2 & c^{-2}\frac{\gamma_{\mathbf{v}}^2}{\gamma_{\mathbf{v}}+1}v_1v_3 \\ \gamma_{\mathbf{v}}v_2 & c^{-2}\frac{\gamma_{\mathbf{v}}^2}{\gamma_{\mathbf{v}}+1}v_1v_2 & 1+c^{-2}\frac{\gamma_{\mathbf{v}}^2}{\gamma_{\mathbf{v}}+1}v_2^2 & c^{-2}\frac{\gamma_{\mathbf{v}}^2}{\gamma_{\mathbf{v}}+1}v_2v_3 \\ \gamma_{\mathbf{v}}v_3 & c^{-2}\frac{\gamma_{\mathbf{v}}^2}{\gamma_{\mathbf{v}}+1}v_1v_3 & c^{-2}\frac{\gamma_{\mathbf{v}}^2}{\gamma_{\mathbf{v}}+1}v_2v_3 & 1+c^{-2}\frac{\gamma_{\mathbf{v}}^2}{\gamma_{\mathbf{v}}+1}v_3^2 \end{pmatrix} \tag{102}$$

Employing the matrix representation (102) of the Lorentz transformation boost, the Lorentz boost application to spacetime coordinates takes the form

$$B(\mathbf{v})\begin{pmatrix} t \\ \mathbf{x} \end{pmatrix} = B(\mathbf{v})\begin{pmatrix} t \\ x_1 \\ x_2 \\ x_3 \end{pmatrix} =: \begin{pmatrix} t' \\ x_1' \\ x_2' \\ x_3' \end{pmatrix} = \begin{pmatrix} t' \\ \mathbf{x}' \end{pmatrix} \tag{103}$$

where $\mathbf{v} = (v_1, v_2, v_3)^t \in \mathbb{R}_c^3$, $\mathbf{x} = (x_1, x_2, x_3)^t \in \mathbb{R}^3$, $\mathbf{x}' = (x_1', x_2', x_3')^t \in \mathbb{R}^3$, and $t, t' \in \mathbb{R}$, where exponent t denotes transposition.

A 2-dimensional boost is obtained in the special case when $v_3 = x_3 = 0$ in (102) and (103). For simplicity, the boosts of inertial frames in Fig. 4 are two dimensional boosts, and time coordinates are not shown. In this figure, the spacetime coordinate systems Σ, Σ' and Σ'' (only two space coordinates are shown) are related by boosts. Specifically, in Fig. 4,

1. The application of the boost $B(\mathbf{u})$ to the spacetime coordinate system Σ gives the spacetime coordinate system Σ',
2. The application of the boost $B(\mathbf{v})$ to the spacetime coordinate system Σ' gives the spacetime coordinate system Σ'', and
3. The application of the boost $B(\mathbf{u}\oplus\mathbf{v})$, or $B(\mathbf{v}\oplus\mathbf{u})$, to the spacetime coordinate system Σ, preceded, or followed respectively, by a Thomas precession (see (107) and (108) in Theorem 2 below) gives the spacetime coordinate system Σ''.

The Lorentz boost (102) and (103) can be written vectorially in the form

$$B(\mathbf{u})\begin{pmatrix} t \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \gamma_{\mathbf{u}}(t + \frac{1}{c^2}\mathbf{u}\cdot\mathbf{x}) \\ \gamma_{\mathbf{u}}\mathbf{u}t + \mathbf{x} + \frac{1}{c^2}\frac{\gamma_{\mathbf{u}}^2}{1+\gamma_{\mathbf{u}}}(\mathbf{u}\cdot\mathbf{x})\mathbf{u} \end{pmatrix} \tag{104}$$

Rewriting (104) with $\mathbf{x} = \mathbf{v}t \in \mathbb{R}^3$, $\mathbf{u}, \mathbf{v} \in \mathbb{R}_c^3 \subset \mathbb{R}^3$, we have

$$
\begin{aligned}
B(\mathbf{u})\begin{pmatrix} t \\ \mathbf{x} \end{pmatrix} &= B(\mathbf{u})\begin{pmatrix} t \\ \mathbf{v}t \end{pmatrix} \\
&= \begin{pmatrix} \gamma_{\mathbf{u}}(t + \frac{1}{c^2}\mathbf{u}\cdot\mathbf{v}t) \\ \gamma_{\mathbf{u}}\mathbf{u}t + \mathbf{v}t + \frac{1}{c^2}\frac{\gamma_{\mathbf{u}}^2}{1+\gamma_{\mathbf{u}}}(\mathbf{u}\cdot\mathbf{v}t)\mathbf{u} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\gamma_{\mathbf{u}\oplus\mathbf{v}}}{\gamma_{\mathbf{v}}} t \\ \frac{\gamma_{\mathbf{u}\oplus\mathbf{v}}}{\gamma_{\mathbf{v}}}(\mathbf{u}\oplus\mathbf{v})t \end{pmatrix} \\
&= \begin{pmatrix} t' \\ \mathbf{x}' \end{pmatrix}
\end{aligned}
\qquad (105)
$$

The equations in (105) reveal explicitly the way Einstein velocity addition underlies the Lorentz boost. The third equality in (105) follows from (9a) and (2).

In general, the composition of two boosts is equivalent to a single boost preceded, or followed, by the space rotation that Thomas precession generates, as we see from the following theorem:

Theorem 2 (The Boost Composition Theorem). *Let* $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^3_c$ *be relativistically admissible velocities, let* $\mathbf{x} = \mathbf{w}t$*,* $t > 0$*, and let* $B(\mathbf{u})$ *and* $B(\mathbf{v})$ *be two boosts. Furthermore, let* $\mathrm{Gyr}[\mathbf{u}, \mathbf{v}]$ *be the spacetime gyration of space coordinates, given by*

$$
\mathrm{Gyr}[\mathbf{u}, \mathbf{v}]\begin{pmatrix} t \\ \mathbf{x} = \mathbf{w}t \end{pmatrix} := \begin{pmatrix} t \\ (\mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{w})t \end{pmatrix} \qquad (106)
$$

Then, boost composition is given by each of the two equations

$$
B(\mathbf{u})B(\mathbf{v}) = B(\mathbf{u}\oplus\mathbf{v})\mathrm{Gyr}[\mathbf{u}, \mathbf{v}] \qquad (107)
$$

$$
B(\mathbf{u})B(\mathbf{v}) = \mathrm{Gyr}[\mathbf{v}, \mathbf{u}]B(\mathbf{v}\oplus\mathbf{u}) \qquad (108)
$$

Proof. We will show that (107) follows from the gyroassociative law of Einstein addition and that that (108) follows from (107) and the gyrocommutative law of Einstein addition.

Let us consider the chain of equations below, which are numbered for subsequent explanation:

$$\begin{aligned}
B(\mathbf{u})B(\mathbf{v})\begin{pmatrix} t \\ \mathbf{x} \end{pmatrix} &\overbrace{=\!=}^{(1)} B(\mathbf{u})B(\mathbf{v})\begin{pmatrix} t \\ \mathbf{w}t \end{pmatrix} \\
&\overbrace{=\!=}^{(2)} B(\mathbf{u})\begin{pmatrix} \frac{\gamma_{\mathbf{v}\oplus\mathbf{w}}}{\gamma_{\mathbf{w}}} t \\ \frac{\gamma_{\mathbf{v}\oplus\mathbf{w}}}{\gamma_{\mathbf{w}}} (\mathbf{v}\oplus\mathbf{w})t \end{pmatrix} \\
&\overbrace{=\!=}^{(3)} B(\mathbf{u})\begin{pmatrix} t' \\ (\mathbf{v}\oplus\mathbf{w})t' \end{pmatrix} \\
&\overbrace{=\!=}^{(4)} \begin{pmatrix} \frac{\gamma_{\mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w})}}{\gamma_{\mathbf{v}\oplus\mathbf{w}}} t' \\ \frac{\gamma_{\mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w})}}{\gamma_{\mathbf{v}\oplus\mathbf{w}}} \{\mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w})\}t' \end{pmatrix} \\
&\overbrace{=\!=}^{(5)} \begin{pmatrix} \frac{\gamma_{\mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w})}}{\gamma_{\mathbf{w}}} t \\ \frac{\gamma_{\mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w})}}{\gamma_{\mathbf{w}}} \{\mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w})\}t \end{pmatrix} \\
&\overbrace{=\!=}^{(6)} \begin{pmatrix} \frac{\gamma_{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}{\gamma_{\mathbf{w}}} t \\ \frac{\gamma_{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}{\gamma_{\mathbf{w}}} \{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}\}t \end{pmatrix}
\end{aligned} \tag{109}$$

□

Derivation of the numbered equalities in (109) follows:

1. Follows from the definition $\mathbf{x} = \mathbf{w}t$.
2. Follows from 1 by a boost application to spacetime coordinates according to (105).
3. Follows from 2 by the obvious definition

$$t' = \frac{\gamma_{\mathbf{v}\oplus\mathbf{w}}}{\gamma_{\mathbf{w}}} t \tag{110}$$

4. Follows from 3 by a boost application to spacetime coordinates according to (105).
5. Follows from 4 by the substitution of (110) for t'.
6. Follows from 5 by the gyroassociative law of Einstein addition.

Hence, following (109) we have the equation

$$B(\mathbf{u})B(\mathbf{v})\begin{pmatrix} t \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \frac{\gamma_{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}{\gamma_{\mathbf{w}}} t \\ \frac{\gamma_{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}{\gamma_{\mathbf{w}}} \{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}\}t \end{pmatrix} \tag{111}$$

Now, let us consider another chain of equations, which are numbered for subsequent explanation:

$$
\begin{aligned}
B(\mathbf{u}\oplus\mathbf{v})\mathrm{Gyr}[\mathbf{u},\mathbf{v}]\begin{pmatrix} t \\ \mathbf{x} \end{pmatrix} &\overbrace{=\!=\!=}^{(1)} B(\mathbf{u}\oplus\mathbf{v})\begin{pmatrix} t \\ \mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{x} \end{pmatrix} \\
&\overbrace{=\!=\!=}^{(2)} B(\mathbf{u}\oplus\mathbf{v})\begin{pmatrix} t \\ (\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w})t \end{pmatrix} \\
&\overbrace{=\!=\!=}^{(3)} \begin{pmatrix} \dfrac{\gamma_{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}{\gamma_{\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}t \\ \dfrac{\gamma_{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}{\gamma_{\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}\{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}\}t \end{pmatrix} \\
&\overbrace{=\!=\!=}^{(4)} \begin{pmatrix} \dfrac{\gamma_{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}{\gamma_{\mathbf{w}}}t \\ \dfrac{\gamma_{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}{\gamma_{\mathbf{w}}}\{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}\}t \end{pmatrix}
\end{aligned}
\tag{112}
$$

Derivation of the numbered equalities in (112) follows:

1. Follows from the definition of the spacetime gyration $\mathrm{Gyr}[\mathbf{u},\mathbf{v}]$ in terms of the space gyration $\mathrm{gyr}[\mathbf{u},\mathbf{v}]$ in (106).
2. Follows from 1 by definition, $\mathbf{x} = \mathbf{w}t$.
3. Follows from 2 by a boost application to spacetime coordinates according to (105).
4. Follows from 3 by the identity $\gamma_{\mathbf{w}} = \gamma_{\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}$ that, in turn, follows from the definition of gamma factors in (3) along with the invariance (37) of relativistically admissible velocities under gyrations.

Hence, following (112) we have the equation

$$
B(\mathbf{u}\oplus\mathbf{v})\mathrm{Gyr}[\mathbf{u},\mathbf{v}]\begin{pmatrix} t \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \dfrac{\gamma_{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}{\gamma_{\mathbf{w}}}t \\ \dfrac{\gamma_{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}{\gamma_{\mathbf{w}}}\{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}\}t \end{pmatrix}
\tag{113}
$$

Finally, (107) follows from (111) and (113).

In order to verify (108), let us now consider the chain of equations below, which are numbered for subsequent explanation:

$$
\begin{aligned}
\mathrm{Gyr}[\mathbf{v},\mathbf{u}]B(\mathbf{v}\oplus\mathbf{u})\begin{pmatrix} t \\ \mathbf{x} \end{pmatrix} &\overbrace{=\!=\!=}^{(1)} \mathrm{Gyr}[\mathbf{v},\mathbf{u}]B(\mathbf{v}\oplus\mathbf{u})\begin{pmatrix} t \\ \mathbf{w}t \end{pmatrix} \\
&\overbrace{=\!=\!=}^{(2)} \mathrm{Gyr}[\mathbf{v},\mathbf{u}]\begin{pmatrix} \dfrac{\gamma_{(\mathbf{v}\oplus\mathbf{u})\oplus\mathbf{w}}}{\gamma_{\mathbf{w}}}t \\ \dfrac{\gamma_{(\mathbf{v}\oplus\mathbf{u})\oplus\mathbf{w}}}{\gamma_{\mathbf{w}}}\{(\mathbf{v}\oplus\mathbf{u})\oplus\mathbf{w}\}t \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
&\overbrace{=\!=}^{(3)} \begin{pmatrix} \dfrac{\gamma_{(\mathbf{v}\oplus\mathbf{u})\oplus\mathbf{w}}}{\gamma_{\mathbf{w}}} t \\ \dfrac{\gamma_{(\mathbf{v}\oplus\mathbf{u})\oplus\mathbf{w}}}{\gamma_{\mathbf{w}}} \mathrm{gyr}[\mathbf{v},\mathbf{u}]\{(\mathbf{v}\oplus\mathbf{u})\oplus\mathbf{w}\}t \end{pmatrix} \\
&\overbrace{=\!=}^{(4)} \begin{pmatrix} \dfrac{\gamma_{\mathrm{gyr}[\mathbf{u},\mathbf{v}]\{(\mathbf{v}\oplus\mathbf{u})\oplus\mathbf{w}\}}}{\gamma_{\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}} t \\ \dfrac{\gamma_{\mathrm{gyr}[\mathbf{u},\mathbf{v}]\{(\mathbf{v}\oplus\mathbf{u})\oplus\mathbf{w}\}}}{\gamma_{\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}} \mathrm{gyr}[\mathbf{v},\mathbf{u}]\{(\mathbf{v}\oplus\mathbf{u})\oplus\mathbf{w}\}t \end{pmatrix} \\
&\overbrace{=\!=}^{(5)} \begin{pmatrix} \dfrac{\gamma_{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}{\gamma_{\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}} t \\ \dfrac{\gamma_{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}}{\gamma_{\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}} \{(\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}\}t \end{pmatrix} \\
&\overbrace{=\!=}^{(6)} B(\mathbf{u}\oplus\mathbf{v}) \begin{pmatrix} t \\ \mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}t \end{pmatrix} \\
&\overbrace{=\!=}^{(7)} B(\mathbf{u}\oplus\mathbf{v}) \begin{pmatrix} t \\ \mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{x} \end{pmatrix} \\
&\overbrace{=\!=}^{(8)} B(\mathbf{u}\oplus\mathbf{v})\mathrm{Gyr}[\mathbf{u},\mathbf{v}] \begin{pmatrix} t \\ \mathbf{x} \end{pmatrix}
\end{aligned}
\tag{114}
$$

The chain of equations (114) is valid for all spacetime events $(t, \mathbf{x})^t$, $t \in \mathbb{R}$, $\mathbf{x} = \mathbf{w}t$, $\mathbf{w} \in \mathbb{R}_c^3$, thus verifying (108)¡ and the proof of Theorem 2 is complete. Derivation of the numbered equalities in (114) follows:

1. Follows by definition, $\mathbf{x} = \mathbf{w}t$.
2. Follows from 1 by a boost application to spacetime coordinates according to (105).
3. Follows from 2 by the definition of the spacetime gyration Gyr$[\mathbf{v}, \mathbf{u}]$ in terms of the space gyration gyr$[\mathbf{v}, \mathbf{u}]$ in (106).
4. Follows from 3 by the identity $\gamma_{\mathbf{w}} = \gamma_{\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w}}$, $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}_c^3$, that, in turn, follows from the definition of gamma factors in (3) along with the invariance (37) of relativistically admissible velocities under gyrations.
5. Follows from 4 by the linearity of gyrations along with the gyrocommutative law of Einstein addition.
6. Follows from 5 by a boost application to spacetime coordinates according to (105).
7. Follows from 6 by definition, $\mathbf{x} = \mathbf{w}t$.
8. Follows from 7 by the definition of the spacetime gyration Gyr$[\mathbf{v}, \mathbf{u}]$ in terms of the space gyration gyr$[\mathbf{v}, \mathbf{u}]$ in (106).

□

The Boost Composition Theorem 2 and its proof establish the following two results:

1. The composite velocity of frame Σ'' relative to frame Σ in Fig. 4 may, paradoxically, be both $\mathbf{u}\oplus\mathbf{v}$ and $\mathbf{v}\oplus\mathbf{u}$. Indeed, we see from the result (107) and (108) of Theorem 2 that
   (a) The composite velocity of frame Σ'' relative to frame Σ in Fig. 4 is $\mathbf{u}\oplus\mathbf{v}$ in the sense that Σ'' is obtained from Σ by a boost of velocity $\mathbf{u}\oplus\mathbf{v}$ *preceded* by the gyration $\mathrm{Gyr}[\mathbf{u}, \mathbf{v}]$; and, equivalently,
   (b) The composite velocity of frame Σ'' relative to frame Σ in Fig. 4 is $\mathbf{v}\oplus\mathbf{u}$ in the sense that Σ'' is obtained from Σ by a boost of velocity $\mathbf{v}\oplus\mathbf{u}$ *followed* by the gyration $\mathrm{Gyr}[\mathbf{v}, \mathbf{u}]$.
2. The relationships (107) and (108) between boosts and Thomas precession are equivalent to the gyroassociative law and the gyrocommutative law of Einstein velocity addition as we see from the proof of Theorem 2.

In view of these two results of the Boost Composition Theorem, the validity of the Thomas precession frequency, as shown graphically in Fig. 4, and the relationship between the Thomas precession angle ϵ and its generating angle θ stem from the gyroassociative law of Einstein velocity addition. Hence, in particular, the result that ϵ and θ have opposite signs is embedded in the gyroassociative law of Einstein addition. In the next section we will present a convincing numerical demonstration that interested readers may follow to determine that, indeed, ϵ and θ in Fig. 4 have opposite signs.

15 Thomas Precession Angle and Generating Angle Have Opposite Signs

As in Fig. 4, let ϵ and θ be the Thomas Precession Angle and its generating angle, respectively. As verified analytically, and as shown graphically in Fig. 4, the angles ϵ and θ are related by (71) and, hence, they have opposite signs.

Without loss of generality, as in Fig. 4, we limit our considerations to two space dimensions. Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2_s$ be two nonzero relativistically admissible velocities with angle θ between their directions, as shown in Fig. 4. Then, they are related by the equation

$$\frac{\mathbf{v}}{\|\mathbf{v}\|} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \frac{\mathbf{u}}{\|\mathbf{u}\|} \tag{115}$$

Let $\mathbf{w} \in \mathbb{R}^2_s$ be the velocity of an object relative to frame Σ'' in Fig. 4. Then, the velocity of the object relative to frame Σ in Fig. 4 is

$$\mathbf{u}\oplus(\mathbf{v}\oplus\mathbf{w}) = (\mathbf{u}\oplus\mathbf{v})\oplus\mathrm{gyr}[\mathbf{u}, \mathbf{v}]\mathbf{w} \tag{116}$$

so that the velocity **w** of the object is rotated relative to Σ by the Thomas precession gyr[**u**, **v**], which corresponds to the rotation angle ϵ given by (71). Hence,

$$\mathrm{gyr}[\mathbf{u},\mathbf{v}]\mathbf{w} = \begin{pmatrix} \cos\epsilon & -\sin\epsilon \\ \sin\epsilon & \cos\epsilon \end{pmatrix} \mathbf{w} \tag{117}$$

where ϵ is given by (71).

Substituting **v** from (115) into (117), we obtain the equation

$$\mathrm{gyr}[\mathbf{u}, \frac{\|\mathbf{v}\|}{\|\mathbf{u}\|} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \mathbf{u}]\mathbf{w} = \begin{pmatrix} \cos\epsilon & -\sin\epsilon \\ \sin\epsilon & \cos\epsilon \end{pmatrix} \mathbf{w} \tag{118}$$

In (118), θ is the angle shown in the left part of Fig. 4, which generates the Thomas precession angle ϵ shown in the right part of Fig. 4, where ϵ is determined by θ according to (71) and, hence, where θ and ϵ have opposite signs. The validity of (118) can readily be corroborated numerically. The numerical corroboration of the validity of (118), in turn, provides a simple way to convincingly confirm our claim that indeed θ and ϵ have opposite signs.

References

1. L. Belloni, C. Reina, Sommerfeld's way to the Thomas precession. Eur. J. Phys. **7**, 55–61 (1986)
2. E. Borel, *Introduction Géométrique a Quelques Théories Physiques* (Gauthier, Paris, 1914)
3. J.-L. Chenm, A.A. Ungar, The Bloch gyrovector. Found. Phys. **32**(4), 531–565 (2002)
4. M. Chrysos, The non-intuitive $\frac{1}{2}$ thomas factor: a heuristic argument with classical electromagnetism. Eur. J. Phys. **27**(1), 1–4 (2006)
5. A.S. Eddington, *The Mathematical Theory of Relativity*. (University press, Cambridge, 1924)
6. J. Ehlers, W. Rindler, I. Robinson, Quaternions, bivectors, and the Lorentz group, in *Perspectives in Geometry (Essays in Honor of V. Hlavatý)* (Indiana University Press, Bloomington, 1966), pp. 134–149
7. A. Einstein, Zur Elektrodynamik Bewegter Körper [on the electrodynamics of moving bodies] (We use the English translation in, *Einstein's Miraculous Years: Five Papers that Changed the Face of Physics*, or in *The Principle of Relativity*, or in http://www.fourmilab.ch/etexts/einstein/specrel/www/. Ann. Phys. (Leipzig). **17**, 891–921 (1905)
8. A. Einstein, *Einstein's Miraculous Years: Five Papers that Changed the Face of Physics* (Princeton University Press, Princeton, 1998) Edited and introduced by John Stachel. Includes bibliographical references. Einstein's dissertation on the determination of molecular dimensions – Einstein on Brownian motion – Einstein on the theory of relativity – Einstein's early work on the quantum hypothesis. A new English translation of Einstein's 1905 paper on pp. 123–160
9. T. Feder, Strong near subgroups and left gyrogroups. J. Algebra **259**(1), 177–190 (2003)
10. M. Ferreira, G. Ren, Möbius gyrogroups: a Clifford algebra approach. J. Algebra **328**(1), 230–253 (2011)
11. V. Fock, *The Theory of Space, Time and Gravitation*, 2nd revised edn. (Macmillan, New York, 1964). Translated from the Russian by N. Kemmer. A Pergamon Press Book

12. T. Foguel, A.A. Ungar, Involutory decomposition of groups into twisted subgroups and subgroups. J. Group Theory **3**(1), 27–46 (2000)
13. T. Foguel, A.A. Ungar, Gyrogroups and the decomposition of groups into twisted subgroups and subgroups. Pac. J. Math **197**(1), 1–11 (2001)
14. H. Gelman, The second orthogonality conditions in the theory of proper and improper rotations. I. Derivation of the conditions and of their main consequences. J. Res. Nat. Bur. Stand. Sect. B **72B**, 229–237 (1968)
15. H. Gelman, The second orthogonality conditions in the theory of proper and improper rotations. II. The intrinsic vector. J. Res. Nat. Bur. Stand. Sect. B **73B**, 125–138 (1969)
16. H. Gelman, The second orthogonality conditions in the theory of proper and improper rotations. III. The conjugacy theorem. J. Res. Nat. Bur. Stand. Sect. B **73B**, 139–141 (1969)
17. H. Gelman, The second orthogonality conditions in the theory of proper and improper rotations. IV. Solution of the trace and secular equations. J. Res. Nat. Bur. Stand. Sect. B **73B**, 215–223 (1969)
18. H. Goldstein, *Classical Mechanics*. Addison-wesley series in physics, 2nd edn. (Wesley, Reading, 1980)
19. R.M. Jonson, Gyroscope precession in special and general relativity from basic principles. Am. J. Phys. **75**(5), 463–471 (2007)
20. D. Kalman, The axis of a rotation: analysis, algebra, geometry. Math. Mag. **62**(4), 248–252 (1989)
21. E. Kreyszig, *Differential Geometry* (Dover, New York, 1991). Reprint of the 1963 edition
22. H.A. Lorentz, A. Einstein, H. Minkowski, H. Weyl, *The Principle of Relativity* (Dover, New York, 1952). With notes by A. Sommerfeld, Translated by W. Perrett and G. B. Jeffery, A collection of original memoirs on the special and general theory of relativity
23. P.K. MacKeown, Question 57: Thomas precession. Am. J. Phys. **65**(2), 105 (1997)
24. G.B. Malykin, Thomas precession: correct and incorrect solutions. Phys.-Uspekhi **49**(8), 837–853 (2006)
25. J.E. Marsden, Steve Smale and geometric mechanics, in *From Topology to Computation: Proceedings of the Smalefest (Berkeley, CA, 1990)* (Springer, New York, 1993), pp. 499–516
26. J. McCleary, *Geometry From a Differentiable Viewpoint* (Cambridge University Press, Cambridge, 1994)
27. C. Møller, *The Theory of Relativity* (Clarendon Press, Oxford, 1952)
28. Th.M. Rassias, Book review: a gyrovector space approach to hyperbolic geometry, by Abraham A. Ungar. J. Geom. Symm. Phys. **18**, 93–106 (2010)
29. Th.M. Rassias, G.M. Rassias, *Selected Studies, Physics-Astrophysics, Mathematics, History of Science: A Volume Dedicated to the Memory of Albert Einstein* (North-Holland, Amsterdam, 1982)
30. J.A. Rhodes, M.D. Semon, Relativistic velocity space, Wigner rotation, and thomas precession. Am. J. Phys. **72**(7), 943–960 (2004)
31. W. Rindler, I. Robinson, A plain man's guide to bivectors, biquaternions, and the algebra and geometry of Lorentz transformations, in *On Einstein's Path (New York, 1996)* (Springer, New York, 1999), pp. 407–433.
32. K. Rózga, On central extensions of gyrocommutative gyrogroups. Pac. J. Math. **193**(1), 201–218 (2000)
33. R.U. Sexl, H.K. Urbantke, *Relativity, Groups, Particles*. Springer Physics (Springer, Vienna, 2001). Special relativity and relativistic symmetry in field and particle physics, Revised and translated from the third German (1992) edition by Urbantke
34. L. Silberstein, *The Theory of Relativity* (MacMillan, London, 1914)
35. S. Smale, Differentiable dynamical systems. Bull. Am. Math. Soc. **73**, 747–817 (1967)
36. S. Smale, in *The Collected Papers of Stephen Smale*, vols. 1–3, ed. by F. Cucker, R. Wong (Singapore University Press, Singapore, 2000)
37. J.D.H. Smith, A.A. Ungar, Abstract space-times and their Lorentz groups. J. Math. Phys. **37**(6), 3073–3098 (1996)

38. A. Sommerfeld, Über die Zusammensetzung der Geschwindigkeiten in der Relativtheorie. Physikalische Zeitschrift **10**, 826–829 (1909)
39. J. Stachel, History of relativity, in *Twentieth Century Physics*, vol. I, ed. by L.M. Brown, A. Pais, B. Pippard (Published jointly by the Institute of Physics Publishing, Bristol, 1995), pp. 249–356
40. E.F. Taylor, J.A. Wheeler, *Spacetime Physics* (W.H. Freeman, San Francisco, 1966)
41. L.H. Thomas, The motion of the spinning electron. Nature **117**, 514 (1926)
42. L.H. Thomas, The kinematics of an electron with an axis. Phil. Mag. **3**, 1–23 (1927)
43. L.H. Thomas, Recollections of the discovery of the Thomas precessional frequency, in *AIP Conference Proceedings No. 95, High Energy Spin Physics*, ed. by G.M Bunce (Brookhaven National Lab, Brookhaven, 1982)
44. M. Tsamparlis, *Special Relativity: An Introduction with 200 Problems and Solutions* (Springer, New York, 2010)
45. A.A. Ungar, Thomas rotation and the parametrization of the Lorentz transformation group. Found. Phys. Lett. **1**(1), 57–89 (1988)
46. A.A. Ungar, The relativistic noncommutative nonassociative group of velocities and the Thomas rotation. Result. Math. **16**(1–2), 168–179 (1989). The term “K-loop” is coined here
47. A.A. Ungar, Group-like structure underlying the unit ball in real inner product spaces. Result. Math. **18**(3–4), 355–364 (1990)
48. A.A. Ungar, Quasidirect product groups and the Lorentz transformation group, in *Constantin Carathéodory: An International Tribute*, vols. I, II, ed. by Th.M. Rassias (World Scientific, Teaneck, 1991), pp. 1378–1392
49. A.A. Ungar, Thomas precession and its associated grouplike structure. Am. J. Phys. **59**(9), 824–834 (1991)
50. A.A. Ungar, Thomas precession: its underlying gyrogroup axioms and their use in hyperbolic geometry and relativistic physics. Found. Phys. **27**(6), 881–951 (1997)
51. A.A. Ungar, Gyrovector spaces in the service of hyperbolic geometry, in *Mathematical Analysis and Applications*, ed. by Th.M. Rassias (Hadronic Press, Palm Harbor, 2000) pp. 305–360
52. A.A. Ungar, *Beyond the Einstein Addition Law and Its Gyroscopic Thomas Precession: The Theory of Gyrogroups and Gyrovector Spaces*, volume 117 of Fundamental theories of physics (Kluwer, Dordrecht, 2001)
53. AA. Ungar, Seeing the möbius disc-transformation like never before. Comput. Math. Appl. **45**, 805–822 (2003)
54. A.A. Ungar, *Analytic Hyperbolic Geometry: Mathematical Foundations and Applications* (World Scientific, Hackensack, 2005)
55. A.A. Ungar, Gyrovector spaces and their differential geometry. Nonlinear Funct. Anal. Appl. **10**(5), 791–834 (2005)
56. A.A. Ungar, *Analytic Hyperbolic Geometry and Albert Einstein’s Special Theory of Relativity* (World Scientific, Hackensack, 2008)
57. A.A. Ungar, Einstein’s special relativity: the hyperbolic geometric viewpoint, in *PIRT Conference Proceedings,* 4–6 September 2009, Budapest, pp. 1–35
58. A.A. Ungar *A Gyrovector Space Approach to Hyperbolic Geometry* (Morgan and Claypool, San Rafael, 2009).
59. A.A. Ungar, *Barycentric Calculus in Euclidean and Hyperbolic Geometry: A Comparative Introduction* (World Scientific, Hackensack, 2010)
60. A.A. Ungar, *Hyperbolic Triangle Centers: The Special Relativistic Approach* (Springer, New York, 2010)
61. A.A. Ungar, When relativistic mass meets hyperbolic geometry. Commun. Math. Anal. **10**(1), 30–56 (2011)
62. V. Varičak, Beiträge zur nichteuklidischen geometrie [contributions to non-euclidean geometry]. Jber. Dtsch. Mat. Ver. **17**, 70–83 (1908)
63. V. Varičak, Anwendung der Lobatschefskjschen Geometrie in der Relativtheorie. Physikalische Zeitschrift **11**, 93–96 (1910)

64. J. Vermeer, A geometric interpretation of Ungar's addition and of gyration in the hyperbolic plane. Topology Appl. **152**(3), 226–242 (2005)
65. S. Walter, The non-Euclidean style of Minkowskian relativity, in *The Symbolic Universe: Geometry and Physics 1890–1930*, ed. by J.J. Gray (Oxford University Press, New York, 1999) pp. 91–127
66. S. Walter, Book review: beyond the Einstein addition law and its gyroscopic Thomas precession: the theory of gyrogroups and gyrovector spaces, by Abraham A. Ungar. Found. Phys. **32**(2), 327–330 (2002)