Ken'ichi Ohshika
Athanase Papadopoulos *Eds.*

In the Tradition of Thurston II

Geometry and Groups

Springer

In the Tradition of Thurston II

Ken'ichi Ohshika • Athanase Papadopoulos
Editors

In the Tradition of Thurston II

Geometry and Groups

Springer

Editors
Ken'ichi Ohshika
Department of Mathematics
Gakushuin University
Tokyo, Japan

Athanase Papadopoulos
Institut de Recherche Mathématique
Avancée
Université de Strasbourg et CNRS
Strasbourg, France

ISBN 978-3-030-97559-3 ISBN 978-3-030-97560-9 (eBook)
https://doi.org/10.1007/978-3-030-97560-9

Cover Text: William Thurston at the Clay Research Conference on the Poincaré conjecture, Institut Océanographique, Paris, June 8, 2010. Photo © François Tisseyre, Atelier EcoutezVoir.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

William Thurston's ideas have had a profound and permanent influence on mathematics since the last third of the twentieth century, and this influence is still growing today, in geometry, topology, dynamics, and other fields, by the works of a whole community of mathematicians who are inspired by these ideas.

This second volume of the collection *In the Tradition of Thurston*, with the subtitle *Geometry and Dynamics*, covers several topics which originate in or are strongly motivated by Thurston's work. When we asked the various authors to contribute to this volume, we told them (like we did for the previous volume) that we wish the articles to be inspired by Thurston's work, and to be preferably a survey. On the other hand, we gave them complete freedom as to the choice of their topics, because we knew that, given Thurston's broad range of interests, there was little chance to have any significant overlap between the various contributions.

The topics covered in the present volume include complex hyperbolic Kleinian groups, Möbius structures, hyperbolic ends, cone 3-manifolds, Thurston's norm, surgeries in representation varieties, triangulations, spaces of polygons and of singular flat structures on surfaces, combination theorems in the theories of Kleinian groups, hyperbolic groups and holomorphic dynamics, iteration of rational maps, automatic groups, and the combinatorics of right-angled Artin groups.

We thank all the authors for their valuable collaboration, and Elena Griniari for her editorial support. We hope that the various chapters in this volume will be useful for learning the kind of mathematics that Thurston transmitted to us.

Tokyo, Japan — Ken'ichi Ohshika
Strasbourg, France — Athanase Papadopoulos
July 2022

Contents

Chapter 1
Introduction

Ken'ichi Ohshika and Athanase Papadopoulos

Keywords Thurston · Geometry · Topology · Group actions

MSC 75-06

William Thurston passed away in 2012, ten years ago. Since the beginning of his career as a mathematician, in the early 1970s, Thurston continually introduced new ideas, reviving old theories and making them very fresh and inspiring. He reshaped many fields, including group theory, geometry, topology, complex analysis, computer science, combinatorics and others, always trying to go deep in his understanding of patterns and forms. He was an artist in every sense of the word, creating beauty and sharing it with others. He loved mathematics, and transmitted his way of thinking to several generations of mathematicians who work in a tradition he established, until now and for many years to come.

Our objective in this volume, like in the other volumes of this series to which we have given the generic name "In the tradition of Thurston", is to make available to the wide community of mathematicians some highlights of the beautiful ideas which Thurston brought at the forefront of mathematical research. The main topics considered here include complex hyperbolic Kleinian groups, Möbius structures, hyperbolic ends and κ-surfaces in hyperbolic space, the Thurston norm, surgeries in representation varieties, triangulations and other cell decompositions of surfaces, combination theorems, iterations of rational maps, automatic groups, and the combinatorics and the geometry of right-angled Artin groups.

K. Ohshika
Department of Mathematics, Gakushuin University, Tokyo, Japan
e-mail: ohshika@math.gakushuin.ac.jp

A. Papadopoulos (✉)
Institut de Recherche Mathématique Avancée, Université de Strasbourg et CNRS, Strasbourg, France
e-mail: papadop@math.unistra.fr

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_1

To give a more precise idea of the material presented in the volume, let us make an overview of the content of each chapter.

Chapter 2, by Michael Kapovich, is titled *A survey of complex hyperbolic Kleinian groups*. It consists in a survey of the geometry and dynamics of discrete groups of isometries of complex hyperbolic spaces, with an emphasis on the interactions between this theory and the function theory of complex hyperbolic manifolds. Besides a fairly detailed exposition of these theories, the chapter contains an introduction to several basic topics such as discrete subgroups of the group $PU(n, 1)$ of holomorphic isometries of complex projective space, geometrically finite groups, the theory of ends, cone 3-manifolds, and critical exponents. The chapter also contains a list of open problems on this theme.

In Chap. 3, titled *Möbius structures, hyperbolic ends and κ-surfaces in hyperbolic space*, Graham Smith studies a generalisation of the correspondence between the boundary of the convex core of a quasi-Fuchsian manifold and the surface at infinity, which was introduced in Thurston's Princeton 1975–76 lecture notes. The author proves general theorems giving the correspondence between Möbius structures on a surface and hyperbolic ends, the latter being hyperbolic manifolds carrying height functions, that is, locally strictly convex functions whose gradient flow lines are unit speed geodesics and whose sub-level sets are complete. He shows the existence and the uniqueness of the k-surface corresponding to a given Möbius structure under certain reasonable assumptions, and constructs operators which allow the passage back and forth between families of hyperbolic ends and Möbius surfaces.

Chapter 4, by Joan Porti, titled *Cone 3-manifolds*, is concerned with the deformation theory of cone 3-manifolds, more precisely, of 3-manifolds equipped with metrics of constant curvature with singularities along embedded graphs. The author describes the phenomena that occur by deforming the cone angles, including the operations of cusp opening and collapsing, under special restrictions on the cone angles. He reviews the role of cone 3-manifolds with small cone angles appearing in the proof of Thurston's orbifold geometrisation theorem. A central ingredient in this proof involves a variation of the cone angles and an examination of the phenomena that can occur during this process. The chapter also contains a survey of the theory of geometric convergence of sequences of pointed cone manifolds and a review of the related compactness theorems. At the same time, the author mentions the appearance of cone manifolds in works of several authors other than Thurston, including his own work with Boileau and Leeb, the works of Hodgson and Kerckhoff in their study of Dehn fillings, of Bromberg in his proof of Bers's density conjecture, of Mednykh and Rasskazov on geometric structures on knot and link complements, and of other authors working in Thurston's tradition.

In Chap. 5, titled *A survey of the Thurston norm*, Takahiro Kitayama surveys recent progress in 3-dimensional topology which involves the Thurston norm. This is a semi-norm defined on the first cohomology group with real coefficient of a 3-manifold, whose introduction has had a huge impact on 3-dimensional topology. In this chapter, the author gives an overview of applications of the Thurston norm, putting a special emphasis on its relation with topological invariants of 3-manifolds, such as the (twisted) Alexander polynomial, the Seiberg–Witten invariant, and Heegaard–Flower homology, among others.

In Chap. 6, titled *From hyperbolic Dehn filling to surgeries in representation varieties*, Georgios Kydonakis surveys gluing constructions in representation spaces of surface groups into higher-rank Lie groups, as analogues of the operation of hyperbolic Dehn filling invented by Thurston. The gluing construction in the representation spaces consists in gluing Higgs bundles via the non-abelian Hodge correspondence. Thurston, in his hyperbolic Dehn filling theorem, gave a family of hyperbolic 3-manifolds which were unknown before. In the same way, the gluing construction in the setting of this chapter gives subsets of "model representations" which, presumably, could not be obtained by other methods. The author gives specific examples in the case where the target Lie groups are $\mathrm{SO}(p, p+1)$ and $\mathrm{Sp}(4, \mathbb{R})$.

In Chap. 7, titled *Acute geodesic triangulations of manifolds*, Sang-Hyun Kim gives a survey on the problem of existence of acute geodesic triangulations of Riemannian manifolds. In the first part of this chapter, the author surveys known results on acute triangulations of manifolds of dimension greater than 2: Kalai's theorem on the non-existence of acute geodesic triangulations of manifolds with dimension greater than 5, and some partial results for dimensions 3 and 4. The main part of the chapter deals with the case of surfaces. Starting from Colin de Verdière's result on the existence of geodesic triangulations on Riemannian surfaces and Thurston's famous theorem on the moduli space of flat metrics with cone singularities on the two-sphere, the author presents important topics such as the Koebe–Andreev–Thurston theorem on circle packings and a related result by Hodgson–Rivin on hyperbolic 3-polytopes.

In Chap. 8, titled *Signature calculation of the area Hermitian form on some spaces of polygons*, Ismail Sağlam studies Hermitian forms defined on the moduli spaces of singular flat metrics on the sphere having unit area and with prescribed curvature data at the singular points, arising from the area form. The work is motivated by Thurston's study of the moduli space of singular flat metrics on the sphere with n cone-singular points having angles less than 2π, where he showed that the area equation induces a Hermitian form of signature $(1, n-3)$ on the moduli space of such structures. The author in this chapter generalises Thurston's construction to the case where one singular point has angle greater than 2π, and he calculates the signature of the Hermitian form induced by the area form.

In Chap. 9, titled *Equilateral convex triangulations of* $\mathbb{R}P^2$ *with three conical points of equal defect*, Mikhail Chernaviskikh, Altan Erdnigor, Nikita Kalinin, and Alexandr Zakharov study convex equilateral triangulations of the real projective plane. The authors determine the top term of the growth function of the number of isometry classes of equilateral triangulations of the real projective plane with at most n vertices whose vertices have valency 6 except for three with valency 4. The coefficient of the top term is expressed by the Lobachevsky function which appears as the volume of a certain region in $\mathbb{R}^4$.

Chapter 10 by Mahan Mj and Sabyasachi Mukherjee, is titled *Combination theorems in groups, geometry and dynamics*. The authors survey a series of results, which they call “combination theorems”, and which are analogues of the Klein–Maskit combination theorems for Kleinian groups. The fields visited include hyperbolic geometry, geometric group theory and holomorphic dynamics. Thurston’s influence on all these fields is highlighted. The survey starts with the tools which Thurston developed in his proof of the uniformisation theorem for Haken manifolds and then pass through combination theorems for Gromov-hyperbolic groups due to Bestvina–Feign, Agol–Wise, Haglund–Wise and others. They review the Mahan–Reeves combination theorems in the relative hyperbolicity setting, Thurston’s topological characterisation of rational maps which led to the Douady–Hubbard theory of mating in holomorphic dynamics, and the Mahan–Mukherjee theory of combination of rational maps and Kleinian groups via orbit-equivalence.

Chapter 11 by Kevin Pilgrim is titled *On the pullback relation on curves induced by a Thurston map*. We recall here that a Thurston map is an orientation-preserving branched covering of the sphere by itself which is of degree ≥ 2 and which is postcritically finite, that is, the union of the forward orbit of its set of critical points is finite. A key result in the theory of holomorphic dynamics is a theorem of Thurston which provides a topological characterisation of (conjugacy classes of) post-critically finite rational maps of the sphere among Thurston maps. In this chapter, the author presents several results, including works of Floyd, Parry, himself and others on the action of a Thurston map by inverse images on the set of isotopy classes of simple closed curves in the complement of its post-critical set, and he proposes several open questions in this setting.

Chapter 12, by William Floyd, is titled *The pullback map on Teichmüller space induced from a Thurston map*. Thurston’s topological characterisation of rational maps is based on a pullback map σ_f acting on a certain Teichmüller space associated with a postcritically finite branched covering f of the Riemann sphere. Thurston proved that such a branched covering is combinatorially equivalent (in some precise sense) to a rational map if and only if σ_f has a fixed point. Furthermore, there is a bijection between the fixed points of σ_f and the conjugacy classes of rational maps equivalent to f. In this chapter, the author describes Thurston’s characterisation theorem and discusses the recent development in understanding the pullback map σ_f, made by Pilgrim, Selinger, Bartholdi, Nekrashevych, Buff, Epstein, Koch and others.

Chapter 13, by Russell Lodge, Yauhen Mikulich and Dierk Schleicher, is titled *A classification of postcritically finite Newton maps*. Here, a *Newton map* is a rational function $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ for which there exists some complex polynomial $p(z)$ such that $f(z) = z - \frac{p(z)}{p'(z)}$ for all $z \in \mathbb{C}$. Newton maps arise in the application of Newton’s method to a polynomial, and they form a natural family to be studied from the dynamical perspective. In this chapter, the authors work out, for the family of postcritically finite Newton maps, a complete combinatorial classification in terms of a finite connected graph satisfying certain explicit conditions. The classification relies on Thurston’s characterisation and rigidity theorem for postcritically finite branched covers of the sphere. This is a remarkable instance where a whole family of rational functions is subject to a classification.

Chapter 14, by Sarah Rees, is titled *The development of the theory of automatic groups*. After an overview of the basic notions of automatic, bi-automatic and combable groups, the author explains why hyperbolic groups and fundamental groups of compact 3-manifolds based on six of Thurston's eight geometries are automatic. She then describes algorithmic aspects of these groups and explains how actions of groups on spaces satisfying various notions of negative curvature can be used to prove automaticity or bi-automaticity. She shows how these results have been used to derive such properties in the setting of mapping class groups, Coxeter groups, braid groups and other families of Artin groups. All along the text, the author surveys the important problems which were tackled by the theory of automatic groups, mentioning several among those which remain open. We recall that the notion of automatic group was introduced by Thurston, after Cannon published his paper 1984 paper *The combinatorial structure of cocompact discrete hyperbolic groups*, in which the latter showed that Cayley graphs of cocompact discrete groups of isometries of n-dimensional hyperbolic space have a finite recursive description. Cannon, in the introduction to this paper, expresses his debt to Thurston who promoted a return to geometric considerations for the study of groups, that is, to the classical methods of Dehn and Cayley.

Chapter 15, by Thomas Koberda, is titled *Geometry and combinatorics via right-angled Artin groups*. We recall that for a given finite simplicial graph Γ, the associated right-angled Artin group is the group whose generators are the vertices of Γ and whose relations are the commutators corresponding to the edges Γ. As the author explains, these groups interpolate between free groups and abelian groups. Starting with right-angled Artin groups, he addresses several questions pertaining to the fields of combinatorial group theory, graph theory, complexity theory, mathematical logic, mapping class groups, hyperbolicity and others. Thurston's seal is visible on several of these questions. The author points out relations between objects from these various fields and formulates some open problems. In the last section, he provides a list of questions concerning what remains to be understood on the relationship between combinatorics and algebra, from the perspective of right-angled Artin groups.

The broad range of the contributions in this volume gives only a small idea of the scope of Thurston's impact on mathematics. It is certain that his insight, which is unique in the modern history of mathematics, will continue to influence succeeding generations of mathematicians.

Chapter 2
A Survey of Complex Hyperbolic Kleinian Groups

Michael Kapovich

Abstract This survey of discrete subgroups of isometries of complex hyperbolic spaces is aimed to discuss interactions between function theory on complex hyperbolic manifolds and the theory of discrete groups. We present a number of examples and basic results about complex-hyperbolic Kleinian groups. The appendix to the paper written by Mohan Ramachandran includes a proof of a result known as "Burns' Theorem" about ends of complex-hyperbolic manifolds.

Keywords Discrete groups · Complex-hyperbolic geometry

MSC (2020) 22E40, 32Q05

2.1 Introduction

This survey is based on a series of lectures I gave at the workshop "Progress in Several Complex Variables," held in KIAS, Seoul, Korea, in October of 2019. It is useful to read it in conjunction with my (longer) survey of discrete isometry groups of real hyperbolic spaces, [51], since most issues in the real and complex hyperbolic settings are quite similar. The theory of complex hyperbolic manifolds and complex hyperbolic Kleinian groups (aka discrete holomorphic isometry groups of complex hyperbolic spaces $\mathbb{H}^n_{\mathbb{C}}$) is a rich mixture of Riemannian and complex geometry, topology, dynamics, symplectic geometry and complex analysis. The choice of topics covered in the survey is governed by my personal taste and is, by no means, canonical: It is geared towards a discussion of interactions between the function theory on complex hyperbolic manifolds and the geometry/dynamics of complex hyperbolic Kleinian groups (Sects. 2.9 and 2.10). I refer the reader to [15, 33–35, 66, 70–72, 79] for further discussion of the geometry of complex

M. Kapovich (✉)
Department of Mathematics, UC Davis, Davis, CA, USA
e-mail: kapovich@math.ucdavis.edu

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_2

hyperbolic spaces and their discrete isometry groups. The bibliography of complex hyperbolic Kleinian groups appearing at the end of the survey is long but is not meant to be exhaustive, my apologies to everybody whose papers are omitted.

It should be pointed out that the development of the theory of complex hyperbolic manifolds and complex hyperbolic Kleinian groups was influenced in part by the work of William Thurston on real hyperbolic manifolds and Kleinian groups (especially Thurston's approach to geometric finiteness and orbifolds), as well as by Thurston's work on general geometric structures on manifolds. On a personal note, Thurston's way of thinking about hyperbolic manifolds, discrete groups and geometric structures, was the single most importance influence on my mathematical work. Much of my past and present research is driven by attempting to understand and generalize Thurston's results.

2.2 Complex Hyperbolic Space

Most of the basic material on geometry of complex hyperbolic spaces can be found in Goldman's book [35]; I also refer the reader to [33, 70, 72] for shorter introductions.

Consider the vector space $V = \mathbb{C}^{n+1}$ equipped with the pseudo-Hermitian bilinear form

$$\langle z, w\rangle = -z_0\bar{w}_0 + \sum_{k=1}^{n} z_k\bar{w}_k.$$

Set $q(z) := \langle z, z\rangle$. This quadratic form has signature $(n, 1)$. Define the *negative light cone* $V_- := \{z : q(z) < 0\}$. Consider the complex projective space $\mathbb{P}^n := PV$, the projectivization of V, and the projection $p : z \mapsto [z] \in \mathbb{P}^n$. The projection $\mathbf{B}^n := p(V_-)$ is an open ball in $\mathbb{P}^n$. In order to see this, consider the affine hyperplane in $\mathbb{C}^{n+1}$ given by $A = \{z_0 = 1\}$ (and equipped with the standard Euclidean Hermitian metric). Then $V_- \cap A$ is the open unit ball in A centered at the origin. This intersection projects diffeomorphically to $p(V_-)$.

The tangent space $T_{[z]}\mathbb{P}^n$ is naturally identified with $z^\perp$, the orthogonal complement of $\mathbb{C}z$ in V, taken with respect to $\langle\cdot,\cdot\rangle$. If $z \in V_-$, then the restriction of q to $z^\perp$ is positive-definite, hence, $\langle\cdot,\cdot\rangle$ project to a Hermitian metric h (also denoted $\langle\cdot,\cdot\rangle_h$) on $\mathbf{B}^n$. From now on, I will always equip $\mathbf{B}^n$ with the Hermitian metric h and let d denote the corresponding distance function on $\mathbf{B}^n$.

Definition 2.1 The *complex hyperbolic n-space* $\mathbb{H}^n_{\mathbb{C}}$ is $(\mathbf{B}^n, h)$.

I next describe the Hermitian metric h on $\mathbf{B}^n$ using the coordinates $(z_1, ..., z_n)$ on A. First, regarding $\mathbf{B}^n$ as a subset of the affine hyperplane A, for a vector $y \in T_x\mathbf{B}^n$ we have

$$\langle y, y\rangle_h = \frac{\langle x, x\rangle\langle y, y\rangle - \langle x, y\rangle\langle y, x\rangle}{-\langle x, x\rangle^2}.$$

Setting $x = (1, z)$, $z \in \mathbb{C}^n$, and denoting $u \cdot v$ the standard Euclidean Hermitian inner product on $\mathbb{C}^n$, we obtain:

$$\langle y, y\rangle_h = \frac{(-1+|z|^2)|y|^2 - (z\cdot y)(y\cdot z)}{-(-1+|z|^2)^2}, \qquad y \in T_z\mathbf{B}^n.$$

In the differential form, the metric h is, thus, given by

$$ds_h^2 = \frac{1}{1-|z|^2}\sum_{k=1}^{n} dz_k d\bar{z}_k + \frac{1}{(1-|z|^2)^2}\sum_{j,k=1}^{n} z_j\bar{z}_k dz_k d\bar{z}_j.$$

This Hermitian metric is Kähler, with the Kähler potential (centered at the origin) equal to

$$f(z) = \log(1-|z|^2),$$

and the Kähler form $\omega = \frac{i}{2}\partial\bar{\partial} f$ equal to

$$\omega = \frac{1}{1-|z|^2}\sum_{k=1}^{n} dz_k \wedge d\bar{z}_k + \frac{1}{(1-|z|^2)^2}\sum_{j,k=1}^{n} z_j\bar{z}_k dz_k \wedge d\bar{z}_j.$$

The complex hyperbolic metric on $\mathbf{B}^n$ (the unit ball in $\mathbb{C}^n$) is the Bergman metric with the Bergman kernel $K(z, \zeta)$ equal to

$$K(z,\zeta) = \frac{n!}{2\pi^n}(1-(z\cdot\zeta))^{-n-1},$$

where, as before, $z \cdot \zeta$ is the standard Hermitian inner product on $\mathbb{C}^n$.

The distance function d on $\mathbb{H}^n_{\mathbb{C}}$ satisfies

$$\cosh^2(d([x],[y])) = \frac{\langle x, y\rangle\langle x, y\rangle}{\langle x, x\rangle\langle y, y\rangle}.$$

For example, specializing to the case when $[x]$ is the center of $\mathbf{B}^n$ and $[y]$ is represented by a point $z \in \mathbf{B}^n$, we obtain:

$$\cosh^2(d(0,z)) = (1-|z|^2)^{-1}.$$

See [35, pp. 72–79] and [59, §1.4]; note however that Goldman uses a different normalization of the metric on the complex hyperbolic space; with his normalization sectional curvature varies in the interval $[-2, -\frac{1}{2}]$.

A real linear subspace $W \subset V$ is said to be *totally real with respect to the form* $\langle\cdot,\cdot\rangle$ if for any two vectors $z, w \in W$, $\langle z, w\rangle \in \mathbb{R}$. Such a subspace is automatically totally real in the usual sense: $JW \cap W = \{0\}$, where J is the almost complex structure on V.

Real geodesics in $\mathbf{B}^n$ are projections (under p) of totally real indefinite (with respect to q) 2-planes in V (intersected with V_-). For instance, geodesics through the origin $0 \in \mathbf{B}^n$ are Euclidean line segments in $\mathbf{B}^n$.

More generally, totally-geodesic real subspaces in $\mathbf{B}^n$ are projections of totally real indefinite subspaces in V (intersected with V_-). They are isometric to the real hyperbolic space $\mathbb{H}^n_{\mathbb{R}}$ of constant sectional curvature -1. Boundaries of real hyperbolic planes are called *real circles* in S^{2n-1}.

Complex geodesics in $\mathbf{B}^n$ are projections of indefinite complex 2-planes; boundaries of complex geodesics are called *complex circles* in S^{2n-1}. Complex geodesics are isometric to the unit disk with the Hermitian metric

$$\frac{dzd\bar{z}}{(1-|z|^2)^2},$$

which has constant curvature -4. These are the extremal disks for the Kobayashi metric on $\mathbf{B}^n$, which coincides with the complex hyperbolic distance function d. It is also equal to the Caratheodory's metric on $\mathbf{B}^n$ (as is the case for all bounded convex domains in $\mathbb{C}^n$).

More generally, complex hyperbolic k-dimensional subspaces $\mathbb{H}^k_{\mathbb{C}}$ in $\mathbf{B}^n$ are projections of indefinite complex $k+1$-dimensional subspaces (intersected with V_-).

All complete totally-geodesic submanifolds in $\mathbb{H}^n_{\mathbb{C}}$ are either real or complex hyperbolic subspaces.

The holomorphic bisectional curvature of $\mathbb{H}^n_{\mathbb{C}}$ is constant, equal to -1. It turns out that $\mathbb{H}^n_{\mathbb{C}}$ has negative sectional curvature which varies in the interval $[-4, -1]$. Thus, $\mathbb{H}^n_{\mathbb{C}}$ is a *negatively pinched Hadamard manifold*:

Definition 2.2

1. A Hadamard manifold X is a simply-connected complete nonpositively curved Riemannian manifold.
2. A Hadamard manifold X is said to have *strictly negative curvature* if there exists $a < 0$ such that the sectional curvature of X is $\leq a$.
3. A Hadamard manifold X is said to be *negatively pinched* (has *pinched negative curvature*) if there exist two negative numbers $b \leq a < 0$ such that the sectional curvature of X lies in the interval $[b, a]$.

The group $U(n, 1) = U(q)$ of (complex) automorphisms of q projects to the group $G = PU(n, 1) = Aut(\mathbf{B}^n)$ of complex (biholomorphic) automorphisms of $\mathbf{B}^n$. This group acts transitively, with the stabilizer of the center of $\mathbf{B}^n$ equal to $K = U(n)$. Consequently, the metric d on $\mathbf{B}^n$ is complete. The group G is a Lie group, its Lie topology is equivalent to the topology of pointwise convergence, equivalently, the topology of uniform convergence on compacts in $\mathbf{B}^n$, equivalently, the quotient topology of the matrix group topology on $U(n, 1)$. The group G is linear, its matrix representation is given, for instance, by the adjoint representation, which is faithful since G has trivial center.

The Lie group G is connected and has real rank 1. Its Cartan decomposition is

$$G = KA_+K,$$

where A_+ is the semigroup of positive translations (transvections) along a chosen geodesic through 0.

Let $\overline{\mathbf{B}^n}$ denote the closure of $\mathbf{B}^n$ in $\mathbb{P}^n$. The boundary sphere $S^{2n-1} = \partial\mathbf{B}^n$ of $\mathbf{B}^n$ is the projection to $\mathbb{P}^n$ of the null-cone of the form q. The sphere S^{2n-1} is a *CR manifold*: It is equipped with a smooth totally nonintegrable hyperplane distribution $H_z, z \in S^{2n-1}$,

$$H_z = T_z S^{2n-1} \cap J(T_z S^{2n-1}),$$

where J is the almost complex structure on $\mathbb{P}^n$. The subspace H_z is a (complex) hyperplane in $T_z\mathbb{P}^n$. We let P_z denote the unique projective subspace in $\mathbb{P}^n$ passing through z and tangent to H_z. Thus, $P_z \cap \overline{\mathbf{B}^n} = \{z\}$.

One defines a *sub-Riemannian metric* d_C on S^{2n-1} as follows. Given points $\xi, \eta \in S^{2n-1}$, define $C_{p,q}$ as the collection of smooth paths $c : [0, 1] \to S^{2n-1}$ connecting p to q such that c is a *contact path*, i.e. $c'(t) \in H_{c(t)}$ for all $t \in [0, 1]$. Then the *Carnot metric* d_C on S^{2n-1} is

$$d_C(\xi, \eta) = \inf_{c \in C_{\xi,\eta}} \int_0^1 ||c'(t)||dt,$$

where $|| \cdot ||$ is a background Riemannian metric on S^{2n-1}, say, the unique metric of sectional curvature $+1$ invariant under $O(2n)$. It turns out that d_C is indeed a metric which topologizes S^{2n-1}. However, unlike a Riemannian metric on S^{2n-1}, which has Hausdorff dimension equal to the topological dimension, the metric space (S^{2n-1}, d_C) is *fractal*, its Hausdorff dimension $\dim_H$ is equal to

$$\dim_H(S^{2n-1}, d_C) = 2n.$$

Most of the following discussion is valid for general negatively pinched Hadamard spaces; I refer to the paper by Bowditch [11] for details, especially in the context of discrete isometry groups.

Since $\mathbb{H}^n_{\mathbb{C}}$ is a Hadamard manifold X, it has an *intrinsically defined* ideal (visual) boundary $\partial_\infty X$, defined as the set of equivalence classes of geodesic rays, where two rays are equivalent if and only if they are within finite Hausdorff distance. Every geodesic ray is equivalent to a geodesic ray emanating from a chosen base-point $o \in X$. The topology on $\partial_\infty X$ is the quotient topology, where the space of geodesic rays is equipped with the topology of uniform convergence on compacts. Equivalently, since the map from the unit tangent sphere UT_oX at o to $\partial_\infty X$ is bijective, $\partial_\infty X$ is homeomorphic to UT_oX. The union $\overline{X} := X \cup \partial_\infty X$ also has a natural topology with respect to which it is homeomorphic to the closed ball. Given a subset $Y \subset X$, we define $\partial_\infty Y$ as the intersection of the closure of Y in $\overline{X}$ with $\partial_\infty X$.

If X is strictly negatively curved, it satisfies the *visibility property*: Any two distinct points $\xi, \eta \in \partial_\infty X$ are connected by a unique geodesic, denoted $\xi\eta$.

In the case $X = \mathbb{H}^n_{\mathbb{C}}$, this abstract compactification is naturally homeomorphic to the closed ball compactification $\overline{\mathbf{B}^n}$: Two geodesic rays c_1, c_2 are equivalent if and only if they terminate at the same point of the boundary sphere S^{2n-1}.

Suppose that X is a Hadamard manifold. Given a closed subset $\Lambda \subset \partial_\infty X$, one defines the *closed convex hull*, denoted $\mathrm{hull}(\Lambda)$, of Λ in X as the intersection of all closed subsets $C \subset X$ such that $\partial_\infty C \supset \Lambda$. For $\eta > 0$ we will use the notation $\mathrm{hull}_\eta(\Lambda)$ to denote the closed η-neighborhood of $\mathrm{hull}(\Lambda)$ in X.

Theorem 2.1 (M. Anderson [3]) *If X has pinched negative curvature then for every closed subset $\Lambda \subset \partial_\infty X$ which is not a singleton,* $\mathrm{hull}(\Lambda)$ *is a (closed, convex) subset of X such that* $\partial_\infty \mathrm{hull}(\Lambda) = \Lambda$.

Remark 2.1

(a) Assuming that X is negatively curved:

1. $\mathrm{hull}(\Lambda) = \emptyset$ if and only if Λ consists of at most one point.
2. For any two distinct points $\xi, \eta \in \partial_\infty X$, $\mathrm{hull}(\{\xi, \eta\}) = \xi\eta$.

(b) Anderson's theorem fails for the Euclidean plane $X = E^2$.

Anderson's theorem requires negative pinching: It fails if X merely has strictly negative curvature, see [2].

The geometry of convex hulls remains a bit of a mystery, for instance we still do not entirely understand volumes of convex hulls of finite subsets. The best known result seems to be:

Theorem 2.2 (A. Borbély [9]) *If X is m-dimensional, has curvature in the interval $[-k^2, -1]$ and Λ has cardinality $\le n$, then $Vol(\mathrm{hull}(\Lambda)) \le Cn^{2-\eta}$, where $C = C(m, k)$, while*

$$\eta = \frac{1}{1 + 4k(m-1)}.$$

For a closed subset $\Lambda \subset \partial \mathbf{B}^n$, define its *tangent hull* $\hat{\Lambda}$ as the union of hyperplanes $P_\lambda, \lambda \in \Lambda$. I will refer to the hyperplanes $P_\lambda, \lambda \in \Lambda$ as the *complex support hyperplanes* of Λ. Similarly, for an open subset $\Omega = \partial \mathbf{B}^n - \Lambda$, define

$$\check{\Omega} = \mathbb{P}^n - \hat{\Lambda}. \tag{2.1}$$

Remark 2.2 $\hat{\Lambda}$ is also closed and $\hat{\Lambda} \cap \overline{\mathbf{B}^n} = \Lambda$.

See Appendix A for a discussion of *horospheres* and *horoballs* in Hadamard manifolds X and the *horofunction compactification* of X, which leads to an alternative description of the topology on $\overline{X}$.

Isometries of X extend to homeomorphisms of $\overline{X}$; in the setting of $\mathbf{B}^n$, this is just the fact that all automorphisms of $\mathbf{B}^n$ are restrictions of projective transformations:

$$PU(n, 1) < PGL(n+1, \mathbb{C}).$$

The group $G = PU(n, 1)$ acts doubly transitively on the boundary sphere S^{2n-1}: Given two pairs of distinct points $\xi_i, \eta_i, i = 1, 2$, we connect these points by unique biinfinite (unit speed) geodesics $c_i = \xi_i\eta_i$. Set $z_i := c_i(0), v_i := c'(0) \in T_{z_i}\mathbf{B}^n$. Then, since G acts transitively on the unit tangent bundle $UT\mathbf{B}^n$, there exists $g \in G$ sending $v_1 \mapsto v_2$. Thus, $g(c_1) = c_2$ and, consequently, $g(\xi_1) = \xi_2$, $g(\eta_1) = \eta_2$.

Classification of Isometries Every isometry $g \in G = Aut(\mathbf{B}^n)$ is continuous on the closed ball $\overline{\mathbf{B}^n}$ and, hence, has at least one fixed point there. Accordingly, automorphisms $g \in G$ are classified as:

1. **Elliptic**: g has a fixed point z in $\mathbf{B}^n$. After conjugating g via $h \in Aut(\mathbf{B}^n)$ which sends z to 0,

$$hgh^{-1} \in K = U(n).$$

2. **Parabolic**: g has a unique fixed point in $\overline{\mathbf{B}^n}$ and this is a boundary point $z \in S^{2n-1}$. Equivalently,

$$\inf\{d(z, gz) : z \in \mathbf{B}^n\} = 0$$

and the infimum is not realized.

3. **Hyperbolic**: g has exactly two fixed points ξ, η in $\overline{\mathbf{B}^n}$, both are in S^{2n-1}. (In particular, g preserves the unique geodesic $\xi\eta$ in $\mathbf{B}^n$ and acts as a translation along this geodesic. This geodesic is called the *axis* of g.) Equivalently,

$$\inf\{d(z, gz) : z \in \mathbf{B}^n\} \neq 0.$$

This infimum is realized by any point on the axis of g.

The fixed point λ of a hyperbolic isometry γ is called *attractive* (resp. *repulsive*) if for some (every) $x \in X$, $\gamma^i(x) \to \lambda$ as $i \to \infty$ (resp. $i \to -\infty$).

An elliptic automorphism of $\mathbf{B}^n$ is called a *complex reflection* if its fixed-point set is a complex hyperbolic hyperplane in $\mathbb{H}^n_{\mathbb{C}}$.

As any strictly negatively curved Hadamard manifold, $\mathbb{H}^n_{\mathbb{C}}$ satisfies the *convergence property:*

Theorem 2.3 *For every sequence $g_i \in G = PU(n, 1)$, after extraction, the following dichotomy holds:*

(a) Either g_i converges to an isometry $g \in G$.
(b) Or there is a pair of points $\xi, \eta \in S^{2n-1}$ such that $g_i|_{\overline{\mathbf{B}^n} - \{\eta\}}$ converges uniformly on compacts to the constant ξ.

Proof First, one proves the convergence property for sequences of hyperbolic isometries with a common axis. The Cartan decomposition of G then concludes the proof.

In the case (b), I will say that (g_i) converges to the *quasiconstant map* ξ_η. (The point η is the *indeterminacy point* of ξ_η.)

It turns out that most elementary properties of discrete isometry groups of strictly negatively curved Hadamard manifolds can be derived just from the Convergence Property! See [12, 85, 86] for a development of the theory of *convergence group actions* on compact metrizable spaces, i.e. topological group actions satisfying the Convergence Property.

Remark 2.3

1. If $g_i \to \xi_\eta$ then $g_i^{-1} \to \eta_\xi$.
2. If $g_i \to \xi_\eta$, then (g_i) converges (again, uniformly on compacts) to the constant map ξ on $\mathbb{P}^n - P_\eta$.

2.3 Basics of Discrete Subgroups of $PU(n, 1)$

Almost all the properties of discrete subgroups $\Gamma < G = PU(n, 1)$ stated in this section hold for discrete isometry groups of negatively pinched Hadamard manifolds.

Definition 2.3 A subgroup $\Gamma < \mathrm{Isom}(X)$ of isometries of a Riemannian manifold X is called *discrete* if it is discrete as a subset of $\mathrm{Isom}(X)$. Discrete subgroups $\Gamma < PU(n, 1)$ are *complex hyperbolic Kleinian groups*.

Here, all *reasonable* topologies on $\mathrm{Isom}(X)$ agree. For instance, one can use the topology of uniform convergence on compact subsets, or the topology of pointwise convergence.

Recall that a group Γ of homeomorphisms of a topological space X is said to act *properly discontinuously* on X if for every compact $C \subset X$,

$$\mathrm{card}\{\gamma \in \Gamma : \gamma C \cap C \neq \emptyset\} < \infty.$$

Remark 2.4 Suppose that X is a Riemannian manifold and $G = \mathrm{Isom}(X)$ is the isometry group of X.

(a) The following are equivalent for subgroups $\Gamma < G$:

1. Γ is a discrete subgroup of G.
2. Γ acts properly discontinuously on X.

3. For one (equivalently, every) $x \in X$ the function $\Gamma \to \mathbb{R}_+$, $\gamma \mapsto d(x, \gamma x)$ is proper (with Γ equipped with discrete topology), i.e. if γ_i is a sequence consisting of distinct elements of Γ, then

$$\lim_{i\to\infty} d(x, \gamma_i x) = \infty.$$

(b) Every discrete subgroup of G is at most countable.

A group Γ is said to act *freely* on X is for every $x \in X$, the Γ-stabilizer

$$\Gamma_x = \{\gamma \in \Gamma : \gamma x = x\}$$

is the trivial subgroup of Γ.

If X is a manifold and Γ is a group acting freely and properly discontinuously, then the quotient space X/Γ is a manifold and the projection map $X \to X/\Gamma$ is a covering map. If one does not assume freeness of the action then X/Γ is an *orbifold* and the projection map $X \to X/\Gamma$ is an *orbi-covering map*. If X is simply-connected, the group Γ is the (orbifold) fundamental group of X/Γ. See Appendix D for a discussion of orbifolds and related concepts.

In the case when X is a Hadamard manifold, a subgroup $\Gamma < \mathrm{Isom}(X)$ acts freely on X if and only if Γ is *torsion-free*, i.e. every nontrivial element of Γ has infinite order. If Γ acts on X isometrically/holomorphically, the Riemannian metric/complex structure on X descends to the quotient manifold (orbifold) X/Γ.

Definition 2.4 A complex hyperbolic n-dimensional orbifold (manifold) is the quotient of $\mathbb{H}^n_{\mathbb{C}}$ by a discrete (torsion-free) subgroup of $PU(n, 1)$, $M_\Gamma = \mathbb{H}^n_{\mathbb{C}}/\Gamma$.

Remark 2.5 If X is a Hadamard manifold and $\Gamma < \mathrm{Isom}(X)$ is discrete, then Γ is torsion-free if and only if it contains no elliptic elements, besides the identity.

For *finitely generated subgroups* $\Gamma < PU(n, 1)$, one can eliminate torsion by passing to a finite index subgroup:

Theorem 2.4 (Selberg's Lemma, See E.g. [32] or [74]) *If* $\mathbf{k}$ *is a field and* $\Gamma < GL(n, \mathbf{k})$ *is a finitely generated subgroup, then* Γ *is virtually torsion-free, i.e. contains a torsion-free subgroup of finite index.*

In particular, every complex hyperbolic orbifold $\mathcal{O}$ with finitely generated (orbifold) fundamental group, admits a finite-sheeted manifold orbi-covering $M \to \mathcal{O}$.

Remark 2.6 Selberg's theorem fails for discrete finitely generated groups of isometries of negatively pinched Hadamard manifolds, see [52].

Definition 2.5 Given a Hadamard manifold X, a discrete subgroup $\Gamma < \mathrm{Isom}(X)$ and a point $x \in X$, the *limit set* $\Lambda = \Lambda_\Gamma$ is the accumulation set of the orbit Γx in $\partial_\infty X$, i.e.

$$\Lambda = \partial_\infty(\Gamma x).$$

The complement $\Omega := \partial_\infty X - \Lambda$ is called the *discontinuity domain* of Γ.

Remark 2.7 Suppose that Γ is a discrete subgroup of $\text{Isom}(X)$ and X is strictly negatively curved. Then:

1. Λ is independent of $x \in X$.[1]
2. Λ is closed and Γ-invariant. Accordingly, Ω is open in $\partial_\infty X$ and is Γ-invariant as well.
3. Ω is either empty or is dense in $\partial_\infty X$.
4. Either Λ consists of at most two points or it is perfect, i.e. contains no isolated points.
5. If Γ' is a subgroup of Γ, then $\Lambda_{\Gamma'} \subset \Lambda_\Gamma$.
6. If $\Gamma' \lhd \Gamma$ is an infinite normal subgroup then $\Lambda_{\Gamma'} = \Lambda_\Gamma$.
7. If $\Gamma' < \Gamma$ is a subgroup of finite index then $\Lambda_{\Gamma'} = \Lambda_\Gamma$.

Example 2.1 Let $\gamma \in \text{Isom}(X)$ be a non-elliptic element. Then the limit set of the group $\Gamma = \langle\gamma\rangle$ generated by γ is equal to the fixed-point set of γ in $\partial_\infty X$.

Lemma 2.1 *If* $\Gamma < \text{Isom}(X)$ *is a discrete subgroup and* X *is a strictly negatively curved Hadamard manifold, then* Γ *acts properly discontinuously on* $Y = X \cup \Omega$.

Proof Let C be a compact subset of Y. Suppose there exists a sequence consisting of distinct elements $\gamma_i \in \Gamma$ such that for each i, $\gamma_i C \cap C \neq \emptyset$. In view of the Convergence Property, after extraction, the sequence γ_i either converges to an isometry $\gamma \in \text{Isom}(X)$ (which would contradict the discreteness of Γ) or to a quasiconstant map ξ_η, with $\xi, \eta \in \Lambda$. Since (γ_i) converges to ξ uniformly on compacts in $\overline{X} - \{\eta\}$ and $C \subset Y \subset \overline{X} - \{\eta\}$ is compact, there exists a neighborhood U of ξ disjoint from C; thus, for all but finitely many values of i, $\gamma_i(C) \subset U$. A contradiction.

A more difficult result is

Theorem 2.5 (A. Cano and J. Seade, See [15, 16]) *Every discrete subgroup* $\Gamma < PU(n, 1)$ *acts properly discontinuously on* $\check{\Omega} := \mathbb{P}^n - \hat{\Lambda}$ *(see (2.1)).*

Remark 2.8 An alternative proof of this result is an application of a proper discontinuity theorem in [55]. More precisely, let $F_{1,n}$ be the flag-manifold consisting of flags (V_1, V_n) in $V = \mathbb{C}^{n+1}$, where V_1 is a line and V_n is a hyperplane (containing V_1). We have a G-equivariant holomorphic fibration $\pi : F_{1,n} \to \mathbb{P}^n$ sending each pair (V_1, V_n) to V_1. The tangent hull $\hat{\Lambda}$ of Λ defines a natural continuous map $\theta : \Lambda \to F_{1,n}$ sending each $\lambda \in \Lambda$ to the pair (V_1, V_n) consisting of the preimages of λ and P_λ in V. Let $\tilde{\Lambda}$ be the image of θ and let $Th(\tilde{\Lambda})$ be the *thickening* of $\tilde{\Lambda}$ in $F_{1,n}$, consisting of flags (V_1', V_n') such that either V_1' belongs to Λ or V_n' is a complex support hyperplane of Λ. Then Γ acts properly discontinuously on $\Omega_{Th} = F_{1,n} - Th(\tilde{\Lambda})$; see [55]. Since $\pi^{-1}(\check{\Omega}) \subset \Omega_{Th}$, the action of Γ on $\check{\Omega}$ is properly discontinuous as well.

[1] This also holds for general Hadamard manifolds even though the convergence property fails.

In particular, the quotient $\overline{M}_\Gamma := (\mathbf{B}^n \cup \Omega)/\Gamma$ embeds as an orbifold with boundary in the complex orbifold without boundary $\check{\Omega}/\Gamma$. The boundary of $\overline{M}_\Gamma$ (equal to Ω/Γ) is strictly Levi-convex in $\check{\Omega}/\Gamma$.

Notation The *boundary* ∂M_Γ of a complex hyperbolic orbifold M_Γ is Ω_Γ/Γ; in other words, this is the boundary of $\overline{M}_\Gamma$. □

We now return to the discussion of discrete subgroups of general negatively pinched Hadamard manifolds X.

Theorem 2.6 *If α, β are hyperbolic elements of a discrete subgroup of* Isom(X), *then their fixed-point sets are either equal or disjoint.*

Corollary 2.1 *If $\Gamma <$ Isom(X) is discrete and fixes a point $\lambda \in \partial_\infty X$ then Λ_Γ either equals to $\{\lambda\}$ and Γ contains no hyperbolic element, or Λ_Γ consists of two points, $\Lambda_\Gamma = \{\lambda, \lambda'\}$ and Γ contains no parabolic element.*

Definition 2.6 A discrete subgroup $\Gamma <$ Isom(X) is called *elementary* if card$(\Lambda_\Gamma) \leq 2$. It is said to be *nonelementary* otherwise.

Elementary subgroups are, in many ways, exceptional, among discrete subgroups.

In view of Remark 2.7(4), the limit set of every nonelementary subgroup is perfect. In particular, it has the cardinality of the continuum. Hence:

Proposition 2.1 *The limit set of a discrete subgroup of* Isom(X) *consists of 0, 1, 2 or a continuum of points.*

Proposition 2.2 *The limit set of a nonelementary discrete group Γ is the smallest nonempty closed Γ-invariant subset of $\partial_\infty X$. In particular, every orbit in Λ_Γ is dense.*

Proof Suppose that $L \subsetneq \Lambda_\Gamma$ is a closed nonempty and Γ-invariant subset. Take a point $\xi \in \Lambda_\Gamma - L$ and let (γ_i) be a sequence in Γ converging to a quasiconstant map ξ_η. Then for every $\lambda \in L - \{\eta\}$, $\lim_{i\to\infty} \gamma_i(\lambda) = \xi$. Since L is closed and $\xi \notin L$, for all sufficiently large i, $\gamma_i(\lambda) \notin L$, contradicting the invariance of L. This leaves us with the possibility that L is the singleton $\{\xi\}$ and ξ is fixed by the entire Γ. It then follows that Γ is elementary.

Theorem 2.7 *Suppose that Γ is an elementary subgroup of* Isom(X).

1. *If Λ_Γ is a singleton then every element of Γ is elliptic or parabolic.*
2. *If Λ_Γ consists of two points then every element of Γ is elliptic or hyperbolic. Hyperbolic elements fix Λ_Γ pointwise. Elliptic elements can swap the two limit points.*
3. *Γ is a virtually nilpotent[2] group.*

[2] I.e. contains a nilpotent subgroup of finite index.

See [6] for a more detailed discussion of elementary groups and their quotient spaces M_Γ. Here we only note that discrete elementary subgroups of $PU(n, 1)$ are virtually *2-step* nilpotent.

Proposition 2.3 *Suppose that X is a strictly negatively curved Hadamard manifold. If ξ, η are distinct limit point of a discrete subgroup $\Gamma < \mathrm{Isom}(X)$ then there exists a sequence $\gamma_k \in \Gamma$ of hyperbolic elements whose attractive (resp. repulsive) fixed points converge to ξ (resp. η).*

Proof Since ξ, η are limit points of Γ, there exist sequences (g_i), (h_j) in Γ which converge, respectively, to the quasiconstant maps ξ_α and β_η. By precomposing these sequences with suitable elements of Γ, we can assume that the points ξ, η, α, β are pairwise distinct. Let $U_\alpha, U_\beta, U_\xi, U_\eta$ be pairwise disjoint open ball neighborhoods in $\partial_\infty X$ of α, β, ξ, η respectively. In view of the convergence $g_i \to \xi_\alpha, h_j \to \beta_\eta$, for all sufficiently large i we have

$$h_i(\partial_\infty X - U_\eta) \subset U_\beta, \quad g_i(\partial_\infty X - U_\alpha) \subset U_\xi,$$

and, hence,

$$g_i \circ h_i(\partial_\infty X - U_\eta) \subset U_\xi.$$

In particular, the composition $f_i = g_i \circ h_i$ has an attractive fixed point in U_ξ. Similarly, f_i^{-1} has an attractive fixed point in U_η.

Corollary 2.2 *If Γ is nonelementary then the set of hyperbolic fixed points of elements of Γ is dense in Λ_Γ.*

Corollary 2.3 *If a discrete group Γ contains a parabolic element then parabolic fixed points are dense in Λ_Γ.*

The following theorem provides a converse to Theorem 2.7(3):

Theorem 2.8 *Each nonelementary discrete subgroup $\Gamma < \mathrm{Isom}(X)$ contains a nonabelian free subgroup whose limit set is homeomorphic to the Cantor set.*

Definition 2.7 The *convex core*, $Core(M)$, of $M = M_\Gamma = X/\Gamma$ is the projection to M_Γ of the closed convex hull $\mathrm{hull}(\Lambda_\Gamma)$ of the limit set of Γ.

Given $\eta > 0$, define $Core_\eta(M)$ as the projection to M_Γ of $\mathrm{hull}_\eta(\Lambda_\Gamma)$. Intrinsically, the convex core can be defined as:

Remark 2.9 $Core(M)$ is the intersection of all closed convex suborbifolds $M' \subset M$ such that $\pi_1(M') \to \pi_1(M)$ is surjective.

Conical Limit Points I conclude this section with a discussion of a classification of limit points of discrete subgroups of $\mathrm{Isom}(X)$.

Definition 2.8 A sequence (x_i) in X is said to converge to a point $\xi \in \partial_\infty X$ *conically* if there exists a geodesic ray $x\xi$ in X and a constant $R < \infty$ such that: $d(x_i, x\xi) \leq R$ for all i and $\lim_{i\to\infty} x_i = \xi$.

Remark 2.10 Let $\lambda \in \Lambda_\Gamma$ be a limit point. The following are equivalent:

1. There exists a sequence $\gamma_i \in \Gamma$ such that the sequence $(\gamma_i(x))$ converges to ξ conically.
2. The projection of the ray $x\lambda$ to M_Γ defines a non-proper map $\mathbb{R}_+ \to M_\Gamma$.

Definition 2.9 A limit point $\lambda \in \Lambda_\Gamma$ is called *conical* or *radial* if it satisfies one of the two equivalent properties in the above remark. The set of conical limit points of Γ is denoted $\Lambda^c = \Lambda^c_\Gamma$.

Example 2.2

1. If Γ is an elementary hyperbolic subgroup of $\mathrm{Isom}(X)$ then $\Lambda_\Gamma = \Lambda^c_\Gamma$.
2. If Γ is an elementary parabolic subgroup of $\mathrm{Isom}(X)$ then $\Lambda^c_\Gamma = \emptyset$.

2.4 Margulis Lemma and Thick-Thin Decomposition

In this section, X is a negatively pinched Hadamard manifold. For each discrete subgroup $\Gamma < \mathrm{Isom}(X)$, a point $x \in X$ and a number $\epsilon > 0$, define $\Gamma_{x,\epsilon}$ to be the subgroup of Γ generated by the (necessarily finite) set

$$\{\gamma \in \Gamma : d(x, \gamma x) < \epsilon\}.$$

This subgroup is the "almost-stabilizer" of x in Γ.

Let $U_{\Gamma,\epsilon}$ denote the subset of X consisting of points x for which the almost-stabilizer $\Gamma_{x,\epsilon}$ is infinite.

The components of $U_{\Gamma,\epsilon}$ need not be convex (already for $X = \mathbb{H}^2_{\mathbb{C}}$), but each component is contractible:

Proposition 2.4 *Each component of $U_{\Gamma,\epsilon}$ is contractible.*

In view of the contractibility of X and of hull Λ_Γ, it follows that $X - U_{\Gamma,\epsilon}$ and hull $\Lambda_\Gamma - U_{\Gamma,\epsilon}$ are both contractible. Furthermore, if X has curvature ≤ -1, each component U of $U_{\Gamma,\epsilon}$ is *uniformly quasiconvex*:

Theorem 2.9 *There exist universal constants δ_0, η_0 such that each component U of $U_{\Gamma,\epsilon}$ satisfies:*

1. *For any two points $x, y \in U$, the geodesic xy is contained in the δ_0-neighborhood of U.*
2. *The η_0-neighborhood of U is convex.*

Theorem 2.10 (Kazhdan–Margulis; Margulis; See E.g. [4]) *Let X be an n-dimensional Hadamard manifold of sectional curvature bounded below by $b \leq 0$.*

Then there exists $\epsilon = \epsilon(n, b)$ *such that for every discrete subgroup* $\Gamma < \mathrm{Isom}(X)$ *and every* $x \in X$*, the subgroup* $\Gamma_{x,\epsilon}$ *is virtually nilpotent. In particular, if* X *is negatively curved, then* $\Gamma_{x,\epsilon}$ *is elementary.*

Corollary 2.4 *For each discrete torsion-free subgroup* $\Gamma < \mathrm{Isom}(X)$*, the set* $U_{\Gamma,\epsilon}$ *breaks into connected components* $X_{\Gamma,\epsilon,i}$ *each of which is stabilized by some elementary subgroup* Γ_i *of* Γ *and for each* $x \in X_{\Gamma,\epsilon,i}$ *the stabilizer* Γ_i *contains the "almost stabilizer"* $\Gamma_{x,\epsilon}$*. (The index can be infinite.)*

As a corollary, one obtains the *thick-thin decomposition* of the orbifold $M = M_\Gamma$: $M_{(0,\epsilon)}$ is the projection of $U_{\Gamma,\epsilon}$ to M. It consists of all points $y \in M$ for which there exists a homotopically nontrivial loop based at y of length $< \epsilon$. Define also $M_{(0,\epsilon]}$ as the closure of $M_{(0,\epsilon)}$ in M. Both $M_{(0,\epsilon)}$ and $M_{(0,\epsilon]}$ are called the ϵ-*thin* parts of M. The complement $M_{[\epsilon,\infty)} = M - M_{(0,\epsilon)}$ and its interior $M_{(\epsilon,\infty)}$ are called the ϵ-*thick* parts of M.

One defines the ϵ-thick, resp. thin, part of the convex core $Core(M)$ as the intersection $Core(M) \cap M_{[\epsilon,\infty)}$, resp. $Core(M) \cap M_{(0,\epsilon)}$.

Components of the thin parts M and $Core(M)$ come in two shapes:

(a) **Tubes**. Suppose that U is a component of $U_{\Gamma,\epsilon}$ whose stabilizer Γ_U in Γ is virtually hyperbolic, i.e. contains a cyclic hyperbolic subgroup of finite index. In other words, the limit set of Γ_U consists of two points ξ, η. The geodesic $\xi\eta$ is then invariant under Γ_U; it is also contained in U and projects to a closed geodesic $c \subset U/\Gamma_U$. The quotient U/Γ_U is a *tube*: If Γ_U is torsion-free then this quotient is homeomorphic to an $\mathbb{R}^k$-bundle over S^1, with the base of the fibration corresponding to the closed geodesic c.

(b) **Cusps**. Suppose that U is a component of $U_{\Gamma,\epsilon}$ whose stabilizer Γ_U in Γ is virtually parabolic, i.e. contains a parabolic subgroup of finite index. In other words, the limit set of Γ_U consists of a single point η. The group Γ_U preserves horoballs B_η based at η. The subsets $U_{\Gamma,\epsilon}$ are typically strictly smaller (not even Hausdorff-close) than any of the horoballs B_η.

2.5 Geometrically Finite Groups

The notion of geometrically finite Kleinian group was introduced by Lars Ahlfors in the mid 1960s for the real hyperbolic space and later generalized (by William Thurston and Brian Bowditch) to manifolds of negative curvature: The discrete groups in this class are the *nicest-behaving* among discrete isometry groups of negatively pinched Hadamard manifolds.

Definition 2.10 Let X be a negatively pinched Hadamard manifold. A discrete subgroup $\Gamma < G = \mathrm{Isom}(X)$ is called *geometrically finite* if:

(a) The orders of elliptic elements of Γ are uniformly bounded (from above), and
(b) the volume of $Core_\eta(M_\Gamma)$ is finite for some (equivalently, every, $\eta > 0$).

A discrete subgroup $\Gamma < G$ is called *convex-cocompact* if $\mathrm{card}(\Lambda_\Gamma) \neq 1$ and $Core(M_\Gamma)$ is compact.

For instance, if $\Lambda_\Gamma = \partial_\infty X$ then $\mathrm{hull}(\Lambda_\Gamma) = X$ and, thus, Γ is geometrically finite if and only if $\Gamma < G$ is a *lattice*, i.e. $vol(M_\Gamma) < \infty$. Under the same assumption, Γ is convex-cocompact if and only if $\Gamma < G$ is a *uniform lattice*, i.e. M_Γ is compact.

Theorem 2.11

1. *(B. Bowditch, [11]) A discrete subgroup $\Gamma < G$ is geometrically finite if and only if the ϵ-thick part of $Core(M_\Gamma)$ is compact.*
2. *(B. Bowditch, [11]) A discrete subgroup $\Gamma < G$ is convex-cocompact if and only if $\overline{M}_\Gamma$ is compact.*
3. *(B. Bowditch, [11]) A discrete subgroup $\Gamma < G$ is convex-cocompact if and only if every limit point of Γ is conical.*
4. *(M. Kapovich, B. Liu, [54]) A discrete subgroup $\Gamma < G$ is geometrically finite if and only if every limit point of Γ is either conical or a parabolic fixed point.*

In particular, (1) implies that geometrically finite groups are finitely presentable (since $\mathrm{hull}\,\Lambda_\Gamma - U_{\Gamma,\epsilon}$ is contractible).

In particular, a convex-cocompact subgroup $\Gamma < PU(n,1)$ acts properly discontinuously and cocompactly on $\mathbb{H}^n_{\mathbb{C}} \cup \Omega$. The action of Γ on $\check{\Omega}$ is properly discontinuous but not cocompact. If becomes cocompact if we lift it to the flag-manifold $F_{1,n}$ (see [55]):

Theorem 2.12 *The Γ-action on the domain $\Omega_{Th} \subset F_{1,n}$ is properly discontinuous and cocompact.*

2.6 Ends of Negatively Curved Manifolds

Let X be a negatively pinched Hadamard manifold and let Λ be a closed subset of $\partial_\infty X$ consisting of at least two points. Set $\Omega = \partial_\infty X - \Lambda$. The nearest-point projection $\Pi : X \to \mathrm{hull}(\Lambda)$ extends continuously to a map $\Pi : X \cup \Omega \to \mathrm{hull}(\Lambda)$: While for $x \in X$, $\Pi(x)$ is defined by minimizing the distance function $d_x = d(x, \cdot)$ on $\mathrm{hull}(\Lambda)$, for $\xi \in \Omega$, the projection $\Pi(\xi)$ is defined by minimizing the Busemann function b_ξ based at ξ. For a component $\Omega_0 \subset \Omega$ we define a subset $X_0 \subset X$ as the union of open geodesic rays $x\xi - \{x\}$, where $\xi \in \Omega_0, x = \Pi(\xi)$. The union of these geodesic rays is an open subset of $X - \mathrm{hull}(\Lambda)$ whose closure in $X \cup \Omega$ equals $X_0 \cup \Omega_0 \cup \Pi(\Omega_0)$.

We now specialize to the setting when $\Lambda = \Lambda_\Gamma$ is the limit set of a discrete subgroup $\Gamma < \mathrm{Isom}(X)$. If Ω_0 has cocompact stabilizer Γ_0 in Γ, then Γ_0 also acts cocompactly on $X_0 \cup \Omega_0 \cup \Pi(\Omega_0)$. Thus, M_Γ has an *isolated end* E_0 corresponding to Ω_0/Γ_0, with isolating neighborhood X_0/Γ_0.

Definition 2.11 Ends E_0 of $M = M_\Gamma$ which have this form are called *convex ends* of M.

From the analytical viewpoint, the advantage of working with convex ends E_0 is that they admit *convex exhaustion functions*: For every convex end E_0 there exists a convex function $\phi : M \to \mathbb{R}_+$ which is proper on the closure of E_0 and vanishes on $M - E_0$.

Suppose that C is an unbounded component of the thin part $M_{(0,\epsilon)}$ of $M = M_\Gamma$, and C has compact boundary. Then C also defines an isolated end E_C with an isolating neighborhood given by $C \cap M_{(0,\epsilon)}$.

Definition 2.12 Ends E_C of M_Γ which have this form are called *cuspidal ends* of M_Γ.

Remark 2.11

1. Γ is convex-cocompact if and only if M_Γ has only convex ends.
2. If M_Γ has only convex and cuspidal ends then Γ is geometrically finite.

One can refine (cf. [49]) the above definitions in two ways:

(a) Considering unbounded components of the thin part of $Core(M_\Gamma)$ and, thus, defining cuspidal ends of the convex core.
(b) Removing from M_Γ its cuspidal ends and their preimages under the nearest-point projection $M_\Gamma \to Core(M_\Gamma)$, one then defines *relative* convex ends of M_Γ.

One can also classify ends of M_Γ using potential theory as *hyperbolic* and *parabolic* ends, see [68]. Note that if $M = M_\Gamma$ is a complex hyperbolic manifold, then every convex end E of M is hyperbolic.

2.7 Critical Exponent

Notation Let $B(x, r)$ denote the open ball of radius r and center at x in a metric space.

I will discuss the critical exponent mostly in the case of complex hyperbolic Kleinian groups; for a discussion in the broader context of negatively curved Hadamard manifolds and Gromov-hyperbolic spaces see e.g. [18, 24, 25, 60, 75].

The *critical exponent* of a discrete isometry group Γ of a Hadamard manifold X (typically, satisfying some further curvature restrictions) is, probably, the single most important numerical invariant of Γ: It reflects the geometry of Γ-orbits in X, the geometry of the limit set of Γ, the ergodic theory of the action of Γ on the limit set and analytic properties of the quotient space X/Γ. Its origin goes back to the nineteenth century and the work of Poincaré (among others), who was interested in constructing *automorphic functions* (and forms) on the hyperbolic plane by "averaging" a certain holomorphic function (or a form) over a discrete isometry group Γ. The resulting infinite series (the *Poincaré series*) may or may not converge, depending on the *weight* of the form, leading to the notion of *critical exponent* or *exponent of convergence* of Γ.

Let $\Gamma < \mathrm{Isom}(X)$, a discrete isometry group of a Hadamard manifold. Pick points $x, y \in X$. The *entropy* of Γ is defined as

$$\delta = \delta_\Gamma = \lim\sup_{r\to\infty} \frac{1}{r} \mathrm{card}(B(x,r) \cap \Gamma y).$$

Thus, the entropy measures the rate of exponential growth of Γ-orbits in X. It turns out that δ is equal to the *critical exponent* of Γ, defined as

$$\delta = \inf\{s : \sum_{\gamma\in\Gamma} \exp(-sd(x, \gamma y)) < \infty\},$$

i.e. δ is the *exponent of convergence* of the Poincaré series $\sum_{\gamma\in\Gamma} \exp(-sd(x, \gamma y))$. Furthermore, δ is independent of the choice of $x, y \in X$. If

$$\sum_{\gamma\in\Gamma} \exp(-\delta d(x, \gamma y)) < \infty$$

(which depends only on Γ and not on the choice of x, y), then Γ is said to be a subgroup of *convergence type*; otherwise, Γ is said to be of *divergence type*.

Below are equivalent characterizations of δ in the case $X = \mathbb{H}^n_{\mathbb{C}}$:

Theorem 2.13 *Suppose that $\Gamma < PU(n,1)$ is a discrete subgroup.Then:*

1. *(Corlette [20]; Corlette–Iozzi [22], Theorem 6.1) $\delta = \delta_\Gamma$ is equal to the Hausdorff dimension $\dim_H \Lambda^c_\Gamma$, where the conical limit set Λ^c_Γ is equipped with the restriction of the Carnot metric on S^{2n-1}. In particular, if Γ is geometrically finite then $\delta = \dim_H \Lambda$.*
2. *(Elstrodt–Patterson–Sullivan–Corlette–Leuzinger, see [61, Corollary 1]) Let $\lambda = \lambda(M_\Gamma)$ denote the bottom of the L^2-spectrum of the Laplacian on M_Γ. Then*

$$\begin{cases} \lambda = n^2 & \text{if } 0 \le \delta \le n \\ \lambda = \delta(2n-\delta) & \text{if } n \le \delta \le 2n \end{cases}$$

2.8 Examples

I will say that a discrete torsion-free subgroup $\Gamma < G = PU(n,1)$ is *Stein* if the complex manifold M_Γ is Stein.

I will start with two elementary examples.

Example 2.3 (Cyclic Hyperbolic Groups) Let $\gamma \in PU(n,1)$ be a hyperbolic isometry fixing points $\lambda_\pm \in S^{2n-1} = \partial_\infty \mathbb{H}^n_{\mathbb{C}}$ and let $\Gamma = \langle\gamma\rangle$ be the cyclic subgroup of $PU(n,1)$ it generates. Then Γ is an elementary subgroup with the limit set $\Lambda = \{\lambda_-, \lambda_+\}$. The quotient manifold $M_\Gamma = \mathbb{H}^n_{\mathbb{C}}/\Gamma$ is diffeomorphic to the product $\mathbb{R}^{2n-1} \times S^1$ while $\overline{M}_\Gamma$ is diffeomorphic to the product $\bar{D}^{2n-1} \times S^1$, where $\bar{D}^{2n-1}$ is the closed disk of real dimension $2n - 1$.

Example 2.4 (Integer Heisenberg Groups) Given a natural number n, define the $2n+1$-dimensional real Lie group H_{2n+1} as the group of $(n+2) \times (n+2)$-matrices

$$\begin{bmatrix} 1 & \mathbf{a} & c \\ 0 & I_n & \mathbf{b} \\ 0 & 0 & 1 \end{bmatrix},$$

where I_n is the identity $n \times n$ matrix, $\mathbf{a} \in \mathbb{R}^n$ is a row-vector, $\mathbf{b} \in \mathbb{R}^n$ is a column-vector and $c \in \mathbb{R}$. This group is 2-step nilpotent with 1-dimensional center consisting of the matrices with $\mathbf{a} = \mathbf{b} = 0$ and $c \in \mathbb{R}$. The quotient of H_{2n+1} by its center is the $2n$-dimensional commutative Lie group isomorphic to $\mathbb{R}^{2n}$. The real Heisenberg group H_{2n+1} contains the integer Heisenberg group $H_{2n+1}(\mathbb{Z})$, defined as the intersection

$$H_{2n+1} \cap SL(n+2, \mathbb{Z}).$$

The quotient $N = H_{2n+1}/H_{2n+1}(\mathbb{Z})$ is a compact *nil-manifold*, which is a nontrivial circle over the torus T^{2n}. Algebraically, in terms of its presentation, $H_{2n+1}(\mathbb{Z})$ is given by

$$\langle x_1, y_1, ..., x_n, y_n, t | [x_i, t] = [y_j, t] = 1, [x_i, y_i] = t, i = 1, ..., n, j = 1, ..., n \rangle.$$

The Heisenberg group H_{2n+1} embeds in $PU(n+1, 1)$, fixing a point ξ in $\partial_\infty \mathbb{H}^{n+1}_{\mathbb{C}}$ and acting simply-transitively on every horosphere in $\mathbb{H}^{n+1}_{\mathbb{C}}$ centered at ξ. Thus, $H_{2n+1}(\mathbb{Z})$ embeds as a discrete elementary subgroup $\Gamma < PU(n+1, 1)$ such that M_Γ is diffeomorphic to $N \times (0, \infty)$. The partial compactification $\overline{M}_\Gamma$ is diffeomorphic to $N \times [0, \infty)$.

The rest of our examples are nonelementary.

Example 2.5 (Schottky Groups) These are convex-cocompact subgroups $\Gamma < G$ isomorphic to free nonabelian groups F_k of finite rank k. The limit set Λ_Γ is homeomorphic to the Cantor set. Its Hausdorff dimension is positive but can be arbitrarily close to 0. Schottky groups are always Stein. Every nonelementary discrete subgroup contains a Schottky subgroup. Schottky subgroups can be found via the following procedure. Let $\gamma_1, ..., \gamma_k$ be hyperbolic isometries with pairwise disjoint fixed-point sets. Then there exists t_0 such that for each integer $t \geq t_0$, the subgroup generated by $s_1 = \gamma_1^t, ..., s_k = \gamma_k^t$ is a Schottky group with free generating set $s_1, ..., s_k$.

Example 2.6 (Schottky-Type Groups) These are geometrically finite subgroups $\Gamma < G$ isomorphic to free products of elementary subgroups of G, such that the limit set Λ_Γ is homeomorphic to the Cantor set. Schottky-type subgroups can be

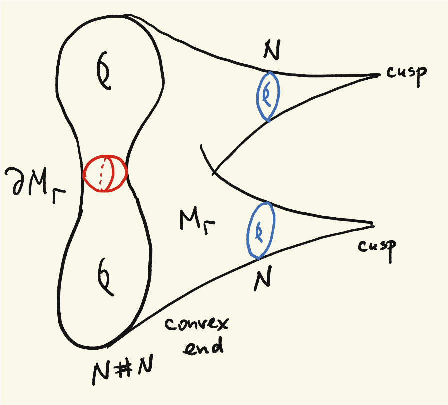


Fig. 2.1 Quotient manifold of a Schottky-type group with $k = 2$

found via the following procedure. Let $\Gamma_1, ..., \Gamma_k$ be elementary subgroups with pairwise disjoint limit sets. Then there exist torsion-free finite-index subgroups $\Gamma_i^\ell < \Gamma_i, i = 1, ..., k$, such that the subgroup generated by

$$\Gamma_1^\ell, ..., \Gamma_k^\ell$$

is Schottky-type and the homomorphism

$$\Gamma_1^\ell \star ... \star \Gamma_k^\ell \to \Gamma = \langle \Gamma_1^\ell, ..., \Gamma_k^\ell \rangle$$

sending $\Gamma_i^\ell \to \Gamma_i^\ell, i = 1, ..., k$, is an isomorphism. For instance, suppose that $\Gamma_1, ..., \Gamma_k$ are integer Heisenberg subgroups of G. Then M_Γ has k cuspidal ends (diffeomorphic to $N \times (0, \infty)$) and one convex end, with ∂M_Γ diffeomorphic to the k-fold connected sum of N with itself, where $N = H_{2n-1}/H_{2n-1}(\mathbb{Z})$. See Fig. 2.1.

Real and complex Fuchsian groups defined below were introduced by Burns and Shnider in [14].

Example 2.7 (Real-Fuchsian Subgroups) Let $\mathbb{H}^2_\mathbb{R} \subset \mathbb{H}^n_\mathbb{C}$ be a real hyperbolic plane in $\mathbb{H}^n_\mathbb{C}$. Let $\Gamma < PU(n, 1)$ be a geometrically finite subgroup whose limit set is $\partial_\infty \mathbb{H}^2_\mathbb{R}$. Then Γ preserves $\mathbb{H}^2_\mathbb{R}$ and acts on it with quotient of finite area. The quotient surface-orbifold Σ is the convex core of M_Γ. The limit set of Γ has Hausdorff dimension 1. Assume now that $n = 2$, Γ is torsion-free and Σ is compact. Then M_Γ is diffeomorphic to the tangent bundle of Σ and is Stein.

Example 2.8 (Real Quasi-Fuchsian Subgroups) Let $\Gamma_t, t \in [0, 1]$, be a continuous family of discrete convex-cocompact subgroups of $PU(n, 1)$ such that Γ_0 is real-Fuchsian but other subgroups $\Gamma_t, t > 0$ are not.[3] The subgroups Γ_t are real-quasi-

[3] Such deformations exist as long as Γ_t is, say, torsion-free. More generally, such deformations exist if Γ has trivial center and is not isomorphic to a von Dyck group. See e.g. [89].

Fuchsian subgroups. Their limit sets are topological circles of Hausdorff dimension >1.

Assume that $n = 2$, Γ is torsion-free and Σ is compact. Then M_Γ is diffeomorphic to the tangent bundle of Σ and is Stein.

Example 2.9 (Complex-Fuchsian Subgroups) Let $\mathbb{H}^1_{\mathbb{C}} \subset \mathbb{H}^n_{\mathbb{C}}$ be a complex hyperbolic line in $\mathbb{H}^n_{\mathbb{C}}$. Let $\Gamma < PU(n,1)$ be a geometrically finite subgroup whose limit set is $\partial_\infty \mathbb{H}^1_{\mathbb{C}}$. Then Γ preserves $\mathbb{H}^1_{\mathbb{C}}$ and acts on it with quotient of finite area. The quotient surface-orbifold Σ is the convex core of M_Γ. The limit set of Γ has Hausdorff dimension 2. Let $W \subset V = \mathbb{C}^{n+1}$ be the 2-dimensional complex linear subspace such that the projection of $W \cap V_-$ to $\mathbf{B}^n$ equals $\mathbb{H}^1_{\mathbb{C}}$. The $W^\perp \subset V$ (the complex orthogonal complement with respect to the form q on V) has the property that q restricted to $W^\perp$ is positive-definite. The projection $[W^\perp]$ of $W^\perp$ to $\mathbb{P}^n$ is Γ-invariant. The pair $([W], [W^\perp])$ defines a linear holomorphic fibration of $\mathbb{P}^n - [W^\perp]$ over $[W]$: The fiber through $x \in \mathbb{P}^n - [W^\perp]$ is the unique projective hyperplane passing through x and intersecting transversally both $[W]$ and $[W^\perp]$. Restricting to $\mathbf{B}^n$ we obtain a Γ-invariant holomorphic fibration $\mathbf{B}^n \to \mathbb{H}^1_{\mathbb{C}}$. Projecting to M_Γ we obtain a holomorphic orbi-fibration $M_\Gamma \to \Sigma$, whose fibers are biholomorphic to quotients of $\mathbf{B}^{n-1}$ by finite subgroups of $Aut(\mathbf{B}^{n-1})$. Assume now that $n = 2$, Γ is torsion-free and Σ is compact. Then M_Γ is diffeomorphic to the square root of the tangent bundle of Σ (the spin-bundle) and is not Stein (it contains the compact complex curve Σ).

Convex-cocompact complex Fuchsian groups are locally rigid in the sense that any small deformation of such a group is again complex Fuchsian, [84]. The complex Fuchsian examples generalize to the case of geometrically finite subgroups of $PU(n,1)$ whose limit sets are ideal boundaries of k-dimensional complex hyperbolic subspaces $\mathbb{H}^k_{\mathbb{C}} \subset \mathbb{H}^n_{\mathbb{C}}$. The rigidity theorem holds in this case as well, see [13, 19, 36].

Example 2.10 (Hybrid Groups) One can combine, say, torsion-free, real and complex Fuchsian groups in a variety of ways. For instance, one can form free products of such groups. The nature of the quotient manifolds will depend on the precise way in which the free factors are interacting with each other. For instance, in the case $n = 2$ the boundary of M_Γ can be either a connected sum, or the *toral sum* of certain circle bundles over surfaces. One can also break real and complex Fuchsian groups into smaller pieces and consider amalgams over $\mathbb{Z}$ of these pieces. As a result, one can get for instance, circle bundles over surfaces other than the unit tangent bundle and its square root, see [37] and [1] for more detail.

Example 2.11 (AGG Subgroups: Anan'in–Grossi–Gusevskii, [1]) These interesting examples of convex-cocompact subgroups of $PU(2,1)$ are isomorphic images of *von Dyck groups* $D(2,n,n)$, for $n \in \{10\} \cup [13, 1001]$. None of these subgroups is complex Fuchsian or real quasi-Fuchsian. According to Proposition C.6, these subgroups are locally rigid in $PU(2,1)$: Every small deformation is conjugate in $PU(2,1)$ to the original subgroup. The limit set is a topological circle but is neither a complex nor a real circle. Fix a (unique up to conjugation) discrete, faithful and isometric action of $D(2, n, n)$ on $\mathbb{H}^1_{\mathbb{C}}$. For each embedding $\rho : D(2, n, n) \to \Gamma < PU(2, 1)$ constructed in Section 3.3 of [1], the complex hyperbolic orbifold M_Γ is diffeomorphic to the total space of an orbifold bundle over the complex 1-dimensional orbifold $\mathcal{B} = \mathbb{H}^1_{\mathbb{C}}/D(2, n, n)$ with fibers given by projections to M_Γ of some complex geodesics in $\mathbb{H}^2_{\mathbb{C}}$. It follows from the local rigidity of each ρ, combined with [81, Lemma 4.5], that there exists an equivariant holomorphic map

$$\tilde{f} : \mathbb{H}^1_{\mathbb{C}} \to \mathbb{H}^2_{\mathbb{C}}.$$

(I owe this observation to Ludmil Katzarkov.)[4] Since the orbi-bundle $M = M_\Gamma \to \mathcal{B}$ has holomorphic fibers, it follows that $\tilde{f}$ descends to a holomorphic map $f : \mathcal{B} \to M$ which has only positive, zero-dimensional intersections with the fibers. Composing with the projection $M \to \mathcal{B}$, we obtain a self-map $h : \mathcal{B} \to \mathcal{B}$ which is a branched covering. Since $\mathcal{B}$ is a hyperbolic orbifold, it follows that $h = \mathrm{id}$. In other words, $M \to \mathcal{B}$ admits a holomorphic section. In particular, M (and any of its finite manifold-covering spaces, given by Selberg's Lemma) is non-Stein.

Example 2.12 (Polygon-Groups, J. Granier, [39]) The polygon-group $\Gamma_{6,3}$ (see Example C.17) embeds as a convex-cocompact subgroup in $PU(2, 1)$ via the reflection representation $\rho_{6,3}$. Thus, the limit set of $\Gamma_{6,3} < PU(2, 1)$ is homeomorphic to the Menger curve.

Conjecturally, the same holds for all polygon-groups $\Gamma_{n,3}$, $n \geq 6$, cf. [10, 29, 50] for a discussion of isometric actions on real hyperbolic spaces.

Example 2.13 Complex-hyperbolic manifolds which are singular fibrations with compact fibers.

Definition 2.13 A *singular Kodaira fibration* is a surjective holomorphic map with connected fibers $f : M \to B$ between connected complex manifolds/orbifolds, where $0 < \dim B < \dim M$. (Usually, it is required that no two generic fibers are biholomorphic to each other, but, in order to simplify the discussion, I will omit this condition.)

Singular Kodaira fibrations need not be locally trivial in the holomorphic or even topological sense; a *Kodaira fibration* is a holomorphic map $f : M \to B$ which is a smooth fiber bundle.

In the context of complex hyperbolic manifolds, the first example of a singular Kodaira fibration appeared in Ron Livne's PhD thesis, [63]. Many more examples are now known. Below we discuss one example which (to my knowledge) first appeared in the work of Mostow [67] with an explicit description of the quotient-orbifold given by Hirzebruch in [46].

[4] I refer the reader to the book [17] for a gentle introduction to Simpson's results, and for a discussion of variations of Hodge structures and period domains.

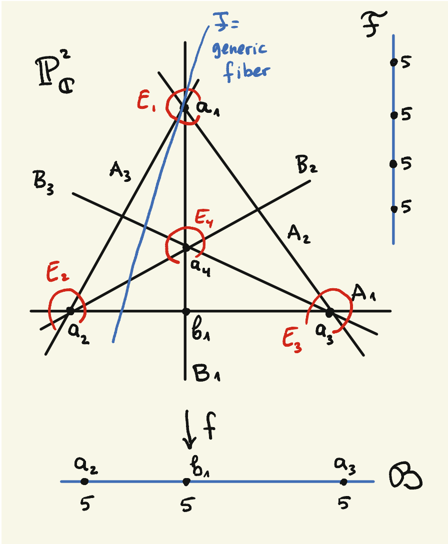


Fig. 2.2 Orbi-Kodaira fibration

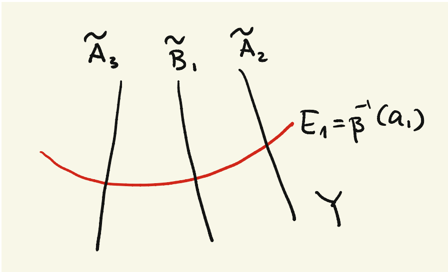


Fig. 2.3 Blow-up

Consider the *standard quadrangle* in $\mathbb{P}^2_{\mathbb{C}}$, which is a configuration A of six lines $A_1, A_2, A_3, B_1, B_2, B_3$ with four triple intersection points a_1, a_2, a_3, a_4 and three double intersection points b_1, b_2, b_3, see Fig. 2.2. Let Y denote the complex surface obtained via blow-up of the four triple intersection points of A; let $\beta : Y \to \mathbb{P}^2_{\mathbb{C}}$ denote the blow-down map. Then Y contains a configuration $\tilde{A}$ of eight distinguished smooth rational curves $C_1, ..., C_{10}$: The four exceptional divisors $E_1, ..., E_4$ of the blow-up and six lifts $\tilde{A}_i, \tilde{B}_i, i = 1, 2, 3$, of the original projective lines in the arrangement A. (See Fig. 2.3.) The configuration $\tilde{A}$ is a divisor D with simple normal crossings: Any two curves intersect in at most one point and at every intersection point only two curves intersect. Our next goal is to define a *complex orbifold* O with underlying space Y and the singular/orbifold locus Σ_O equal to the union of curves in $\tilde{A}$ (the preimage under β of the union of lines in A). The local complex orbifold-charts of O are defined as follows.

1. At every point $z \in O - \Sigma_O$ the local chart is given by the restriction of β to a suitable neighborhood of z.

2. At every point $z \in \Sigma_O$ which is not a (double) intersection point of the divisor D but $z \in C_i$, $i = 1, ..., 10$, the local chart is the holomorphic five-fold branched covering over a suitable neighborhood of z, ramified over C_i.
3. Suppose that z is an intersection point of D, $z \in C_i \cap C_j$, $i \neq j$. Choose local holomorphic coordinates at z where C_i, C_j correspond to the coordinate lines in $\mathbb{C}^2$ and z corresponds to the origin; $\mathbb{C}^2 = \mathbb{C} \times \mathbb{C}$. Each factor $\mathbb{C}$ in this product decomposition is biholomorphic to the quotient $\mathbb{C}/\mathbb{Z}_5$, with $\mathbb{Z}_5$ acting linearly on $\mathbb{C}$. Thus, a small neighborhood U of z in Y is biholomorphic to the quotient of the bi-disk, $\Delta^2/\mathbb{Z}_5^2$. This yields the local orbifold-chart at z, $\Delta^2 \to \Delta^2/\mathbb{Z}_5^2 \cong U$.

The result is a complex orbifold O with the underlying space Y. Hirzebruch then proves that the orbifold O is biholomorphic to the orbifold-quotient $M_\Gamma = \mathbf{B}^2/\Gamma$ of the complex 2-ball, by appealing to Yau's Uniformization Theorem, [90]: He verifies that the orbifold O admits a finite holomorphic orbifold-covering $M \to O$, where M is a complex surface of general type satisfying the equality of characteristic classes $3c_2 = c_1^2$; equivalently, $3\sigma(M) = \chi(M)$, where σ is the signature and χ is the Euler characteristic. Yau's theorem implies that M admits a Kähler metric of constant bisectional curvature -1, i.e. is a ball-quotient. Mostow Rigidity Theorem then implies that O is a complex hyperbolic orbifold as well. A bit more streamlined version of this argument was later developed by Barthel–Hirzebruch–Hofer [5], and Holzapfel, [47], who defined orbifold-characteristic classes directly computable from lines arrangement A in $\mathbb{P}^2_{\mathbb{C}}$ (as well as $\mathbb{P}^1_{\mathbb{C}} \times \mathbb{P}^1_{\mathbb{C}}$) and the orbifold-ramification numbers assigned to rational curves in the corresponding post-blow-up divisor.

I next describe a singular orbi-Kodaira fibration on O. Pick one of the triple intersection points, say, a_1, of the arrangement A and let A_1 be a line in A not passing through a_1. Consider the pencil of projective lines passing through a_1. This pencil defines a (nonsingular) holomorphic fibration of $\mathbb{P}^2_{\mathbb{C}} - \{a_1\}$ with the base A_1; the fibration map sends $z \in \mathbb{P}^2_{\mathbb{C}} - \{a_1\}$ to the point of intersection of the line za_1 with the line A_1. This fibration becomes a holomorphic map $f : Y \to \tilde{A}_1$ when we lift it to Y. Some fibers of f are, however, singular: These are the three singular fibers corresponding to the lifts of the three lines A_2, A_3, B_1 passing through a_1 and other points of triple intersection of A: a_2, a_3, a_4. The corresponding fibers are reducible rational curves (with the extra components corresponding to the exceptional divisors E_2, E_3, E_4). The line A_1 has an orbifold structure induced from O: The corresponding orbifold $\mathcal{B}$ has three singular points a_2, a_3, b_1, with local isotropy groups $\mathbb{Z}_5$ for each of them. The map f defined above respects the orbifold structure of O and $\mathcal{B}$ and, hence, we obtain a singular Kodaira orbi-fibration $f : O \to \mathcal{B}$. This fibration is nonsingular away from the preimages of the points a_2, b_1, a_3, with the generic fiber(s) $\mathcal{F}$ diffeomorphic to the orbifold with the underlying space $\mathbb{P}^1_{\mathbb{C}}$ and four singular points of order 5.

The restriction of f to $O' = f^{-1}(\{a_2, b_1, a_3\})$ is a nonsingular Kodaira fibration, i.e. a smooth (orbifold) fiber bundle; accordingly, $\pi_1(\mathcal{F})$ embeds as a normal subgroup in $\pi_1(O')$. Since the inclusion $O' \to O$ induces an epimorphism of fundamental groups $\pi_1(O') \to \pi_1(O) = \Gamma$, the image N of $\pi_1(\mathcal{F})$ in $\pi_1(O) = \Gamma$ is a normal finitely-generated subgroup $N \lhd \Gamma$. By passing to the universal covering of $\mathcal{B}$, we obtain a holomorphic map $h : \mathbb{H}^2_{\mathbb{C}}/N \to \mathbb{H}^1_{\mathbb{C}}$. The fibers of this map are compact and, generically, diffeomorphic to $\mathcal{F}$. The map h has infinitely many critical values in $\mathbb{H}^1_{\mathbb{C}}$ which break into finitely many $\pi_1(\mathcal{B})$-orbits and accumulate to the entire circle $\partial_\infty \mathbb{H}^1_{\mathbb{C}}$. Lifting h further to an N-invariant holomorphic function $\mathbb{H}^2_{\mathbb{C}} \to \mathbb{H}^1_{\mathbb{C}}$ and extending this function to a measurable N-invariant nonconstant function $S^3 = \partial_\infty \mathbb{H}^2_{\mathbb{C}} \to S^1$, we conclude that the action of N on S^3 is non-ergodic.

The group Γ in the above example is a special case of:

Example 2.14 (Arithmetic Lattices of Simplest Type) Let K be a totally real number field, i.e. a finite extension of $\mathbb{Q}$ such that the image of every embedding $K \to \mathbb{C}$ lies in $\mathbb{R}$. Take an imaginary quadratic extension L/K, i.e. an extension which does not embed in $\mathbb{R}$. Since K is totally-real and L is its imaginary extension, all embeddings $L \to \mathbb{C}$ come in complex conjugate pairs:

$$\sigma_1, \bar{\sigma}_1, \ldots, \sigma_k, \bar{\sigma}_k.$$

Next, take a Hermitian quadratic form in $n+1$ variables

$$\varphi(z, \bar{z}) = \sum_{p,q=1}^{n+1} a_{pq} z_p \bar{z}_q$$

with coefficients in L. I require the forms $\varphi^{\sigma_1}, \varphi^{\sigma_2}$ to have signature $(n, 1)$ and the forms $\varphi^{\sigma_j}, \varphi^{\bar{\sigma}_j}$ to be definite for the rest of the embeddings. I will identify L with $\sigma_1(L)$, so $\sigma_1 = \mathrm{id}$. Let $SU(\varphi)$ denote the group of special unitary automorphisms of the form φ on L^{n+1}. The embedding σ_1 defines a homomorphism $SU(\varphi) \to SU(n, 1)$ with relatively compact kernel.

A subgroup Γ of $SU(n, 1)$ is said to be an arithmetic lattice of the *simplest type* (or *of type I*) if it is commensurable[5] to $SU(\varphi, \mathcal{O}_L) = SU(\varphi) \cap SL(n+1, \mathcal{O}_L)$, where $\mathcal{O}_L$ is the ring of integers of L. For every such Γ the quotient $\mathbb{H}^n_{\mathbb{C}}/\Gamma$ has finite volume. I refer to [66] for more detail on arithmetic subgroups of $SU(n, 1)$.

It is known that every arithmetic lattice Γ of the simplest type contains a finite index congruence-subgroup Γ' with infinite abelianization, [56] (see also [88]). Equivalently, the quotient-space $\mathbf{B}^n/\Gamma'$ has positive 1st Betti number. In contrast, Rogawski, [76], proved that for *type II* arithmetic lattices in $SU(2, 1)$, every congruence-subgroup has finite abelianization. It is unknown if such a lattice contain finite index subgroups with infinite abelianization. Furthermore, certain classes of non-arithmetic lattices in $SU(2, 1)$ (the ones violating the integrality condition for arithmetic groups) are proven to have positive virtual first Betti number by the work of S.-K. Yeung, [91].

I now discuss the existence of (singular) Kodaira fibrations of *compact* complex hyperbolic manifolds $M = \mathbb{H}^n_{\mathbb{C}}/\Gamma$.

[5] I.e. the intersection of the two groups has finite index in both.

1. Suppose that $b_1(M) > 0$. Since M is Kähler, $b_1(M)$ is even, hence, there exists an epimorphism $\phi : \Gamma \to \mathbb{Z}^2$. If the kernel of ϕ is not finitely-generated, then, according to a theorem of Delzant, [27], the manifold M admits a singular Kodaira fibration over a 1-dimensional complex hyperbolic orbifold.
2. If M and B are both complex hyperbolic, then there are no (nonsingular) Kodaira fibrations $M \to B$: This was first proven in the case when M is a surface by Liu, [62], and then generalized to arbitrary dimensions by Koziarz and Mok, [58]. They also prove nonexistence of Kodaira fibrations $M \to B$ when $\dim(B) \geq 2$ and M merely has finite volume. Furthermore, if M is 2-dimensional, for every singular Kodaira fibration $M \to B$, the kernel of the homomorphism $\pi_1(M) = \Gamma \to \pi_1(B)$ is finitely generated but is not finitely-presentable [45, 48].

Problem 2.1 Is there a discrete subgroup $\Gamma < PU(2, 1)$ isomorphic to the fundamental group of a compact real hyperbolic surface, such that $M = M_\Gamma$ admits a Kodaira fibration (with compact fibers) $M \to \mathbb{H}^1_{\mathbb{C}}$? Is there a singular Kodaira fibration (with compact fibers) $\mathbb{H}^2_{\mathbb{C}}/\Gamma \to \mathbb{H}^1_{\mathbb{C}}$ which has only finitely many singular fibers?

2.9 Complex Hyperbolic Kleinian Groups and Function Theory on Complex Hyperbolic Manifolds

In this section we discuss some interesting interactions between the general theory of holomorphic functions on complex manifolds (which I review in Sect. 2.10) and geometry/topology of complex Kleinian groups.

Proposition 2.5 *If $\Gamma < PU(n, 1)$ is a discrete, torsion-free subgroup such that $M = M_\Gamma$ admits a surjective holomorphic map with compact fibers $f : M \to B$, where B is a complex manifold satisfying $\dim(B) < n$, then $\Omega_\Gamma = \emptyset$. In particular, M cannot have convex ends.*

Proof Suppose, to the contrary, that $\Omega_\Gamma \neq \emptyset$. Then $Core_\eta(M)$ is a proper submanifold (with boundary) in M. Since $\mathbb{H}^n_{\mathbb{C}}$ is strictly negatively curved, the nearest-point projection $\Pi : \mathbb{H}^n_{\mathbb{C}} \to \mathrm{hull}_\eta(\Lambda_\Gamma)$ is strictly contracting away from $\mathrm{hull}_\eta \Lambda_\Gamma$. By the Γ-equivariance, Π descends to a strictly contracting projection $\pi : M \to Core_\eta(M)$. Therefore, if Y is a compact complex k-dimensional subvariety in M of positive dimension not contained in $Core_\eta(M)$ then $\pi(Y)$ has k-volume strictly smaller than that of Y. This is a contradiction since $\pi : Y \to \pi(Y)$ is homotopic to the identity inclusion map $\mathrm{id}_Y : Y \to M$ and compact complex subvarieties in Kähler manifolds are volume-minimizers in their homology classes. Taking a generic fiber Y of $f : M \to B$ through a point $x \in M - \mathrm{hull}_\eta(\Lambda_\Gamma)$ concludes the proof.

I next discuss the geometry and topology of quotient-orbifolds M_Γ, primarily for convex-cocompact subgroups $\Gamma < PU(n, 1)$.

A classical example of a complex submanifold with strictly Levi-convex boundary is a closed unit ball $\overline{\mathbf{B}^n}$ in $\mathbb{C}^n$. Suppose that $\Gamma < Aut(\mathbf{B}^n)$ is a discrete torsion-free subgroup of the group of holomorphic automorphisms of $\mathbf{B}^n$ with (nonempty) domain of discontinuity $\Omega \subset \partial\mathbf{B}^n$. The quotient

$$\overline{M}_\Gamma = (\mathbf{B}^n \cup \Omega)/\Gamma$$

is a smooth submanifold with strictly Levi-convex boundary in the complex manifold $\check{\Omega}_\Gamma/\Gamma$ (see (2.1)). Thus, we conclude:

Lemma 2.2 *If* $\overline{M}_\Gamma = (\mathbf{B}^n \cup \Omega)/\Gamma$ *has compact boundary, then* M *is strongly pseudoconvex.*

Consequently:

Theorem 2.14 *Let* $\Gamma < PU(n,1)$, $n \geq 2$, *be a convex-cocompact discrete subgroup. Then* ∂M_Γ *is connected.*

Proof Since Γ is convex-cocompact, it is also finitely generated. Hence, by Selberg's Lemma, the orbifold M_Γ is very good. Therefore, it suffices to consider the case when Γ is torsion-free, i.e. M_Γ is a complex n-manifold. Since $\overline{M}_\Gamma$ is strongly pseudoconvex, connectedness of its boundary is an immediate consequence of Theorem F.24.

Theorem 2.15 *Let* $\Gamma < PU(n,1)$, $n \geq 2$, *be a convex-cocompact discrete subgroup which is not a lattice, i.e.* $\Omega_\Gamma \neq \emptyset$. *Then* $\dim(\Lambda_\Gamma) \leq 2n-3$, *equivalently,* $\mathrm{cd}_{\mathbb{Q}}(\Gamma) \leq 2n-2$.

Proof As before, it suffices to consider the case of torsion-free groups Γ. According to Corollary F.6, M_Γ is homotopy-equivalent to a CW complex of dimension $\leq 2n-2$. It follows that $\mathrm{cd}_{\mathbb{Q}}(\Gamma) \leq 2n-2$ and, by the Bestvina-Mess theorem, $\dim(\partial_\infty\Gamma) \leq 2n-3$. Since $\partial_\infty\Gamma$ is homeomorphic to Λ_Γ, $\dim(\Lambda_\Gamma) \leq 2n-3$ as well.

In particular, Λ_Γ does not separate S^{2n-1} (even locally) and, hence, Ω_Γ is connected, which gives another proof of the fact that ∂M_Γ is connected.

Specializing to the case $n = 2$, we obtain: If $\Gamma < PU(2,1)$ (for simplicity, torsion-free) is convex-cocompact and is not a lattice, then Λ_Γ is at most 1-dimensional. In particular, according to [53], Γ admits an iterated amalgam decomposition over trivial and cyclic subgroups, so that the terminal groups are either cyclic, or isomorphic to Fuchsian groups (and the limit set is a topological circle) or groups whose limit sets are Sierpinski carpets or Menger curves.

Theorem 2.16 *Suppose that* Γ *is torsion-free convex cocompact,* $n > 1$ *and* M_Γ *contains no compact complex subvarieties of positive dimension. Then* M_Γ *is Stein.*

Proof This is an immediate consequence of Theorem F.25.

One way to prove that M_Γ contains no compact complex subvarieties of positive dimension is to argue that $\Gamma = \pi_1(M)$ is free: This implies that $H_i(M_\Gamma) = 0, i \geq 2$, but, since M_Γ is Kähler, every compact complex k-dimensional subvariety of

M_Γ would define a nonzero $2k$-dimensional homology class. For instance, if Γ is convex-cocompact, $\delta_\Gamma < 1$ then $\dim \Lambda_\Gamma \le \dim_H(\Lambda_\Gamma) < 1$, which implies that $\dim \Lambda_\Gamma = 0$ and, hence, Γ is a virtually free group. However, even when $H_2(M) \neq 0$, one can still, sometimes, prove that M_Γ contains no compact complex curves. For instance, let $L \to M_\Gamma$ be the canonical line bundle. If $C \subset M_\Gamma$ is an (even singular) complex curve, the pull-back of L to C has nonzero 1st Chern class. Assuming that $H_2(M_\Gamma) \cong \mathbb{Z}$ (e.g. if Γ is isomorphic to the fundamental group of a compact Riemann surface), if the 1st Chern class of L evaluated on the generator of $H_2(M_\Gamma)$ is zero, then M_Γ contains no complex curves. This argument applies in the case of real-Fuchsian groups and their quasi-Fuchsian deformations.

Observe that if $\Gamma < PU(2,1)$ is a complex Fuchsian group, then $\dim_H(\Lambda_\Gamma) = 2$.

Theorem 2.17 (S. Dey and M. Kapovich [30]) *If $\Gamma < PU(n,1)$ is discrete, torsion-free and M_Γ contains a compact complex subvariety of positive dimension, then $\delta_\Gamma \ge 2$.*

Corollary 2.5 *Suppose that $\Gamma < PU(n,1)$ is torsion-free, convex-cocompact and $\delta_\Gamma < 2$, then M_Γ is Stein.*

Burns' Theorem I now drop the convex-cocompactness assumption and consider general discrete, torsion-free subgroups $\Gamma < PU(n,1)$. Theorem 2.14 has the following striking generalization. It was first stated by Dan Burns, who, as it appears, never published a proof; a published proof is due to Napier and Ramachandran, [69, Theorem 4.2]:

Theorem 2.18 *Suppose that $n \ge 3$, $\Gamma < PU(n,1)$ is discrete, torsion-free and ∂M_Γ has at least one compact component S. Then:*

1. $\partial M_\Gamma = S$.
2. Γ *is geometrically finite.*

A good example illustrating this theorem is that of a Schottky-type group (Example 2.6), where the limit set is totally disconnected, the quotient manifold Ω_Γ/Γ is compact and M_Γ has k cusps. In particular, $\overline{M_\Gamma}$ is noncompact in this example.

It is unknown if Burns' theorem holds for $n = 2$, but Mohan Ramachandran proved the following:

Theorem 2.19 *Suppose that $\Gamma < PU(2,1)$ is discrete, torsion-free, the injectivity radius of M_Γ is bounded away from zero, and ∂M_Γ has at least one compact component. Then Γ is convex-cocompact.*

The proof of this theorem is given in Appendix G.

2.10 Conjectures and Questions

In this section I collect some conjectures and questions in addition to those scattered throughout the survey.

The first conjecture is a generalization of Burns' theorem, Theorem 2.18:

Conjecture 2.1 Suppose that $\Gamma < PU(n,1), n \geq 2$, is such that for $M = M_\Gamma$ the thick part $M_{[\epsilon,\infty)}$ has a convex end. Then Γ is geometrically finite and Ω_Γ is connected.

The next two conjectures are motivated by Theorem 2.17:

Conjecture 2.2 If $\Gamma < PU(n,1)$ is discrete, torsion-free, $\delta_\Gamma = 2$ and M_Γ contains a compact complex subvariety of positive dimension, then Γ is a complex Fuchsian group.

Conjecture 2.3 If $\Gamma < PU(n,1)$ is discrete, torsion-free and $\delta_\Gamma < 2k$, then M_Γ cannot contain a compact complex subvariety of dimension k.

Conjecture 2.4 (Chengbo Yue's Gap Conjecture, [92]) Suppose that $\Gamma < G = Aut(\mathbf{B}^n)$ is a convex-cocompact torsion-free subgroup. Then either Γ is a uniform lattice in G (and, thus, $\delta_\Gamma = 2n$) or $\delta_\Gamma \leq 2n-1$.

Note that the two other conjectures about nonelementary convex-cocompact subgroups $\Gamma < PU(n,1)$ made in the introduction to [92] fail already in dimension $n = 2$:

(a) The inequality $\dim_H \Lambda_\Gamma > n-1$ does not imply that M_Γ is non-Stein. For instance, a real-hyperbolic quasifuchsian subgroup of $PU(2,1)$ serves as an example.
(b) Even if M_Γ is non-Stein, a compact complex curve in M_Γ need not be a finite union of totally geodesic complex curves, as it is shown by the AGG-examples.

Problem 2.2

1. Investigate which polygon-groups embed discretely in $PU(2,1)$.
2. Is there a convex-cocompact subgroup of $PU(2,1)$ with the limit set homeomorphic to Sierpinski carpet?

While "most" compact 3-dimensional manifolds are hyperbolic, very few examples of hyperbolic 3-manifolds which are of the form Ω_Γ/Γ, $\Gamma < PU(2,1)$ are known, see the book by Richard Schwartz [79] for further discussion.

Conjecture 2.5 The Menger curve limit set in Example 2.12 is "unknotted" in S^3, i.e. is ambient-isotopic to the standard Menger curve $\mathcal{M} \subset \mathbb{R}^3 \subset S^3 = \mathbb{R}^3 \cup \{\infty\}$. Furthermore, in this example, the quotient 3-dimensional manifold Ω_Γ/Γ is hyperbolic.[6]

[6] It suffices to show that Ω_Γ/Γ contains no incompressible tori, which is closely related to the unknottedness problem of the Menger-curve limit set.

Problem 2.3 Prove the existence of discrete geometrically infinite subgroups of $PU(2, 1)$ which are isomorphic to fundamental groups of compact surfaces.[7]

Note that such subgroups do not exist in $PU(1, 1)$ but abound in $PO(3, 1)$. Furthermore, the only known examples of finitely generated geometrically infinite subgroups of $PU(2, 1)$ come from singular Kodaira fibrations and are not finitely-presentable, see Example 2.13.

The conjectures and questions appearing above, deal with discrete subgroups Γ of $PU(n, 1)$ which are not lattices, i.e. the $\mathbb{H}^n_{\mathbb{C}}/\Gamma$ has infinite volume. Below, I discuss two problems regarding lattices.

Arithmeticity The most famous open problem regarding lattices in $PU(n, 1)$ deals with the existence problem of nonarithmetic subgroups and was first raised in Margulis' ICM address [64]. It is known (due to the work of Margulis [65], Corlette [21], Gromov–Schoen [42], and Gromov–Piatetski-Shapiro [41]) that:

(a) For each n, the Lie group $SO(n, 1)$ contains non-arithmetic lattices.
(b) For every simple noncompact connected linear Lie group G which is not locally isomorphic to $SO(n, 1)$ and $SU(n, 1)$, every lattice $\Gamma < G$ is arithmetic.

This leaves out the series of Lie groups $PU(n, 1)$, $n \geq 2$. Currently, primarily due to the work of Deligne and Mostow, see [26], there are known examples of nonarithmetic lattices in $PU(2, 1)$ and $PU(3, 1)$. Loosely speaking there are three approaches to constructing nonarithmetic lattices:

(a) As monodromy groups of some linear holomorphic ODEs, see [23, 26], as well as [83] for a geometric interpretation.
(b) By constructing the corresponding complex hyperbolic orbifolds M_Γ whose underlying space is a blown-up $\mathbb{P}^n$, see [5, 23, 28, 80].
(c) By constructing a Dirichlet fundamental domain of Γ in $\mathbb{H}^2_{\mathbb{C}}$, see [31, 67].

But using these techniques becomes increasingly difficult (or even impossible) as the dimension n increases, which means that different approaches are needed.

Conjecture 2.6 For each n, $PU(n, 1)$ contains a nonarithmetic lattice.

By analogy with the construction of non-arithmetic lattices in [41], one can hope for a similar "hybrid" construction of nonarithmtic lattices in $PU(n, 1)$, leading to a conjecture due to Bruce Hunt:

Conjecture 2.7 For every $n \geq 2$, there exists a quadruple of arithmetic lattices $\Gamma_1, \Gamma_2 < SU(n-1, 1)$ and $\Gamma_3 < SU(n-2, 1)$ such that:

(1) Γ_3 is isomorphic to subgroups in Γ_1, Γ_2; hence, we obtain an amalgam $\Gamma_0 = \Gamma_1 \star_{\Gamma_3} \Gamma_2$.
(2) There exists an epimorphism $\rho : \Gamma_0 \to \Gamma < SU(n, 1)$ injective on Γ_1, Γ_2, whose image is a nonarithmetic lattice $\Gamma < SU(n, 1)$.

[7] Cf. section 11.4 in [51].

Unlike [41], where nonarithmetic lattices in $SO(n,1)$ were constructed via a similar process, with an *isomorphism* $\Gamma_0 \to \Gamma < SO(n,1)$, in the complex hyperbolic setting there is no hope for an injective homomorphism ρ (a lattice in $SU(n,1)$ cannot be isomorphic to an amalgam Γ_0 as above).

Nonexistence of Reflection Lattices The known examples of nonarithmetic lattices Γ in $PU(n,1), n = 2, 3$, are all commensurable to *complex reflection subgroups*, i.e. discrete subgroups of $PU(n,1)$ generated by complex reflections. Furthermore, up to commensuration, the underlying spaces of their quotient orbifolds $M_\Gamma = \mathbb{H}^n_{\mathbb{C}}/\Gamma$ are *rational projective varieties*.

Conjecture 2.8 There exists N such that for all $n \geq N$ the following holds:

1. If $\Gamma < PU(n,1)$ is a lattice then Γ cannot be a reflection subgroup.
2. If $\Gamma < PU(n,1)$ is a lattice then the underlying space of the orbifold M_Γ cannot be a rational algebraic variety. More ambitiously, it has to be a variety of general type.

The motivation for this conjecture comes from theorems due to Vinberg [87], and Prokhorov [73], establishing nonexistence of reflection lattices in $PO(n,1)$, when n is sufficiently large.

Appendix A: Horofunction Compactification

A metric space (Y,d) is called *geodesic* if any two points x, y in X are connected by a geodesic segment, denoted xy. (This notation is a bit ambiguous since in many cases such a segment is non-unique.) A *geodesic triangle*, denoted xyz, in a metric space (X,d) is a set of three geodesic segments xy, yz, zx connecting cyclically the points x, y, z, the *vertices* of the triangle; the segments xy, yz, zx are the *edges* of the triangle. Thus, geodesic triangles are 1-dimensional objects.

Let (Y,d) be a locally compact geodesic metric space. For each $y \in Y$ define the 1-Lipschitz function $d_y = d(y,\cdot)$ on Y. This leads to the *Kuratowski embedding* $\kappa : Y \to C(Y) = C(Y,\mathbb{R})$, $y \mapsto d_y$. Let $\mathbb{R} \subset C(Y)$ denote the linear subspace of constant functions. Composing the embedding κ with the projection $C(Y) \to C(Y)/\mathbb{R}$ (where $\mathbb{R}$ acts additively on $C(Y)$) we obtain the *Kuratowski embedding* of Y,

$$Y \hookrightarrow C(Y)/\mathbb{R}.$$

Then $\overline{Y}$, the closure of Y in $C(Y)/\mathbb{R}$, is the *horofunction compactification* of Y. The *horoboundary* of Y is $\overline{Y} \setminus Y$.

Functions representing points in $\partial_\infty Y = \overline{Y} - Y$ are the *horofunctions* on Y. In other words, horofunctions on Y are limits (uniform on compacts in Y) of sequences of normalized distance functions $d_{y_i} - d_{y_i}(o)$, where $y_i \in Y$ are divergent sequences in Y. Each geodesic ray $r(t)$ in Y determines a horofunction in Y called a *Busemann* *function* b_r, which is the subsequential limit

$$\lim_{i\to\infty} d_{r(i)} - d_{r(i)}(o).$$

If Y is a Hadamard manifold, then each limit as above exists (without passing to a subsequence). Furthermore, each horofunction is a Busemann function. This yields a topological identification of the visual compactification of Y and its horofunction compactification. Level sets of Busemann functions are called *horospheres* in X. The point $r(\infty) \in \partial_\infty Y$ is the *center* of the horosphere $\{b_r = c\}$. Sublevel sets $\{b_r < c\}$ are called *horoballs*. The point $r(\infty)$ represented by the ray r is the *center* of the corresponding horospheres/horoballs.

Appendix B: Two Classical Peano Continua

A *Peano continuum* is a compact connected and locally path-connected metrizable topological space. I will need two examples of 1-dimensional Peano continua. Both are obtained via a procedure similar to the construction of the "ternary" Cantor set.

Sierpinski Carpet Let $I = [0, 1]$ denote the unit interval. Start with the unit square $Q_0 = I^2 \subset \mathbb{R}^2$. Divide I in three congruent subintervals and, accordingly, divide I^2 in 9 congruent subsquares. Remove the interior of the "middle subsquare", the one disjoint from the boundary of Q. Call the result Q_1. Now, repeat this procedure for each of the remaining 8 subsquares in Q_1, to obtain a planar region Q_2, etc. The *standard Sierpinski carpet* in $\mathbb{R}^2$ is the intersection

$$\mathcal{S} := \bigcap_{i=0}^{\infty} Q_i.$$

Menger Curve Consider the unit cube $C = I^3 \subset \mathbb{R}^3$. Let $\pi_i, i = 1, 2, 3$ denote the orthogonal projections of $\mathbb{R}^3$ to the coordinate hyperplanes $P_i, i = 1, 2, 3$, in $\mathbb{R}^3$. In all three planes we take the Sierpinski carpets $\mathcal{S}_i \subset P_i$, constructed from the unit squares $Q_i = C \cap P_i, i = 1, 2, 3$. Then the *standard Menger curve* in $\mathbb{R}^3$ is defined as

$$\mathcal{M} := \bigcap_{i=1}^{3} \pi_i^{-1}(\mathcal{S}_i).$$

Alternatively, $\mathcal{M}$ can be described as follows. First, divide $C = C_0$ in 27 congruent subcubes with the edge-length $1/3$ and remove from C the "middle " open cube" Q_1 as well as the 8 open subcubes which share with Q_1 2-dimensional faces; remove those open faces as well. Continue inductively constructing a sequence of nested compacts $C_0 \supset C_1 \supset C_2 \supset ..$ and. Lastly,

$$\mathcal{M} = \bigcap_{i=0}^{\infty} C_i.$$

Appendix C: Gromov-Hyperbolic Spaces and Groups

A geodesic metric space (X, d) is called δ*-hyperbolic* if every geodesic triangle xyz in X is δ-slim, i.e. every edge of xyz is contained in the closed δ-neighborhood of the union of the other two edges. A geodesic metric space is called *Gromov-hyperbolic* if it is δ-hyperbolic for some $\delta < \infty$.

Examples of Gromov-hyperbolic spaces are *strictly negatively curved Hadamard manifolds*: If X is a Hadamard manifold of sectional curvature ≤ -1 then X is δ_0-hyperbolic with $\delta_0 = \text{arccosh}(\sqrt{2})$.

Let Γ be a group with finite generating set S. Given S, one defines the *Cayley graph* $C_{\Gamma,S}$. This graph is connected and Γ acts on it with finite quotient (the quotient graph has a single vertex and card(S) edges). The graph $C_{\Gamma,S}$ has a graph-metric, where every edge has unit length.

Definition C.14 A finitely generated group Γ is called *Gromov-hyperbolic* or simply *hyperbolic* if one (equivalently, every) Cayley graph of Γ is a Gromov-hyperbolic metric space.

The *Gromov boundary* $\partial_\infty \Gamma$ of Γ is the horoboundary of one (any) Cayley graph of Γ: Gromov boundaries corresponding to different Cayley graphs are equivariantly homeomorphic.

Examples of hyperbolic groups are given by:

Example C.15 Let X be a strictly negatively curved Hadamard manifold, $Y \subset X$ is a closed convex subset and $\Gamma < \text{Isom}(X)$ acts properly discontinuously and cocompactly on Y. Then Γ is hyperbolic and the ideal boundary $\partial_\infty Y$ of Y is equivariantly homeomorphic to the Gromov boundary of Γ.

In particular, every convex-cocompact discrete subgroup $\Gamma < \text{Isom}(X)$ is Gromov-hyperbolic and $\partial_\infty \Gamma$ is equivariantly homeomorphic to the limit set of Γ.

Cohomological dimension (with respect to the Chech cohomology) of the Gromov boundary of a hyperbolic group is closely related to the rational cohomological dimension of Γ itself:

Theorem C.20 (Bestvina–Mess [7]) $\dim(\partial_\infty \Gamma) = cd_{\mathbb{Q}}(\Gamma) - 1$.

In particular, by the Stallings–Swan–Dunwoody Theorem, Γ is *virtually free* (i.e. contains a free subgroup of finite index) if and only if $\partial_\infty \Gamma$ is zero-dimensional, if and only if $\partial_\infty \Gamma$ is totally disconnected, equivalently, it is either empty, or consists of two points or is homeomorphic to the Cantor set.

One classifies 1-dimensional boundaries of hyperbolic groups as follows:

Theorem C.21 (Kapovich–Kleiner [53]) *Suppose that* Γ *is a hyperbolic group with connected 1-dimensional Gromov boundary. Then either* $\partial_\infty \Gamma$ *is homeomorphic to* S^1*, or* Γ *splits as a finite graph of groups with virtually cyclic edge groups,*[1] *or* $\partial_\infty \Gamma$ *is homeomorphic to the Sierpinski carpet or the Menger curve.*

Example C.16 Hyperbolic von Dyck groups $D(p, q, r)$,

$$D(p, q, r) = \langle a, b, c | a^p = b^q = c^r = 1, abc = 1 \rangle, \; p^{-1} + q^{-1} + r^{-1} < 1.$$

These are hyperbolic groups with Gromov boundary homeomorphic to S^1. Moreover, each $D(p, q, r)$ admits a unique (up to conjugation in $\mathrm{Isom}(\mathbb{H}^2)$) isometric conformal action on the hyperbolic plane.

Representations of von Dyck Groups to $PU(2, 1)$ Given an element $g \in G = PU(2, 1)$ we let $\zeta(g)$ denote the codimension in G of the centralizer of g in G. In other words, $\zeta(g)$ is the local dimension near g of the subvariety of elements of G conjugate to g. Thus, $\zeta(g) \geq 2$ for every $g \in G$. Furthermore, if g is an involution then $\zeta(g) = 4$. The paper [89] by Andre Weil describes the local geometry of the *character variety*

$$Hom(D(p, q, r), G)//G$$

as follows:

Suppose that $\rho : D(p, q, r) \to G$ is a *generic* representation, i.e. one whose image has trivial centralizer in G. For instance, any representation whose image is discrete, nonelementary, not stabilizing a complex geodesic, will satisfy this condition. Then, near $[\rho]$, the real-algebraic variety $Hom(D(p, q, r), G)//G$ is smooth of dimension

$$\zeta(\rho(a)) + \zeta(\rho(b)) + \zeta(\rho(c)) - 2\dim(G) = \zeta(\rho(a)) + \zeta(\rho(b)) + \zeta(\rho(c)) - 16.$$

Assuming that $p = 2$, $\zeta(\rho(a)) = 4$, which implies that

$$\zeta(\rho(a)) + \zeta(\rho(b)) + \zeta(\rho(c)) - 16 \leq 4 + 12 - 16 = 0.$$

Combined with an easy analysis of non-generic representations, one obtains:

Proposition C.6 *If* $p = 2$ *then* $Hom(D(p, q, r), G)//G$ *is zero-dimensional.*

Example C.17 (Polygon-Groups) Fix two natural numbers $p \geq 5$ and $q \geq 3$. Define the *polygon-group* $\Gamma_{p,q}$ via the presentation

$$\langle a_1, ..., a_p | a_i^q = 1, [a_i, a_{i+1}] = 1, i = 1, ..., p \rangle,$$

[1] Hence, its Gromov boundary can be inductively described using boundaries of vertex groups.

where i is taken mod p. Each $\Gamma_{p,q}$ is hyperbolic with $\partial_\infty \Gamma_{p,q}$ homeomorphic to the Menger curve.

Every $\Gamma_{p,q}$ admits a canonical *reflection representation* $\rho_{p,q}$ to $PU(2,1)$ constructed as follows:

Pick a real hyperbolic plane $\mathbb{H}^2_{\mathbb{R}} \subset \mathbb{H}^2_{\mathbb{C}}$ and a regular right-angled p-gon $P = z_1 ... z_p$ in $\mathbb{H}^2_{\mathbb{R}}$. Let C_i denote the complex geodesic through the edge $z_i z_{i+1}$ of P (i is taken mod p). For each i let g_i be the order q complex reflection with the fixing C_i, with the rotation in the hyperplanes normal to C_i through the angle $2\pi/q$. Then $[g_i, g_{i+1}] = 1$ and, hence, we obtain a representation

$$\rho_{p,q} : \Gamma_{p,q} \to PU(2,1).$$

Appendix D: Orbifolds

The notion of *orbifold* is a generalization of the notion of a *manifold* which appears naturally in the context of properly discontinuous non-free actions of groups on manifolds. Orbifolds were first invented by Satake [78] in the 1950s under the name of V-manifolds, they were reinvented under the name of orbifolds by Thurston in the 1970s (see [82]) as a technical tool for proving his Hyperbolization Theorem. I refer the reader to [8] for a detailed treatment of orbifolds.

Before giving a formal definition we start with basic examples of orbifolds. Suppose that M is a smooth connected manifold and G a discrete group acting smoothly, faithfully[2] and properly discontinuously on M. Then the quotient $O = M/G$ is an *orbifold*, such orbifolds are called *good*. The quotient M/G, considered as a topological space X_O, is the *underlying space* of this orbifold. If S is a set of points in M where the action of G is not free, then its projection $\Sigma = S/G$ is the *singular locus* of the orbifold O.

To be more concrete, consider 2-dimensional orbifolds. Suppose that $M = \mathbb{H}^2_{\mathbb{R}}$ and G is a discrete subgroup of $PSL(2,\mathbb{R})$. Then the quotient $O = \mathbb{H}^2/G$ is a Riemann surface X_O with a discrete collection of *cone points* z_j which form the singular locus Σ of the orbifold O. The projection $p : \mathbb{H}^2 \to O$ is the *universal cover* of the orbifold O. The Riemann surface X_O has a natural hyperbolic metric which is singular in the discrete set Σ. Metrically, the points z_j are characterized by the property that the total angles around these points are $2\pi/n_j$. The numbers n_j are the orders of cyclic subgroups G_{z_j} of G which stabilize the points in $p^{-1}(z_j)$, they are called *the local isotropy groups*. The projection p is a *ramified covering* from the point of view of Riemann surfaces. From the point of view of orbifolds this is an (orbi) *covering*. Thus, the singular locus of the orbifold O consists of the points z_j in Σ equipped with the extra data: The $PSL(2,\mathbb{R})$-conjugacy classes of the local

[2] I.e. each nontrivial element of G acts nontrivially.

isotropy groups G_{z_j} (of course, each local isotropy group G_{z_j} is determined by the number n_j).

I now discuss the general definition. A (smooth) n-dimensional orbifold O is a pair: A Hausdorff paracompact topological space X (which is called the *underlying space* of O and is denoted X_O) and an *orbifold-atlas* $\mathcal{A}$ on X. The atlas $\mathcal{A}$ consists of:

- A collection of open sets $U_i \subset X$,which is closed under taking finite intersections, such that $X = \bigcup_i U_i$.
- A collection of open sets $\tilde{U}_i \subset \mathbb{R}^n$.
- A collection of finite groups of diffeomorphisms Γ_j acting on $\tilde{U}_i$ so that each nontrivial element of Γ_j acts nontrivially on each component of $\tilde{U}_j$.
- A collection of homeomorphisms

$$\phi_i : U_i \to \tilde{U}_i / \Gamma_i.$$

The atlas $\mathcal{A}$ is required to behave well under inclusions. Namely, if $U_i \subset U_j$, then there is a smooth embedding

$$\tilde{\phi}_{ij} : \tilde{U}_i \to \tilde{U}_j$$

and a monomorphism $f_{ij} : \Gamma_i \to \Gamma_j$ such that $\tilde{\phi}_{ij}$ is f_{ij}-equivariant.

The open sets U_j are the *coordinate* neighborhoods of the points $x \in U_j$ and $\tilde{U}_j$ are their *covering coordinate* neighborhoods.

Similarly to orbifolds, one defines the class of *orbifolds with boundary*; just instead of *open* sets $\tilde{U}_j \subset \mathbb{R}^n$ we use open subsets in

$$\mathbb{R}^n_+ \cup \mathbb{R}^{n-1} = \{(x_1, ..., x_n) : x_n \geq 0\}.$$

The *boundary* of such an orbifold consists of points $x \in X_O$ which correspond to $\mathbb{R}^{n-1}$ under the identification $U_i \cong \tilde{U}_i / \Gamma_i$. As in the case of manifolds, the boundary of each orbifold is an orbifold without boundary. By abusing notation we will call an *orbifold with boundary* simply an *orbifold*. A compact orbifold without boundary is called *closed*.

To each point $x \in X$ we associate a germ of action of a finite group of diffeomorphisms Γ_x at a fixed point $\tilde{x}$. If $\phi_j(x)$ is covered by a point $\tilde{x}_j \in \tilde{U}_j$, then we have the isotropy group $\Gamma_{j,x}$ of $\tilde{x}_j$ in Γ_j. Note that if $U_i \subset U_j$, then the map $\tilde{\phi}_{ij} : \tilde{U}_i \to \tilde{U}_j$ induces an isomorphism from the germ of the action of $\Gamma_{j,x}$ at $\tilde{x}_j$ to the germ of the action of $\Gamma_{i,x}$ at $\tilde{x}_i$. Thus we let the germ $(\Gamma_x, \tilde{x})$ be the equivariant diffeomorphism class of the germ $(\Gamma_{j,x}, \tilde{x}_j)$. The group Γ_x is called the *local isotropy group* of O at x. The set of points x with nontrivial local isotropy group is called *the singular locus* of O and is denoted by Σ_O. Note that the singular locus is nowhere dense in X_O. An orbifold with empty singular locus is called *nonsingular* or a *manifold*.

The main source of examples of orbifolds is:

Example D.18 Let M a smooth connected n-manifold and Γ is a discrete group acting smoothly and faithfully on M. Then $X = M/\Gamma$ has a natural orbifold structure. The atlas A on X is given as follows: Each $y \in M$ admits a coordinate neighborhood $\tilde{U}$ (identified with an open subset of $\mathbb{R}^n$) such that for every $g \in \Gamma$ either $g\tilde{U} \cap \tilde{U} = \emptyset$ or $g \in \Gamma_y$ (the stabilizer of y in Γ) and $g(\tilde{U}) = \tilde{U}$. Then let $\phi : \tilde{U} \to U = \phi(\tilde{U})$ be the quotient map. One verifies that A indeed satisfies axioms of an orbifold-atlas. The groups G_j in the definition of an atlas are just the stabilizers Γ_y as above.

Since Γ_x acts smoothly near the fixed point $\tilde{x}$, the germ $(\Gamma_x, \tilde{x})$ is linearizable: Equip a neighborhood of $\tilde{x}$ with a Γ_x-invariant Riemannian metric. Then the exponential map (with the origin at $\tilde{x}$) conjugates the orthogonal action of Γ_x on the tangent space $T_{\tilde{x}}\mathbb{R}^n$ to the germ of the action of Γ_x at $\tilde{x}$.

Definition D.15 A *Riemannian metric* ρ on an orbifold O is a usual Riemannian metric on $X_O - \Sigma_O$, such that after we lift ρ to the local covering coordinate neighborhoods $\tilde{U}_i$, it extends to a Γ_i-invariant Riemannian metric on the whole $\tilde{U}_i$.

The same definition applies to *complex structures*.

Remark D.12 Each orbifold O admits a Riemannian metric: The proof is the same as for smooth manifolds, using a partition of unity.

A *homeomorphism* (resp. *diffeomorphism*) between orbifolds O, O' is a homeomorphism $h : X_O \to X_{O'}$ such that for all points $x \in O$, $y = h(x) \in O'$, there are coordinate neighborhoods $U_x \cong \tilde{U}_x/\Gamma_x$, $V_y \cong \tilde{V}_y/\Gamma_y$ such that h lifts to an equivariant homeomorphism (resp. diffeomorphism)

$$\tilde{h}_{xy} : \tilde{U}_x \to \tilde{V}_y.$$

Note that to describe a smooth orbifold O up to homeomorphism it suffices to describe the topology of the pair (X_O, Σ_O) and the homeomorphic equivalence classes of the germs $(\Gamma_x, \tilde{x})$ for the points $x \in \Sigma_O$.

Remark D.13 Let O be a connected compact 1-dimensional orbifold without boundary which is not a manifold. Then O is homeomorphic to the closed interval $[a, b]$ where $(\Gamma_a, \tilde{a})$, $(\Gamma_b, \tilde{b})$ are the germs $(\mathbb{Z}_2, 0)$ of the reflection group $\mathbb{Z}_2$ acting isometrically on $\mathbb{R}$ near its fixed point $0 \in \mathbb{R}$.

A *smooth* map between orbifolds O and O' is a continuous map

$$g : O \to O'$$

which can be (locally) lifted to smooth equivariant maps between pairs of coordinate covering neighborhoods

$$\tilde{g}_{ij} : \tilde{U}_j \to \tilde{V}_i$$

Similarly we define *immersions* and *submersions* between orbifolds as smooth maps between orbifolds which locally lift to immersions and submersions respectively.

Suppose that $\mathcal{O}', \mathcal{O}$ are orbifolds and $p : X_{\mathcal{O}'} \to X_{\mathcal{O}}$ is a continuous map. The map p is called a *covering map* or *orbi-covering* between the orbifolds $\mathcal{O}', \mathcal{O}$ if the following property is satisfied:

For each point $x \in X_{\mathcal{O}}$ there exists a chart $U = \tilde{U}/G_x$ such that for every component V_i of $p^{-1}(U)$, the restriction map $p : V_i \to U$ is a quotient map of an equivariant diffeomorphism $h_i : \tilde{V}_i \to \tilde{U}$ (if $y_i = p^{-1}(x) \cap V_i$ then h_i conjugates the action of G_{y_i} on $\tilde{V}_i$ to the action of a subgroup of Γ_x on $\tilde{U}$).

From now on we will assume that the orbifolds under consideration are connected.

The *universal covering* $p : \tilde{\mathcal{O}} \to \mathcal{O}$ of an orbifold $\mathcal{O}$ is the *initial object* in the category of orbifold coverings, i.e. it is a covering such that for any other covering $p' : \mathcal{O}' \to \mathcal{O}$ there exists a covering $\tilde{p} : \tilde{\mathcal{O}} \to \mathcal{O}'$ satisfying $p' \circ \tilde{p} = p$. If $p : \tilde{\mathcal{O}} \to \mathcal{O}$ is the universal covering then the orbifold $\tilde{\mathcal{O}}$ is called the *universal covering orbifold* of $\mathcal{O}$.

The group $Deck(p)$ of *deck transformations* of an orbifold covering $p : \mathcal{O}' \to \mathcal{O}$ is the group of self-diffeomorphisms $h : \mathcal{O}' \to \mathcal{O}'$ such that $p \circ h = p$. A covering $p : \mathcal{O}' \to \mathcal{O}$ is called *regular* if $\mathcal{O}'/Deck(p) = \mathcal{O}$.

The *fundamental group* $\pi_1(\mathcal{O})$ of the orbifold $\mathcal{O}$ is the group of deck transformations of its universal covering. Then $\mathcal{O} = \tilde{\mathcal{O}}/\pi_1(\mathcal{O})$. An alternative definition of the fundamental group based on homotopy-classes of loops in $\mathcal{O}$ see in [74, Chapter 13].

Theorem D.22 *Each orbifold has a universal covering.*

Definition D.16 An orbifold $\mathcal{O}$ is called *good* if its universal covering is a manifold. Orbifolds which are not good are called *bad*. An orbifold is called *very good* if is admits a finite-sheeted manifold-covering space.

Example D.19 Let $\mathcal{O} = M_\Gamma$ be an n-dimensional complex hyperbolic orbifold. Then $\Gamma = \pi_1(\mathcal{O})$ and $\mathcal{O}$ is a good orbifold: Its universal covering space is $\mathbb{H}^n_{\mathbb{C}}$. If Γ is finitely generated then, according to Selberg's Lemma, the orbifold $\mathcal{O}$ is very good.

Orbifold Bundles Instead of defining orbifold bundles in full generality, I will define these only in the case of compact fibers and connected base, since this will suffice for our purposes:

Definition D.17 A smooth orbi-bundle with compact fibers and connected base is a proper submersion $f : \mathcal{O} \to \mathcal{B}$ between orbifolds. Fibers of f are preimages of points under f.

Note that two different fibers need not be isomorphic to each other, but one can prove that they are *commensurable* in the sense that they have a common finite-sheeted orbi-covering.

Appendix E: Ends of Spaces

Let Z be a locally path-connected, locally compact, Hausdorff topological space. The set of *ends* of Z can be defined as follows (see e.g. [32] for details).

Consider an exhaustion (K_i) of Z by an increasing sequence of compact subsets:

$$K_i \subset K_j, \quad \text{whenever } i \le j,$$

and

$$\bigcup_{i \in \mathbb{N}} K_i = Z.$$

Set $K_i^c := Z \setminus K_i$. The ends of Z are equivalence classes of decreasing sequences of connected components (C_i) of $K_i{}^c$:

$$C_1 \supset C_2 \supset C_3 \supset \cdots$$

Two sequences (C_i), (C'_j) of components of $(K_i{}^c)$, $(K'_j{}^c)$ are said to be equivalent if each C_i contains some C'_j and vice-versa. Then the equivalence class of a sequence (C_i) is an *end* e of Z. Each C_i and its closure is called a *neighborhood* of e in Z. The set of ends of Z is denoted $Ends(Z)$. An end e is called *isolated* if it admits a closed 1-ended neighborhood C; such a neighborhood is called *isolating*.

An alternative view-point on the neighborhoods of ends is that there is a natural topology on the union $\hat{Z} = Z \cup Ends(Z)$ which is a compactification of Z and the neighborhoods C of ends e above are intersections of Z with neighborhoods of e in $\hat{Z}$. Then an end e is isolated if and only if it is an isolated point of $\hat{Z}$. A closed neighborhood C of e in Z is isolating if and only if $C \cup \{e\}$ is closed in $\hat{Z}$.

From this definition it is not immediate that the notion of end is independent of the choice of an exhausting sequence (K_i) of compact subsets. The true, but less intuitive, definition of $Ends(Z)$ is by considering the poset (ordered by the inclusion) of all compact subsets $K \Subset Z$. This poset defines the inverse system of sets

$$\{\pi_0(K^c, x) : K \Subset Z\},$$

where the inclusion $K \subset K'$ induces the map

$$\pi_0(Z - K', x') \to \pi_0(Z - K, x'),$$

with $x' \in Z - K' \subset Z - K$. Taking the inverse limit of this system of sets yields $Ends(Z)$ which is, manifestly, a topological invariant. Furthermore, it is an invariant of the proper homotopy type of Z.

In this survey, I adopt the *analyst's viewpoint* on ends of manifolds and conflate isolated ends and their isolating neighborhoods.

Appendix F: Generalities on Function Theory on Complex Manifolds

For a complex manifold M let $\mathcal{O}_M$ denote the ring of holomorphic functions on M. By a *complex manifold with boundary* M I mean a smooth manifold with (possibly empty) boundary ∂M, such that the interior, $\mathrm{int}(M)$, of the manifold M, is equipped with a complex structure, and there exists a smooth embedding $h : M \to X$ to an equidimensional complex manifold X, biholomorphic on $\mathrm{int}(M)$. A holomorphic function on M is a smooth function which admits a holomorphic extension to a neighborhood of M in X.

Suppose that X is a complex manifold and $Y \subset X$ is a codimension 0 smooth submanifold with boundary in X. The submanifold Y is said to be *strictly Levi-convex* if every boundary point of Y admits a neighborhood U in X such that the submanifold with boundary $Y \cap U$ can be written as

$$\{\phi \le 0\},$$

for some smooth submersion $\phi : U \to \mathbb{R}$ satisfying

$$Hess(\phi) > 0.$$

Here $Hess(\phi)$ is the holomorphic Hessian:

$$\left(\frac{\partial^2 \phi}{\partial \bar{z}_i \partial z_j}\right).$$

(Positivity of the Hessian is independent of the local holomorphic coordinates.)

Example F.20 If $X = \mathbb{C}^n$, $Y = \{z \in \mathbb{C}^n : |z| \le 1\}$, then Y is strictly Levi-convex in X: The complex Hessian of the function $\phi(z) = |z|^2 = z \cdot \bar{z}$ is the identity matrix.

Definition F.18 A *strongly pseudoconvex manifold* M is a complex manifold with boundary which admits a strictly Levi-convex holomorphic embedding in an equidimensional complex manifold.

Suppose, in addition, that M is compact and $h : M \to X$ is a holomorphic embedding with strictly Levi-convex image Y. Then there exists a strictly Levi-convex submanifold $Y' \subset X$ such that $Y \subset \mathrm{int}(Y')$. Accordingly, M can be biholomorphically embedded in the interior of a compact strongly pseudoconvex manifold M'.

Definition F.19 A complex manifold Z is called *holomorphically convex* if for every discrete closed subset $A \subset Z$ there exists a holomorphic function $Z \to \mathbb{C}$ which is proper on A.

Alternatively,[3] one can define holomorphically convex manifolds as follows: For a compact K in a complex manifold M, the *holomorphic convex hull* $\hat{K}_M$ of K in M is

$$\hat{K}_M = \{z \in M : |f(z)| \leq \sup_{w \in K} |f(w)|, \forall f \in \mathcal{O}_M\}.$$

Then M is holomorphically convex if and only if for every compact $K \subset M$, the hull $\hat{K}_M$ is also compact.

Definition F.20 A complex manifold is called *Stein* if it admits a proper holomorphic embedding in $\mathbb{C}^N$ for some N.

Equivalently, M is Stein if and only if it is holomorphically convex and any two distinct points $z, w \in M$ can be separated by a holomorphic function, i.e. there exists $f \in \mathcal{O}_M$ such that $f(z) \neq f(w)$. Yet another equivalent definition is: A complex manifold M is Stein if and only if it is strongly pseudoconvex, i.e. it admits an exhaustion by codimension 0 strongly pseudoconvex complex submanifolds with boundary.

In particular:

Theorem F.23 *The interior of every compact strongly pseudoconvex manifold Z is holomorphically convex.*

Therefore, by holomorphically embedding such (connected manifold) Z in the interior of another compact strongly pseudoconvex manifold Z' and applying Grauert's theorem to Z', it follows that Z admits nonconstant holomorphic functions.

Kohn and Rossi in [57] proved a certain extension theorem for CR functions defined on the boundary of a complex manifold to holomorphic functions on the entire manifold. I will state only a weak form of their result which will suffice for our purposes.

Theorem F.24 (Kohn–Rossi) *Suppose that M is a compact strongly pseudoconvex complex manifold of dimension >1 which admits at least one nonconstant holomorphic function. Then every holomorphic function on ∂M extends to a holomorphic function on the entire M.*

As one of the corollaries of this theorem (Corollary 7.3 of [57]), it follows that if such an M is connected then ∂M is also connected. (If ∂M is disconnected, then one can take a nonconstant locally constant function defined near ∂M: Such a function cannot have a holomorphic extension to M.)

[3] And this is the standard definition.

Remark F.14 If M is Kähler, then Theorem F.24 also holds without the assumption on the existence of nonconstant holomorphic functions, see Proposition 4.4 in [69].

Theorem F.25 (Rossi, [77], Corollary on page 20) *Suppose that M is a compact strongly pseudoconvex complex manifold. Then* $\text{int}(M)$ *admits a proper surjective holomorphic map to a Stein space. In particular, if* $\text{int}(M)$ *contains no compact complex subvarieties of positive dimension, then* $\text{int}(M)$ *is Stein.*

I will not define Stein spaces here (strictly speaking, they are not needed for the purpose of this survey), I refer to [40] for various equivalent definitions.

Topology of Stein Manifolds and Spaces Every complex n-dimensional Stein space is homotopy-equivalent to an n-dimensional CW complex, see [43, 44]. More precisely (see Theorem 1.1* on page 153 of [38]):

Theorem F.26 *Let M be a n-dimensional complex manifold which admits a proper holomorphic map $M \to \mathbb{C}^N$ with fibers of positive codimension. Then M is homotopy-equivalent to an n-dimensional CW complex.*

Corollary F.6 *Suppose that M is a connected compact strongly pseudoconvex complex n-manifold with nonempty boundary. Then M is homotopy-equivalent to a CW complex of dimension $2n-2$.*

Appendix G (by Mohan Ramachandran): Proof of Theorem 2.19

Proposition G.7 *Let X be a complex manifold of dimension ≥ 2 and let $M \subset X$ be a domain with compact nonempty smooth strongly pseudoconvex boundary. Then every pluriharmonic function on M which vanishes at ∂M, vanishes identically.*

Proof The proof mostly follows that of Proposition 4.4 in [69]. Suppose that $M = \{x \in X : \varphi(x) < 0\}$ for some smooth function φ, which is strictly plurisubharmonic on a neighborhood of ∂M and such that there exists $\epsilon < 0$ such that $\varphi^{-1}([\epsilon, 0])$ is compact and $\varphi|_{\partial M} = 0$. Let $\beta : M \to \mathbb{R}$ be a pluriharmonic function which vanishes at ∂M. Fix $a \in (\epsilon, 0)$, such that φ is strictly plurisubharmonic on $V = \{x \in M : \varphi(x) > a\}$. If β does not vanish identically on a neighborhood of ∂M, we let $b \in \beta(V)$ denote a regular value of β. Thus, $\beta^{-1}(b)$ is disjoint from ∂M. Since $\varphi^{-1}([\epsilon, 0])$ is compact, the restriction of φ to $\beta^{-1}(b)$ has a maximum at some point $x_0 \in V \cap \beta^{-1}(b)$. The holomorphic 1-form $\partial\beta$ determines a (singular) holomorphic foliation on M. Consider the leaf L through x_0 of this holomorphic foliation: This leaf is contained in $\beta^{-1}(b)$ and, hence, the restriction $\varphi|_L$ has a maximum at x_0 contradicting strict plurisubharmonicity of φ. Therefore, β is identically zero near ∂M and, hence, is identically zero.

The next proposition is proven in [68, Theorem 2.6]:

Proposition G.8 *Suppose now that M has a complete Kähler metric of bounded geometry,*[4] *∂M is connected and M has at least two ends. Then M admits a nonconstant pluriharmonic function $\beta : M \to \mathbb{R}$ which converges to zero at ∂M.*

By combining the two propositions, we conclude:

Corollary G.7 *Suppose that M is a complex manifold of dimension ≥ 2, which admits a holomorphic embedding as a domain with compact nonempty smooth strongly pseudoconvex boundary and which admits a complete Kähler metric of bounded geometry. Then M is 1-ended.*

We can now conclude the proof of Theorem 2.19: Let $M = M_\Gamma$ be a complex hyperbolic manifold of dimension ≥ 2 and of injectivity radius bounded below. Suppose that $E_0 \subset M$ is a convex end. Let $S_0 \subset \partial \overline{M}$ be the component corresponding to the end E_0. Consider the complex manifold $Y = \check{\Omega}_\Gamma / \Gamma$. Remove from Y all the components of $Y - M$ which are disjoint from S_0 and call the result X. Then M embeds in X as a domain with nonempty smooth strongly pseudoconvex boundary, namely, S_0. Then, by the corollary, M is 1-ended. ◻

Acknowledgements During the work on this paper I was partly supported by the NSF grant DMS-16-04241.

References

1. A. Anan'in, C.H. Grossi, N. Gusevskii, Complex hyperbolic structures on disc bundles over surfaces. Int. Math. Res. Notices **19**, 4295–4375 (2011)
2. A. Ancona, Convexity at infinity and Brownian motion on manifolds with unbounded negative curvature. Rev. Mat. Iberoam. **10**, 189–220 (1994)
3. M. Anderson, The Dirichlet problem at infinity for manifolds of negative curvature. J. Differ. Geom. **18**, 701–721 (1983)
4. W. Ballmann, M. Gromov, V. Schroeder, *Manifolds of Nonpositive Curvature*. Progress in Mathematics, vol. 61, (Birkhäuser, Boston, 1985)
5. G. Barthel, F. Hirzebruch, T. Höfer, *Geradenkonfigurationen und Algebraische Flächen*. Aspects of Mathematics, D4 (Friedr. Vieweg & Sohn, Braunschweig, 1987)
6. I. Belegradek, V. Kapovitch, Classification of negatively pinched manifolds with amenable fundamental groups. Acta Math. **196**, 229–260 (2006)
7. M. Bestvina, G. Mess, The boundary of negatively curved groups. J. Amer. Math. Soc. **4**, 469–481 (1991)
8. M. Boileau, S. Maillot, J. Porti, *Three-Dimensional Orbifolds and Their Geometric Structures*. Panoramas and Syntheses, vol. 15 (Société Mathématique de France, 2003)
9. A. Borbély, Some results on the convex hull of finitely many convex sets. Proc. Amer. Math. Soc. **126**(5), 1515–1525 (1998)
10. M. Bourdon, Sur la dimension de Hausdorff de l'ensemble limite d'une famille de sous-groupes convexes co-compacts. C. R. Acad. Sci. Paris Sér. I Math. **325**, 1097–1100 (1997)
11. B.H. Bowditch, Geometrical finiteness with variable negative curvature. Duke Math. J. **77**(1), 229–274 (1995)

[4] I.e. its sectional curvature lies in a finite interval and its injectivity radius is bounded from below.

12. B.H. Bowditch, Convergence groups and configuration spaces, in *Geometric Group Theory Down Under* (Canberra, 1996) (de Gruyter, Berlin, 1999), pp. 23–54
13. M. Burger, A. Iozzi, A measurable Cartan theorem and applications to deformation rigidity in complex hyperbolic geometry. Pure Appl. Math. Q. **4**(1), Special Issue: In honor of Grigory Margulis. Part 2, 181–202 (2008)
14. D. Burns, S. Shnider, Spherical hypersurfaces in complex manifolds. Invent. Math. **33**, 223–246 (1976)
15. A. Cano, J. Navarrete, J. Seade, *Complex Kleinian Groups*. Progress in Mathematics, vol. 303 (Springer, Berlin, 2013)
16. A. Cano, J. Seade, On the equicontinuity region of discrete subgroups of $PU(1,n)$. J. Geom. Analy. **20**, 291–305 (2010)
17. J. Carlson, S. Müller-Stach, C. Peters, *Period Mappings and Period Domains*. Cambridge Studies in Advanced Mathematics (Cambridge University Press, Cambridge, 2003)
18. M. Coornaert, Mesures de Patterson-Sullivan sur le bord d'un espace hyperbolique au sens de Gromov. Pacific J. Math. **159**, 241–270 (1993)
19. K. Corlette, Flat G-bundles with canonical metrics. J. Differ. Geom. **28**(3), 361–382 (1988)
20. K. Corlette, Hausdorff dimensions of limit sets I. Invent. Math. **102**, 521–541 (1990)
21. K. Corlette, Archimedean superrigidity and hyperbolic geometry. Ann. Math. **135**(2), 165–182 (1992)
22. K. Corlette, A. Iozzi, Limit sets of discrete groups of isometries of exotic hyperbolic spaces. Trans. Amer. Math. Soc. **351**(4), 1507–1530 (1999)
23. W. Couwenberg, G. Heckman, E. Looijenga, Geometric structures on the complement of a projective arrangement. Publ. Math. Inst. Hautes Études Sci. **101**, 69–161 (2005)
24. F. Dal'bo, J.-P. Otal, M. Peigné, Séries de Poincaré des groupes géométriquement finis. Israel J. Math. **118**, 109–124 (2000)
25. T. Das, D. Simmons, M. Urbański, *Geometry and Dynamics in Gromov Hyperbolic Metric Spaces. With an Emphasis on Non-proper Settings*. Mathematical Surveys and Monographs, vol. 218 (American Mathematical Society, Providence, 2017)
26. P. Deligne, G.D. Mostow, *Commensurabilities Among Lattices in $PU(1,n)$*. Annals of Mathematics Studies, vol. 132 (Princeton University Press, Princeton, 1993)
27. T. Delzant, L'invariant de Bieri–Neumann–Strebel des groupes fondamentaux des variétés Kählériennes. Math. Ann. **348**(1), 119–125 (2010)
28. M. Deraux, A new non-arithmetic lattice in $PU(3, 1)$. Algebr. Geom. Topol. **20**(2), 925–963 (2020)
29. M. Desgroseilliers, F. Haglund, On some convex cocompact groups in real hyperbolic space. Geom. Topol. **17**, 2431–2484 (2013)
30. S. Dey, M. Kapovich, A note on complex-hyperbolic Kleinian groups. Arnold Math. J. **6**(3–4), 397–406 (2020)
31. M. Deraux, J. Parker, J. Paupert, New non-arithmetic complex hyperbolic lattices. Invent. Math. **203**(3), 681–771 (2016)
32. C. Drutu, M. Kapovich, *Geometric Group Theory* (AMS Colloquium Publications, Providence, 2018)
33. D.B.A. Epstein, Complex hyperbolic geometry, in *Analytical and Geometric Aspects of Hyperbolic Space* (Coventry/Durham, 1984) (Cambridge University Press, Cambridge, 1987), pp. 93–111
34. W. Goldman, Complex hyperbolic Kleinian groups, in *Complex Geometry* (Osaka, 1990). Lecture Notes in Pure and Applied Mathematics, vol. 143 (Dekker, New York, 1993), pp. 31–52
35. W. Goldman, *Complex Hyperbolic Geometry*. Oxford Mathematical Monographs (Clarendon Press, Oxford, 1999)
36. W. Goldman, J.J. Millson, Local rigidity of discrete groups acting on complex hyperbolic space. Invent. Math. **88**(3), 495–520 (1987)
37. W. Goldman, M. Kapovich, B. Leeb, Complex hyperbolic surfaces homotopy-equivalent to a Riemann surface. Commun, Analy. Geom. **9**, 61–96 (2001)

38. M. Goresky, R. MacPherson, *Stratified Morse Theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 14 (Springer, Berlin, 1988)
39. J. Granier, Groupes discrets en géométrie hyperbolique – Aspects effectifs. PhD Thesis, Université de Fribourg, 2015
40. H. Grauert, R. Remmert, *Stein Spaces*. Grundlehren der Mathematischen Wissenschaften, vol. 236 (Springer, Berlin, 1979)
41. M. Gromov, I. Piatetski-Shapiro, Nonarithmetic groups in Lobachevsky spaces. Inst. Hautes Études Sci. Publ. Math. **66**, 93–103 (1988)
42. M. Gromov, R. Schoen, Harmonic maps into singular spaces and p-adic superrigidity for lattices in groups of rank one. Inst. Hautes Études Sci. Publ. Math. **76**, 165–246 (1992)
43. H. Hamm, Zur Homotopietyp Steinscher Räume. J. Reine Angew. Math. **338**, 121–135 (1983)
44. H. Hamm, Zum Homotopietyp q-vollständiger Räume. J. Reine Angew. Math. **364**, 1–9 (1986)
45. J.A. Hillman, D.H. Kochloukova, Finiteness conditions and $PD(r)$-group covers of $PD(n)$-complexes. Math. Z. **256**, 45–56 (2007)
46. F. Hirzebruch, Arrangements of lines and algebraic surfaces, in *Arithmetic and Geometry, Vol. II*, Progress in Mathematics, vol. 36 (Birkhäuser, Boston, Mass., 1983), pp. 113–140
47. R.-P. Holzapfel, *Ball and Surface Arithmetics*. Aspects of Mathematics, vol. E29 (Friedr. Vieweg & Sohn, Braunschweig, 1998)
48. M. Kapovich, On normal subgroups in the fundamental groups of complex surfaces (1998). Preprint math.GT/9808085
49. M. Kapovich, *Hyperbolic Manifolds and Discrete Groups*. Birkhauser's Series "Progress in Mathematics" (Springer, Berlin, 2000). Reprinted in 2009
50. M. Kapovich, Representations of polygons of finite groups. Geom. Topol. **9**, 1915–1951 (2005)
51. M. Kapovich, Kleinian groups in higher dimensions, in *Geometry and Dynamics of Groups and Spaces*. Progress in Mathematics, vol. 265 (Birkhäuser, Basel, 2008), pp. 487–564
52. M. Kapovich, A note on Selberg's lemma and negatively pinched Hadamard manifolds (2018). ArXiv, 1808.01602 . To appear in Journal of Differential Geometry
53. M. Kapovich, B. Kleiner, Hyperbolic groups with low-dimensional boundary. Ann. Sci. École Norm. Sup. **33**(5), 647–669 (2000)
54. M. Kapovich, B. Liu, Geometric finiteness in negatively pinched Hadamard manifolds. Ann. Acad. Sci. Fennicae **44**(2), 841–875 (2019)
55. M. Kapovich, B. Leeb, J. Porti, Dynamics on flag manifolds: domains of proper discontinuity and cocompactness. Geom. Topol. **22**, 157–234 (2017)
56. D. Kazhdan, Some applications of the Weil representation. J. Analy. Mat. **32**, 235–248 (1977)
57. J.J. Kohn, H. Rossi, On the extension of holomorphic functions from the boundary of a complex manifold. Ann. Math. **81**(2), 451–472 (1965)
58. V. Koziarz, N. Mok, Nonexistence of holomorphic submersions between complex unit balls equivariant with respect to a lattice and their generalizations. Amer. J. Math. **132**(5), 1347–1363 (2010)
59. S. Krantz, *Function Theory of Several Complex Variables*. Reprint of the 1992 edition. (AMS Chelsea Publishing, Providence, 2001)
60. F. Ledrappier, *Structure au bord des variétés à courbure négative*, In *"Séminaire de Théorie Spectrale et Géométrie, No. 13, Année 1994–1995"*. Sémin. Théor. Spectr. Géom. , vol. 13 (Univ. Grenoble I, Saint-Martin-d'Hères, 1995), pp. 97–122
61. E. Leuzinger, Critical exponents of discrete groups and L^2-spectrum. Proc. Amer. Math. Soc. **132**(3), 919–927 (2004)
62. K. Liu, Geometric height inequalities. Math. Res. Lett. **3**, 693–702 (1996)
63. R. Livne, On certain covers of the universal elliptic curve. Ph.D. Thesis, Harvard University, 1981
64. G. A. Margulis, Discrete groups of motions of manifolds of nonpositive curvature, in *Proceedings of the International Congress of Mathematicians* (Vancouver, B.C., 1974), vol. 2 (Canadian Mathematical Congress, Montreal, 1975), pp. 21–34
65. G.A. Margulis, *Discrete Subgroups of Semisimple Lie Groups*. Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 17 (Springer, Berlin, 1991)

66. D.B. McReynolds, *Arithmetic lattices in* $SU(n, 1)$ (Preprint, 2005)
67. G.D. Mostow, On a remarkable class of polyhedra in complex hyperbolic space. Pac. J. Math. **86**, 171–276 (1980)
68. T. Napier, M. Ramachandran, Structure theorems for complete Kähler manifolds and applications to Lefschetz type theorems. GAFA **5**, 809–851 (1995)
69. T. Napier, M. Ramachandran, The L^2 $\bar{\partial}$-method, weak Lefschetz theorems, and the topology of Kähler manifolds. J. Amer. Math. Soc. **11**(2), 375–396 (1998)
70. J. Parker, *Hyperbolic Spaces*. Jyväskylä Lectures in Mathematics, vol. 2 (2008)
71. J. Parker, Complex hyperbolic lattices, in *Discrete Groups and Geometric Structures*. Contemporary Mathematics, vol. 501 (American Mathematical Society, Providence, 2009), pp. 1–42
72. J. Paupert, *Introduction to Hyperbolic Geometry*. Arizona State University Lecture Notes, (Springer, Berlin, 2016)
73. M. Prokhorov, Absence of discrete groups of reflections with a noncompact fundamental polyhedron of finite volume in a Lobachevskii space of high dimension. Izv. Akad. Nauk SSSR Ser. Mat. **50**(2), 413–424 (1986)
74. J. Ratcliffe, *Foundations of Hyperbolic Manifolds*. Graduate Texts in Mathematics, vol. 149, 2nd edn. (Springer, New York, 2006)
75. T. Roblin, S. Tapie, Critical exponent and bottom of the spectrum in pinched negative curvature. Math. Res. Lett. **22**(3), 929–944 (2015)
76. J. Rogawski, *Automorphic Representations of Unitary Groups in Three Variables*. Annals of Mathematics Studies, vol. 123 (Princeton University Press, Princeton, 1990)
77. H. Rossi, Strongly pseudoconvex manifolds, in *Lectures in Modern Analysis and Applications, I* (Springer, Berlin, 1969), pp. 10–29
78. I. Satake, On a generalization of the notion of manifold. Proc. Nat. Acad. Sci. U.S.A. **42**, 359–363 (1956)
79. R. Schwartz, *Spherical CR Geometry and Dehn Surgery*. Annals of Mathematics Studies, vol. 165 (Princeton University Press, Princeton, 2007)
80. O. Shvartsman, An example of a nonarithmetic discrete group in the complex ball, in *Lie Groups, Their Discrete Subgroups, and Invariant Theory*. Advances in Soviet Mathematics, vol. 8 (American Mathematical Society, Providence, 1992), pp. 191–196
81. C. Simpson, Higgs bundles and local systems. Inst. Hautes Études Sci. Publ. Math. **75**, 5–95 (1992)
82. W. Thurston, *The Geometry and Topology of Three-Manifolds*, Chap. 13. Princeton University Lecture Notes (Princeton University Press, Providence, 1979), pp. 1978–1981.
83. W. Thurston, Shapes of polyhedra and triangulations of the sphere. Geom. Topol. Monogr. **1**, 511–549 (1998)
84. D. Toledo, Harmonic maps from surfaces to certain Kähler manifolds. Math. Scandinavica **45**, 13–26 (1979)
85. P. Tukia, Convergence groups and Gromov-hyperbolic metric spaces. New Zealand J. Math. **23**, 157–187 (1994)
86. P. Tukia, Conical limit points and uniform convergence groups. J. Reine Angew. Math. **501**, 71–98 (1998)
87. E. Vinberg, Absence of crystallographic groups of reflections in Lobachevskii spaces of large dimension. Trudy Moskov. Mat. Obshch. **47**, 68–102 (1984)
88. N. Wallach, Square integrable automorphic forms and cohomology of arithmetic quotients of $SU(p, q)$. Math. Ann. **266**, 261–278 (1984)
89. A. Weil, Remarks on the cohomology of groups. Ann. Math. **80**(2), 149–157 (1964)
90. S.-T. Yau, Calabi's conjecture and some new results in algebraic geometry. Proc. Nat. Acad. Sci. U.S.A. **74**(5), 1798–1799 (1977)
91. S.-K. Yeung, Virtual first Betti number and integrality of compact complex two-ball quotients. Int. Math. Res. Not. **38**, 1967–1988 (2004)
92. C. Yue, Webster curvature and Hausdorff dimension of complex hyperbolic Kleinian groups, in *Dynamical Systems — Proceedings Of The International Conference In Honor Of Professor Liao Shantao* (World Scientific, Singapore, 1999), pp. 319–328

Chapter 3
Möbius Structures, Hyperbolic Ends and k-Surfaces in Hyperbolic Space

Graham Smith

Abstract Möbius surfaces and hyperbolic ends are key tools used in the study of geometrically finite three-dimensional hyperbolic manifolds. We review the theory of Möbius surfaces and describe a new framework for the theory of hyperbolic ends. We construct the ideal boundary functor sending hyperbolic ends into Möbius surfaces, and the extension functor sending Möbius surfaces into hyperbolic ends. We show that the former is a right inverse of the latter, and we show that every hyperbolic end canonically embeds into the extension of its ideal boundary. We conclude by showing that, for any given Möbius surface, there exists a unique *maximal* hyperbolic end having that Möbius surface for its ideal boundary.

We apply these theories to the study of infinitesimally strictly convex (ISC) surfaces in $\mathbb{H}^3$ which are complete with respect to the sums of their first and third fundamental forms (called quasicomplete in the sequel). We prove a new a priori C^0 estimate for such surfaces. We apply this estimate to provide a complete solution of a Plateau-type problem for surfaces of constant extrinsic curvature in $\mathbb{H}^3$ posed by Labourie in 2000 (Invent Math 141:239–297). We conclude by describing new parametrisations of the spaces of quasicomplete, ISC, constant extrinsic curvature surfaces in $\mathbb{H}^3$ by open subsets of spaces of holomorphic functions.

Keywords Hyperbolic ends · Möbius structures · Flat conformal structures · Complex projective structures · Extrinsic curvature · Asymptotic Plateau problems

Classification AMS 57K32, 53C18, 53C42, 58D10

G. Smith (✉)
Rio de Janeiro, Brazil

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_3

3.1 Overview

3.1.1 Hyperbolic Ends and Möbius Structures

In the words of Thurston, within the family of all three-dimensional manifolds, hyperbolic three-manifolds make up "by far the most interesting, the most complex, and the most useful" class (see [30]). In this chapter, we will only be concerned with two- and three-dimensional manifolds, which we will henceforth refer to simply as *surfaces* and *manifolds* respectively. In addition, in order to avoid an avalanche of unwieldy expressions, we will call a hyperbolic manifold *geometrically finite* whenever it is complete, oriented, of finite topological type and without cusps. Our aim is to present two of the main constructs used in the study of such manifolds, namely hyperbolic ends and Möbius structures.

Hyperbolic manifolds are locally modelled on three-dimensional hyperbolic space $\mathbb{H}^3$. For ease of visualisation, it is helpful to identify this space with the open unit ball $\mathbb{B}_1^3$ in $\mathbb{R}^3$ furnished with the *Beltrami–Klein metric*

$$g_{ij}^{\mathrm{BK}} := \frac{\delta_{ij}}{(1-\|x\|^2)} + \frac{x^i x^j}{(1-\|x\|^2)^2}. \tag{3.1.1}$$

This is called the *Beltrami–Klein model* of $\mathbb{H}^3$ (see [5]). Its most useful property for our purposes is that its metric is affine equivalent to the standard Euclidean metric in the sense that the geodesics of the one coincide, as sets, with the geodesics of the other. In particular, a subset K of the unit ball is convex as a subset of $\mathbb{H}^3$ if and only if it is convex as a subset of $\mathbb{R}^3$.

Let $\partial_\infty\mathbb{H}^3$ denote the ideal boundary of $\mathbb{H}^3$ which, we recall, is defined to be the space of equivalence classes of complete geodesic rays in $\mathbb{H}^3$, where two such rays are deemed equivalent whenever they are asymptotic to one another (see [2]). In the Beltrami–Klein model, equivalence classes are uniquely defined by their end points, so that $\partial_\infty\mathbb{H}^3$ identifies topologically with the unit sphere $\mathbb{S}_1^2$, and the union $\mathbb{H}^3 \cup \partial_\infty\mathbb{H}^3$ likewise identifies topologically with the closed unit ball $\overline{\mathbb{B}}_1^3$.

Let $\mathrm{PSO}_0(3,1)$ denote the group of orientation preserving isometries of $\mathbb{H}^3$. Recall that its action extends uniquely to a continuous action on $\mathbb{H}^3 \cup \partial_\infty\mathbb{H}^3$.

Definition 3.1.1 Let S be a compact, oriented surface of genus at least 2, let Π denote its fundamental group and let $\theta : \Pi \to \mathrm{PSO}_0(3,1)$ be an injective homomorphism with discrete image. We say that θ is a *quasi-Fuchsian representation* whenever it preserves a Jordan curve in $\partial_\infty\mathbb{H}^3$. We say that a hyperbolic manifold X is *quasi-Fuchsian* whenever it is isometric to the quotient of $\mathbb{H}^3$ by the image of some quasi-Fuchsian representation.

Remark 3.1.2 The quasi-Fuchsian manifold X is a complete hyperbolic manifold diffeomorphic to $S \times \mathbb{R}$ (see [5] and [31]).

Remark 3.1.3 The Jordan curve C preserved by $\theta(\Pi)$ coincides with the limit set of the $\theta(\Pi)$-orbit of every point of $\mathbb{H}^3 \cup \partial_\infty \mathbb{H}^3$. In particular, C is uniquely defined by this representation.

Quasi-Fuchsian manifolds are geometrically finite. In fact, they are the archetypical examples of this class of manifold. Of their various interesting properties, two will concern us in particular. The first is a certain natural decomposition, which is constructed as follows. Let $\theta : \Pi \rightarrow \mathrm{PSO}_0(3, 1)$ be a quasi-Fuchsian representation, let $C \subseteq \partial_\infty \mathbb{H}^3$ denote the unique Jordan curve that it preserves, and let $X := \mathbb{H}^3/\theta(\Pi)$ denote the quasi-Fuchsian manifold that it defines. Let $\tilde{K}$ denote the convex hull of C in $\mathbb{H}^3$ and let $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ denote the two connected components of its complement. Since $\theta(\Pi)$ preserves $\tilde{K}$, $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$, their respective quotients K, Ω_1 and Ω_2 identify with subsets of X, and we thus obtain the decomposition

$$X := K \cup \Omega_1 \cup \Omega_2. \tag{3.1.2}$$

Furthermore, K is the minimal, closed, convex subset onto which X retracts (see [5]). More generally (see [15]), every geometrically finite hyperbolic manifold decomposes in this way as the union of such a minimal, closed, convex subset, known as its *Nielsen kernel*, and finitely many unbounded open subsets, of varying topological type, known as its *ends*.

Definition 3.1.4 A *height function* over a hyperbolic manifold Y is defined to be a locally strictly convex, $C^{1,1}_{\mathrm{loc}}$ function $h : Y \rightarrow]0, \infty[$ such that

(1) the gradient flow lines of h are unit speed geodesics; and
(2) for all $t > 0$, $h^{-1}([t, \infty[)$ is complete.

We say that a hyperbolic manifold X is a *hyperbolic end* whenever it carries a height function.

Remark 3.1.5 Height functions, whenever they exist, are unique (see Lemma 3.3.5).

Remark 3.1.6 We are not aware of a similar definition of hyperbolic ends having been used before in the literature. However, we will show in Sect. 3.3 that Definition 3.1.4 yields a rich and coherent theory. We believe that it has the virtues over earlier definitions of being more direct and of lending itself better to potential generalisations.

Consider now the quasi-Fuchsian manifold X and its three components introduced above. By standard properties of convex subsets of hyperbolic space (see [2]), the open sets $\tilde{\Omega}_1$ and $\tilde{\Omega}_2$ are both hyperbolic ends with height functions given by distance in $\mathbb{H}^3$ to $\tilde{K}$. Since the above construction is invariant under the action of $\theta(\Pi)$, the quotients Ω_1 and Ω_2 are also hyperbolic ends. More generally, the connected components of the complement of the Nielsen kernel of any geometrically finite hyperbolic manifold are hyperbolic ends so that the theory of hyperbolic ends encompasses the large scale geometry of geometrically finite hyperbolic manifolds.

The second property of quasi-Fuchsian manifolds that interests us concerns their asymptotic geometry. Indeed, with X as above, we define its *ideal boundary* $\partial_\infty X$ to be the space of equivalence classes of complete geodesic rays in X which are not contained in any compact set where, again, two such rays are deemed equivalent whenever they are asymptotic to one another. The lifts of such rays are complete geodesic rays in $\mathbb{H}^3$ whose end points are not elements of C, so that $\partial_\infty X$ identifies with the quotient of $\partial_\infty \mathbb{H}^3 \setminus C$ under the action of $\theta(\Pi)$.

We now recall that $\partial_\infty \mathbb{H}^3$ naturally identifies with the Riemann sphere $\hat{\mathbb{C}}$ and that the action of $\mathrm{PSO}_0(3, 1)$ on $\partial_\infty \mathbb{H}^3$ identifies with the action of the Möbius group $\mathrm{PSL}(2, \mathbb{C})$ on this space. This identification is immediately visible in the Beltrami–Klein model, since here the natural holomorphic structure of $\partial_\infty \mathbb{H}^3$ is none other than the structure that it inherits as a smooth, embedded submanifold of $\mathbb{R}^3$.

Definition 3.1.7 Let S be a surface. A *Möbius structure* (also known as a *flat conformal structure* or a *complex projective structure*) over S is an atlas $\mathcal{A}$ of S in $\hat{\mathbb{C}}$ all of whose transition maps are restrictions of Möbius maps. A *Möbius surface* is a pair $(S, \mathcal{A})$ where S is a surface and $\mathcal{A}$ is a Möbius structure over S. In what follows, when no ambiguity arises, we will denote the Möbius surface simply by S.

For each i, we denote $\tilde{\Sigma}_i := \partial_\infty \tilde{\Omega}_i$, so that the complement of C in $\partial_\infty \mathbb{H}^3$ decomposes as

$$\partial_\infty \mathbb{H}^3 \setminus C = \tilde{\Sigma}_1 \cup \tilde{\Sigma}_2. \tag{3.1.3}$$

For each i, $\tilde{\Sigma}_i$ is trivially a Möbius surface and, since $\theta(\Pi)$ acts on $\tilde{\Sigma}_i$ by Möbius transformations, the quotient surface

$$\Sigma_i := \partial_\infty \tilde{\Omega}_i / \theta(\Pi) \tag{3.1.4}$$

is also a Möbius surface. In this manner, we obtain a decomposition

$$\partial_\infty X = \Sigma_1 \cup \Sigma_2 \tag{3.1.5}$$

of the ideal boundary of X into the union of two Möbius surfaces, each homeomorphic to S. More generally, the ideal boundary of any geometrically finite hyperbolic manifold consists of the union of finitely many compact Möbius surfaces, one for each end, so that the theory of Möbius surfaces encompasses the asymptotic geometry of geometrically finite hyperbolic manifolds.

We underline, however, that these theories extend beyond the theory of geometrically finite hyperbolic manifolds. Indeed, it is straightforward to construct hyperbolic ends and Möbius surfaces which do not arise respectively as the ends or ideal boundaries of such manifolds. Nevertheless, in Sects. 3.3.4 and 3.3.6, we show that every hyperbolic end X still has a well-defined ideal boundary, denoted by $\partial_\infty X$, given by the space of equivalence classes of complete geodesic rays in X, and that this ideal boundary naturally carries the structure of a Möbius surface. Conversely, in Sects. 3.3.5 and 3.3.6, we show that, for every Möbius surface S, there exists a canonical hyperbolic end, which we denote by $\mathcal{H}S$, and which we call its *extension*, whose ideal boundary is canonically isomorphic to S.[1] It follows that the theories of hyperbolic ends and Möbius structures are naturally developed in tandem. However, in contrast to the presentation of this introduction, we find that the theory of Möbius structures precedes that of hyperbolic ends, and for this reason it will be studied first in the following sections.

In Sects. 3.2 and 3.3, we comprehensively review the foundations of these theories and the relationships between them. We have chosen to derive our results using only classical tools of hyperbolic geometry, such as geodesics, spheres, horospheres, and so on. The reader will notice certain similarities with aspects of the work [16] of Kulkarni. Nonetheless, we find that our approach yields simpler proofs of existing results and useful generalisations of others.

Two main themes will be of particular interest to us. The first concerns the construction and properties of certain special functions which encode global geometry in a local manner. In the case of hyperbolic ends, this function will be none other than the height function defined above, whose analytic properties we will establish in some detail. In the case of Möbius surfaces, it will be a $C^{1,1}_{\text{loc}}$ section of the density bundle of the surface which we call the Kulkarni–Pinkall form. This form, first studied in [17], is naturally related to the horospherical support function of immersed surfaces in $\mathbb{H}^3$ (see [8] and [22]) and for this reason constitutes a key ingredient of useful a priori estimates that we will develop in Sect. 3.4 and which we will discuss presently.

The second main theme that interests us is the construction of the operators ∂_∞ and $\mathcal{H}$ mentioned above. These operators allow us to pass back and forth between the families of hyperbolic ends and Möbius surfaces. In particular, they allow us to compare the geometries of different hyperbolic ends with the same ideal boundaries, and we thereby obtain the following nice result. We say that a hyperbolic end X is *maximal* if it cannot be isometrically embedded in a strictly larger hyperbolic end with the same ideal boundary.

Theorem 3.1.8 (Maximality) *For every Möbius surface S, the extension $\mathcal{H}S$ of S is, up to isometry, the unique maximal hyperbolic end with ideal boundary S.*

Remark 3.1.9 To form a clearer idea of the concept of maximality, consider two half-spaces $H_1, H_2 \subseteq \mathbb{H}^3$ such that H_2 is strictly contained in H_1. Although H_1 and H_2 are both maximal hyperbolic ends, this does not invalidate the definition, since the ideal boundary of the second is strictly contained in that of the first.

Remark 3.1.10 We prove Theorem 3.1.8 in Sect. 3.3.6. In the case where S is compact, this result follows from the work [21] of Scannell via the natural duality between hyperbolic ends and GHMC de Sitter spacetimes (see [9]). An independent proof of the compact case was also provided by the author in [24].

[1] The extension coincides with the hyperbolic end constructed by Kulkarni–Pinkall in Section 8 of [17], where it is called the H-hull of the Möbius surface.

3.1.2 Infinitesimal Strict Convexity, Quasicompleteness and the Asymptotic Plateau Problem

We now discuss the applications of the theories of Möbius surfaces and hyperbolic ends to the study of certain types of immersed surfaces in $\mathbb{H}^3$.

Definition 3.1.11 An *immersed surface* in $\mathbb{H}^3$ is a pair (S, e), where S is an oriented surface and $e : S \to \mathbb{H}^3$ is a smooth immersion. In what follows, we denote the immersed surface sometimes by S and sometimes by e, depending on which is more appropriate to the context.

We first recall some standard definitions of surface theory (c.f. [6]). Let S be an immersed surface. Let $\mathrm{U}\mathbb{H}^3$ denote the bundle of unit vectors over $\mathbb{H}^3$. Let $N_e : S \to \mathrm{U}\mathbb{H}^3$ denote the positively oriented unit normal vector field over e. The *first*, *second* and *third fundamental forms* of e are respectively the symmetric bilinear forms I_e, II_e and III_e defined over S such that, for every pair ξ, ν of vector fields over this surface,

$$\mathrm{I}_e(\xi, \nu) := \langle De \cdot \xi, De \cdot \nu\rangle, \tag{3.1.6}$$

$$\mathrm{II}_e(\xi, \nu) := \langle \nabla_\xi N_e, De \cdot \nu\rangle, \text{ and} \tag{3.1.7}$$

$$\mathrm{III}_e(\xi, \nu) := \langle \nabla_\xi N_e, \nabla_\nu N_e\rangle, \tag{3.1.8}$$

where ∇ here denotes the Levi–Civita covariant derivative of $\mathbb{H}^3$. The *shape operator* of S is the field A_e of endomorphisms of TS defined such that

$$\mathrm{II}_e(\cdot, \cdot) =: \mathrm{I}_e(A_e\cdot, \cdot). \tag{3.1.9}$$

In particular, the shape operator is symmetric with respect to I_e and the third fundamental form of S is expressed in terms of the first fundamental form and the shape operator by

$$\mathrm{III}_e(\cdot, \cdot) = \mathrm{I}_e(A_e^2\cdot, \cdot). \tag{3.1.10}$$

Finally, the *extrinsic curvature* of S is defined by

$$K_e := \mathrm{Det}(A_e). \tag{3.1.11}$$

We now restrict attention to a class of immersed surfaces to which the theories of Möbius surfaces and hyperbolic ends naturally apply.

Definition 3.1.12 Let (S, e) be an immersed surface in $\mathbb{H}^3$. We say that (S, e) is *quasicomplete* whenever the metric $\mathrm{I}_e + \mathrm{III}_e$ is complete and we say that it is *infinitesimally strictly convex* (ISC) whenever its second fundamental form is everywhere positive-definite.

Let (S, e) be a quasicomplete, ISC immersed surface in $\mathbb{H}^3$. We associate a natural hyperbolic end to S as follows. First, we denote $\mathcal{E}S := S \times [0, \infty[$ and we define the function $\mathcal{E}e : \mathcal{E}S \to \mathbb{H}^3$ by

$$\mathcal{E}e(x, t) := \mathrm{Exp}(tN_e(x)). \tag{3.1.12}$$

By local strict convexity of S, $\mathcal{E}e$ is an immersion, and we thus furnish the manifold $\mathcal{E}S$ with the unique hyperbolic structure that makes it into a local isometry. Quasicompleteness then implies that $\mathcal{E}S$ is, in fact, a hyperbolic end (see Lemma and Definition 3.4.1), which we call the *end* of S. In fact, a converse also holds: $\mathcal{E}S$ is a hyperbolic end if and only if S is quasicomplete and II_e is non-negative semi-definite.

In order to describe the natural Möbius structure associated to S, we now recall the concept of developing maps. Let S be a surface and let $\phi : S \to \hat{\mathbb{C}}$ be a local diffeomorphism. For every point x of S, there exists a neighbourhood U of x over which ϕ restricts to a diffeomorphism onto its image V. The set $\mathcal{A} := (U_\alpha, V_\alpha, \phi)_{\alpha \in A}$ forms an atlas of S in $\hat{\mathbb{C}}$ whose transition maps are trivial, and thus a fortiori Möbius. We call $\mathcal{A}$ the *pull-back* Möbius structure of ϕ. Given any Möbius surface S, we say that a local diffeomorphism $\phi : S \to \hat{\mathbb{C}}$ is a *developing map* of S whenever its pull-back Möbius structure is compatible with the initial Möbius structure of the surface. Not every Möbius surface possesses a developing map, although every simply-connected Möbius surface trivially does. We say that a Möbius surface is *developable* whenever a developing map exists. Likewise, we define a *developed* Möbius surface to be a pair (S, ϕ), where S is a surface and $\phi : S \to \hat{\mathbb{C}}$ is a local diffeomorphism. Naturally, in this case, S is furnished with the pull-back Möbius structure of ϕ.

We now return to the case where (S, e) is a quasicomplete ISC immersed surface in $\mathbb{H}^3$. We define the *horizon map* $\mathrm{Hor} : \mathrm{U}\mathbb{H}^3 \to \partial_\infty \mathbb{H}^3$ such that, for every unit speed geodesic $\gamma : \mathbb{R} \to \mathbb{H}^3$,

$$\mathrm{Hor}(\dot{\gamma}(0)) := \lim_{t \to +\infty} \gamma(t). \tag{3.1.13}$$

We define *asymptotic Gauss map* of S by

$$\phi_e := \mathrm{Hor} \circ N_e, \tag{3.1.14}$$

where N_e here denotes the positively oriented unit normal of e. By infinitesimal strict convexity of S, this function is a local diffeomorphism from S into $\partial_\infty \mathbb{H}^3$

(see [2]).[2] In particular, (S, ϕ_e) is a developed Möbius surface which we call the *asymptotic Gauss image* of S.

We have thus seen how hyperbolic ends and Möbius surfaces are associated to quasicomplete, ISC immersed surfaces in $\mathbb{H}^3$. We now show how these constructions yield a useful new a priori estimate for such surfaces. We first require the following parametrisation of the space of open horoballs in $\mathbb{H}^3$ by $\Lambda^2\partial_\infty\mathbb{H}^3$. Let $y \in \partial_\infty\mathbb{H}^3$ be an ideal point. Let $B \subseteq \mathbb{H}^3$ be an open horoball centred on y. Let H be an open half-space whose boundary is an exterior tangent to B at some point. Let $D := \partial_\infty H$ denote the ideal boundary of H and let $\omega(D)$ denote the area form of its Poincaré metric. It turns out that $\omega(D)(y)$ only depends on B. We call y and $\omega_y := \omega(D)(y)$ the *asymptotic centre* and the *asymptotic curvature* of B respectively, and we verify that these data define B uniquely. For all $\omega_y \in \Lambda^2\partial_\infty\mathbb{H}^3$, we henceforth denote by $B(\omega_y)$ the open horoball in $\mathbb{H}^3$ with asymptotic centre y and asymptotic curvature ω_y. In this manner, we obtain the desired parametrisation of the space of open horoballs in $\mathbb{H}^3$ by $\Lambda^2\partial_\infty\mathbb{H}^3$.

In Sect. 3.2.4, we associate to every Möbius surface S a canonical section of $\Lambda^2 S$ which we call its Kulkarni–Pinkall form. The push-forward of this section through any developing map is a function taking values in $\Lambda^2\partial_\infty\mathbb{H}^3$ which, by the preceding discussion, associates to every point of S an open horoball in $\mathbb{H}^3$.

Theorem 3.1.13 (A Priori Estimate) *Let (S, e) be a quasicomplete ISC immersed surface in $\mathbb{H}^3$, let ϕ denote its asymptotic Gauss map and let ω denote the Kulkarni–Pinkall form of the developed Möbius surface (S, ϕ). For all $x \in S$,*

$$e(x) \in \overline{B}(\phi_*\omega(x)). \tag{3.1.15}$$

Remark 3.1.14 Theorem 3.1.13 follows immediately from Theorem 3.4.8.

This estimate in turn yields a complete solution to a problem of Plateau-type concerning surfaces of constant extrinsic curvature, as we now show. First, following the work [19] of Labourie, we make the following two definitions.

Definition 3.1.15 For $k > 0$, a *k-surface* is a quasicomplete, ISC immersed surface in $\mathbb{H}^3$ of constant extrinsic curvature equal to k. In what follows, we denote the k-surface sometimes by S and sometimes by e, depending on which is more appropriate to the context.

Definition 3.1.16 Let (S, ϕ) be a developed Möbius surface. For $k > 0$, we say that a k-surface $e : S \to \mathbb{H}^3$ is a *solution* to the asymptotic Plateau problem (S, ϕ) whenever its asymptotic Gauss image is equal to this Möbius surface.

[2] In fact, it is not necessary for the immersed surface to be infinitesimally strictly convex for its asymptotic Gauss map to be a local diffeomorphism. It is instead sufficient that both of its principal curvatures be different to -1. The properties of surfaces which satisfy this condition are studied in [8, 22].

In other words, Labourie's asymptotic Plateau problem concerns the unique prescription of k-surfaces in terms of their asymptotic Gauss images. In [19], Labourie proved various existence and uniqueness results for solutions of this problem in a more general setting than that studied here. Further existence and continuity results were also obtained by the author in [26]. There is now, scattered across the literature, a rich theory around the asymptotic Plateau problem, which we will review in [28].

In Sect. 3.4, we apply the theories of Möbius surfaces and hyperbolic ends to the study of this problem. In particular, using the a priori estimate of Theorem 3.1.13, we obtain the following new compactness result. First, we say that two developed Möbius surfaces (S, ϕ) and (S', ϕ') are *equivalent* whenever there exists a diffeomorphism $\alpha : S \to S'$ and a Möbius map $\beta \in \mathrm{PSL}(2, \mathbb{C})$ such that

$$\phi' \circ \alpha = \beta \circ \phi. \tag{3.1.16}$$

Theorem 3.1.17 (Monotone Convergence) *Let (S, ϕ) be a developed Möbius surface with universal cover not equivalent to $(\hat{\mathbb{C}}, z)$, $(\mathbb{C}, z)$ or $(\mathbb{C}, \mathrm{Exp}(z))$. Let $(\Omega_m)_{m\in\mathbb{N}}$ be a nested sequence of open subsets of S which exhausts S. If, for $k > 0$ and for all m, $e_m : \Omega_m \to \mathbb{H}^3$ is a k-surface solving the asymptotic Plateau problem $(\Omega_m, \phi|_{\Omega_m})$, then $(e_m)_{m\in\mathbb{N}}$ subconverges in the C^∞_{loc} sense over S to a k-surface $e_\infty : S \to \mathbb{H}^3$ solving the asymptotic Plateau problem (S, ϕ).*

Remark 3.1.18 Theorem 3.1.17 is proven in Theorem 3.4.11.

Upon combining Theorem 3.1.17 with the existence results proven by Labourie in [19], we obtain the main new result of this chapter: a complete solution of the asymptotic Plateau problem for k-surfaces in three-dimensional hyperbolic space.

Theorem 3.1.19 (Existence and Uniqueness) *For all $0 < k < 1$, and for every developed Möbius surface (S, ϕ) with universal cover not equivalent to $(\hat{\mathbb{C}}, z)$, $(\mathbb{C}, z)$ or $(\mathbb{C}, \mathrm{Exp}(z))$, there exists a unique k-surface $e : S \to \mathbb{H}^3$ solving the asymptotic Plateau problem (S, ϕ).*

Remark 3.1.20 Theorem 3.1.19 is proven in Theorem 3.4.16.

Remark 3.1.21 It is an interesting open problem to determine under what conditions a k-surface is complete. We describe an example of a non-complete k-surface in Appendix A.

3.1.3 Schwarzian Derivatives

We conclude this introduction by showing how a reformulation of Theorem 3.1.19 in terms of Schwarzian derivatives yields nice linearisations of the spaces of k-surfaces in $\mathbb{H}^3$.

Let S be a simply-connected Riemann surface. By Riemann's uniformisation theorem, we may suppose that S is the Poincaré disk $\mathbb{D}$, the complex plane $\mathbb{C}$ or the Riemann sphere $\hat{\mathbb{C}}$. For all k, let $\widetilde{\mathrm{Imm}}_k(S)$ denote the space of k-surfaces $e : S \rightarrow \mathbb{H}^3$ whose asymptotic Gauss map is holomorphic. We furnish this space with the C^∞_{loc} topology and we denote by $\mathrm{Imm}_k(S)$ its quotient under the action of post-composition by elements of $\mathrm{PSO}_0(3,1)$. Trivially, every simply connected k-surface is equivalent to a unique element of

$$\mathrm{Imm}_k(\mathbb{D}) \cup \mathrm{Imm}_k(\mathbb{C}) \cup \mathrm{Imm}_k(\hat{\mathbb{C}}). \tag{3.1.17}$$

The space $\mathrm{Imm}_k(\hat{\mathbb{C}})$ will be of little interest to us since, for $k > 1$, it consists of a single equivalence class corresponding to geodesic spheres of radius $\mathrm{arctanh}(1/\sqrt{k})$ whilst, for $k \leqslant 1$, it is empty.

We now show how $\mathrm{Imm}_k(\mathbb{D})$ and $\mathrm{Imm}_k(\mathbb{C})$ are parametrised by open subsets of vector spaces. We first recall the concept of Schwarzian derivative (see [20]). Recall that a function $\phi : S \rightarrow \hat{\mathbb{C}}$ is said to be *locally conformal* whenever it is a holomorphic local diffeomorphism. The *Schwarzian derivative* of any such function $\phi : S \rightarrow \hat{\mathbb{C}}$ is defined by

$$D^{\mathrm{Sch}}\phi := \left(\frac{\phi''}{\phi'}\right)' - \frac{1}{2}\left(\frac{\phi''}{\phi'}\right)^2. \tag{3.1.18}$$

A key property of the Schwarzian derivative is that, for any locally conformal function $\phi : S \rightarrow \hat{\mathbb{C}}$ and for any Möbius map α,

$$D^{\mathrm{Sch}}(\alpha \circ \phi) = D^{\mathrm{Sch}}\phi. \tag{3.1.19}$$

Furthermore, for any holomorphic function $f : S \rightarrow \mathbb{C}$, there exists a locally conformal function $\phi : S \rightarrow \hat{\mathbb{C}}$, unique up to post-composition by Möbius maps, such that

$$D^{\mathrm{Sch}}\phi = f. \tag{3.1.20}$$

Let $\mathrm{Hol}(S)$ denote the space of holomorphic functions over S furnished with the C^0_{loc} topology. For all $k > 0$, let $\tilde{\Sigma} : \widetilde{\mathrm{Imm}}_k(S) \rightarrow \mathrm{Hol}(S)$ denote the function defined such that, for every k-surface $e : S \rightarrow \mathbb{H}^3$,

$$\tilde{\Sigma}[e] := D^{\mathrm{Sch}}\phi_e, \tag{3.1.21}$$

where ϕ_e denotes the asymptotic Gauss map of e. For any k-surface $e \in \widetilde{\mathrm{Imm}}_k(S)$ and for any Möbius map α,

$$\tilde{\Sigma}_\infty[\alpha \circ e] = D^{\mathrm{Sch}}\phi_{\alpha\circ e} = D^{\mathrm{Sch}}(\alpha \circ \phi_e) = D^{\mathrm{Sch}}\phi_e = \tilde{\Sigma}_\infty[e], \tag{3.1.22}$$

so that, for all k, $\tilde{\Sigma}$ descends to a continuous functional $\Sigma : \mathrm{Imm}_k(S) \rightarrow \mathrm{Hol}(S)$.

In [26], we prove an existence and uniqueness result for solutions of asymptotic Plateau problems of hyperbolic conformal type. In the present framework, this is reformulated as follows.

Theorem 3.1.22 (Hyperbolic Asymptotic Plateau Problem) *For all* $0 < k < 1$ *and for all* $f \in \mathrm{Hol}(\mathbb{D})$*, there exists a unique element* $e \in \mathrm{Imm}_k(\mathbb{D})$ *such that*

$$\Sigma[e] = f. \tag{3.1.23}$$

Furthermore, e depends continuously on f*. In other words,* Σ *defines a homeomorphism from* $\mathrm{Hol}(\mathbb{D})$ *into* $\mathrm{Imm}_k(\mathbb{D})$.

Remark 3.1.23 Theorem 3.1.22 is proven in Theorem 3.4.18, below.

Theorem 3.1.19 now yields the corresponding result in the parabolic case.

Theorem 3.1.24 (Parabolic Asymptotic Plateau Problem) *For all* $0 < k < 1$ *and for all* $f \in \mathrm{Hol}(\mathbb{C}) \setminus \mathbb{C}$*, there exists a unique element* $e \in \mathrm{Imm}_k(\mathbb{C})$ *such that*

$$\Sigma[e] = f. \tag{3.1.24}$$

Remark 3.1.25 Theorem 3.1.24 is proven in Theorem 3.4.19, below.

Remark 3.1.26 It is not known in the parabolic case whether the solution e depends continuously on the data f.

Remark 3.1.27 Interestingly, a complementary result holds in the limiting case where $k = 1$. Indeed, by a theorem of Volkov–Vladimirova and Sasaki (see Theorem 46 of [29]), $\mathrm{Imm}_1(\mathbb{D})$ is empty and $\mathrm{Imm}_1(\mathbb{C})$ consists only of horospheres and universal covers of cylinders of constant radius about complete geodesics.[3] When e is a horosphere, $\Sigma[e]$ vanishes and when e is a universal cover of a cylinder, $\Sigma[e]$ is a non-zero constant. For this and other reasons, for $0 < k < 1$, it makes sense to identify complete geodesics and ideal points of $\partial_\infty \mathbb{H}^3$ as degenerate solutions of the asymptotic Plateau problem for $f \in \mathbb{C} \setminus \{0\}$ and $f = 0$ respectively.

Remark 3.1.28 For $k > 1$, we expect both $\mathrm{Imm}_k(\mathbb{D})$ and $\mathrm{Imm}_k(\mathbb{C})$ to be empty. However, we are not aware of any proof of this affirmation.

[3] In fact, Volkov–Vladimirova and Sasaki's result as stated in [29] requires completeness as opposed to mere quasicompleteness. However, as we will show in our forthcoming work [28], a careful analysis of the proof reveals that quasicompleteness is quite sufficient for this result to hold.

3.1.4 Closing Remarks and Acknowledgements

Much of this chapter has been formulated in the language of category theory, which we believe provides the best framework for presenting our results. For the benefit of those who, like the author, have always met this theory with a certain foreboding, we have provided an elementary discussion of its basic principles in Appendix B.

The author is grateful to Sébastien Alvarez, François Fillastre and Andrea Seppi for helpful comments on earlier drafts of this text. Figure 3.5 was prepared by Débora Mondaini.

3.2 Möbius Structures

3.2.1 Möbius Structures

A *Möbius structure* (also known as a *flat conformal structure* or a *complex projective structure*) over a surface S is an atlas $\mathcal{A}$ all of whose transition maps are restrictions of Möbius maps. A *Möbius surface* is a pair $(S, \mathcal{A})$ where S is a surface and $\mathcal{A}$ is a Möbius structure over this surface. In what follows, we will denote the Möbius surface simply by S whenever the atlas is clear from the context. The family of Möbius surfaces forms a category whose morphisms are those functions $\phi : X \to X'$ whose expressions with respect to every pair of coordinate charts are restrictions of Möbius maps. Naturally, we identify Möbius surfaces which are isomorphic.

Every Möbius structure trivially defines a holomorphic structure over the same surface. We call the resulting Riemann surface the *underlying Riemann surface* of the Möbius surface. The operation which associates to a Möbius surface its underlying Riemann surface is trivially a covariant functor. This distinction between Möbius surfaces and their underlying Riemann surfaces is more than a mere abstract formality, and the reader may consult, for example, [7] for an overview of the rich theory concerning the relationship between the two.

The model examples of Möbius surfaces are the open subsets of $\hat{\mathbb{C}}$ and their quotients under actions of subgroups of the Möbius group $\mathrm{PSL}(2, \mathbb{C})$. More generally, given any surface S, and a local diffeomorphism $\phi : S \to \hat{\mathbb{C}}$, a Möbius structure is constructed over S as follows. For every point $x \in S$, there exists a neighbourhood U of x over which ϕ restricts to a diffeomorphism onto its image V. The set $(U_\alpha, V_\alpha, \phi)_{\alpha\in A}$ of all such charts defines an atlas of S in $\hat{\mathbb{C}}$ whose transition maps are trivial, and thus a fortiori Möbius. We call this structure the *pull-back structure* of ϕ and we denote it by $\phi^*\hat{\mathbb{C}}$. It will often be convenient in the sequel to denote the Möbius surface defined by ϕ by (S, ϕ).

Given a Möbius surface S, we say that a local diffeomorphism $\phi : S \to \hat{\mathbb{C}}$ is a *developing map* of S whenever the pull-back Möbius structure of ϕ is compatible with the initial Möbius structure of S. Any two developing maps $\phi, \phi' : S \to \hat{\mathbb{C}}$ are related to one another by

$$\phi' = \alpha \circ \phi, \tag{3.2.1}$$

for some Möbius map α, so that the family of all developing maps over a given Möbius surface can be parametrised by $\mathrm{PSL}(2, \mathbb{C})$ whenever it is non-empty. We say that a Möbius surface is *developable* whenever it has a developing map. In particular, every simply connected Möbius surface has this property. In the following sections, we will mainly be concerned with developable Möbius surfaces. In particular, we will take the developing maps to be given, and we leave the reader to verify that our constructions are independent of the developing maps chosen.

Non-developable Möbius surfaces are studied as follows. Given a Möbius surface S with fundamental group Π and universal cover $\tilde{S}$, any developing map ϕ of $\tilde{S}$ is equivariant with respect to a unique homomorphism $\theta : \Pi \to \mathrm{PSL}(2, \mathbb{C})$ which we call its *holonomy*. Furthermore, given another developing map ϕ' with holonomy θ', there exists a unique Möbius map α such that

$$\theta' = \alpha\theta\alpha^{-1}, \text{ and} \tag{3.2.2}$$

$$\phi' = \alpha \circ \phi. \tag{3.2.3}$$

Although non-developable Möbius surfaces will be of little interest to us in the sequel, their study has produced a deep and fascinating literature. For example, the question of which homomorphisms arise as holonomies of Möbius surfaces is addressed thoroughly by Gallo–Kapovich–Marden in [10]. Likewise, the structure of the space of Möbius surfaces with a given fixed holonomy θ is studied by Goldman in [11]. Finally, branched Möbius structures, for which the developing map is allowed to be a ramified covering, add yet another layer of sophistication to this theory (see, for example, [4]).

We conclude this section by describing a key trichotomy of the theory. We say that a connected Möbius surface is *elliptic* or *parabolic* whenever its universal cover is isomorphic to $(\hat{\mathbb{C}}, z)$ or to $(\mathbb{C}, z)$ respectively and *hyperbolic* otherwise.

Lemma 3.2.1 *Let S be a connected Möbius surface. If S contains an elliptic surface, then S is elliptic. If S contains a parabolic surface, then S is either elliptic or parabolic.*

Proof Upon taking universal covers, we may suppose that S is simply connected. Let S' be an open subset of S. If S' is elliptic then, being compact, it is closed so that, by connectedness, $S = S'$ is also elliptic. Suppose now that S' is parabolic. Let $\phi : S \to \hat{\mathbb{C}}$ be a developing map such that $\phi(S') = \mathbb{C}$. We claim that S' is also simply connected. Indeed, let $\tilde{S}'$ denote its universal cover and let $\pi : \tilde{S}' \to S'$ denote the canonical projection. Since $(\phi \circ \pi)$ is a developing map of $\tilde{S}'$, it is a diffeomorphism from $\tilde{S}'$ onto $\mathbb{C}$. It follows that π is injective and S' is therefore simply connected, as asserted. In particular, ϕ restricts to a diffeomorphism from S' onto $\mathbb{C}$.

Suppose now that $S' \neq S$. In particular, the topological boundary $\partial S'$ of S' in S is non-empty. Since the restriction of ϕ to S' is a diffeomorphism, $\phi(\partial S') = \{\infty\}$. Now let x be a point of $\partial S'$. Let Ω be a connected neighbourhood of x in S over which ϕ restricts to a diffeomorphism. In particular, by injectivity, $\partial S' \cap \Omega = \{x\}$. It follows that $S' \cap (\Omega \setminus \{x\})$ is a non-trivial, open and closed subset of $\Omega \setminus \{x\}$ so that, by connectedness, $\Omega \setminus \{x\} \subseteq S'$. Since $\phi(S \setminus \Omega)$ is uniformly bounded away from ∞, x is in fact the only element of $\partial S'$. We conclude that ϕ defines a diffeomorphism from S onto $\hat{\mathbb{C}}$, so that S is elliptic. This completes the proof. □

We underline that the above trichotomy for Möbius surfaces differs from the elliptic-parabolic-hyperbolic trichotomy for Riemann surfaces. Indeed, although the underlying Riemann surface of any elliptic or parabolic Möbius surface is also respectively elliptic and parabolic, there exist many hyperbolic Möbius surfaces—such as, for example, $(\mathbb{C}^*, z)$, $(\mathbb{C}, e^z)$ and $(\mathbb{C}^*, e^z)$—whose underlying Riemann surfaces are parabolic.

3.2.2 The Möbius Disk Decomposition and the Join Relation

We now introduce a canonical decomposition of Möbius surfaces which will be the main tool used for their study in the sequel. Let S be a developable Möbius surface with developing map ϕ. A *Möbius disk* in S is a pair (D, α) where $D \subseteq \hat{\mathbb{C}}$ is an open disk and $\alpha : D \to S$ satisfies

$$\phi \circ \alpha = \mathrm{Id}. \tag{3.2.4}$$

We call the set $(D_i, \alpha_i)_{i \in I}$ of all Möbius disks in S its *Möbius disk decomposition*. Since ϕ is a local diffeomorphism, every point of S lies in the image of some Möbius disk, so that the Möbius disk decomposition covers S. We define the *join relation* $\sim$ of the Möbius disk decomposition such that, for all $i, j \in I$,

$$i \sim j \;\Leftrightarrow\; \alpha_i(D_i) \cap \alpha_j(D_j) \neq \emptyset. \tag{3.2.5}$$

This relation is trivially reflexive and symmetric, but not transitive. Composing with ϕ, we obtain

$$i \sim j \;\Rightarrow\; D_i \cap D_j \neq \emptyset, \tag{3.2.6}$$

and

$$i \sim j,\; j \sim k,\; D_i \cap D_j \cap D_k \neq \emptyset \;\Rightarrow i \sim k. \tag{3.2.7}$$

We call the pair $((D_i)_{i \in I}, \sim)$ the *combinatorial data* of S. This data is sufficient to recover S uniquely up to isomorphism, as follows from the following general result.

Theorem & Definition 3.2.2 *Let M be a smooth manifold. Let $(\Omega_i)_{i\in I}$ be a family of open subsets of M and let $\sim$ be a reflexive and symmetric relation over I such that*

(1) for all $i, j \in I$, $\Omega_i \cap \Omega_j$ has at most 1 connected component;
(2) $i \sim j \Rightarrow \Omega_i \cap \Omega_j \neq \emptyset$; and
(3) $i \sim j,\ j \sim k,\ \Omega_i \cap \Omega_j \cap \Omega_k \neq \emptyset \Rightarrow i \sim k$.

There exists a (not necessarily second-countable) smooth manifold N, a smooth local diffeomorphism $\phi : N \to M$ and, for all i, a smooth function $\alpha_i : \Omega_i \to N$ such that,

(A) $(\alpha_i(\Omega_i))_{i\in I}$ covers N;
(B) $i \sim j \Leftrightarrow \alpha_i(D_i) \cap \alpha_j(D_j) \neq \emptyset$; and
(C) for all i, $\phi \circ \alpha_i = \mathrm{Id}$.

Furthermore, the triplet $(N, \phi, (\alpha_i)_{i\in I})$ is unique in the sense that if $(N', \phi', (\alpha'_i)_{i\in I})$ is another such triplet, then there exists a unique diffeomorphism $\psi : N \to N'$ such that, for all i, $\alpha'_i = \psi \circ \alpha_i$.

We call N the join of $((\Omega_i)_{i\in I}, \sim)$, we call ϕ the canonical immersion and we call $(\alpha_i)_{i\in I}$ the canonical parametrisations.

Remark 3.2.3 If M possesses any additional structure—such as, say, a hyperbolic structure, a Möbius structure, and so on—then N inherits this structure from M, as follows immediately from the triviality of the transition maps of the atlas constructed in the proof below.

Remark 3.2.4 We do not prove second-countability of N. This will not trouble us, however, since second-countability is only required in manifold theory for constructions involving either Sard's Theorem or partitions of unity, neither of which appear in this chapter. Besides, in every case arising in the sequel, second-countability can be recovered, either by covering N by a countable subfamily of $(\alpha_i(\Omega_i))_{i\in I}$, or by appealing to Radó's Theorem (see [14]).

Proof We first prove existence. Define

$$\tilde{N} := \sqcup_{i\in I}\Omega_i,$$

and define the relation $\approx$ over $\tilde{N}$ such that, for all $x_i \in \Omega_i$ and $y_j \in \Omega_j$,

$$x_i \approx y_j \ \Leftrightarrow\ i \sim j \text{ and } x_i = y_j.$$

It follows by (3) that $\approx$ is an equivalence relation over $\tilde{N}$. Let $N := \tilde{N}/\approx$ denote its quotient space furnished with the quotient topology and let $\alpha : \tilde{N} \to N$ denote the canonical projection. Recall now that a manifold is defined to be a second-countable, Hausdorff space furnished with an atlas. The atlas of N is constructed as follows. For all i, we verify that α restricts to a homeomorphism from Ω_i onto an open subset of N, and we denote

$$U_i := \alpha(\Omega_i),\ \ V_i := \Omega_i,\ \ \alpha_i := \alpha|_{V_i},\ \text{ and } \phi_i := \alpha_i^{-1}.$$

The family $\mathcal{A} := (U_i, V_i, \phi_i)_{i\in I}$ forms an atlas of N all of whose transition maps are trivial, and thus a fortiori smooth, as desired.

Since we are not concerned with second-countability, it only remains to show that N is Hausdorff. For this, let $x_i \in \Omega_i$ and $y_j \in \Omega_j$ be such that there exists a sequence $(p_m)_{m\in\mathbb{N}}$ of points in N converging simultaneously to $\alpha(x_i)$ and to $\alpha(y_j)$. For sufficiently large m, p_m has representative elements $x_{m,i}$ in Ω_i and $y_{m,j}$ in Ω_j respectively, which converge towards x_i and y_j respectively. In particular, $i \sim j$ and, for all m, $x_{m,i} = y_{m,j}$. Upon taking limits, we obtain $x_i = y_j$, so that $x_i \approx y_j$ and therefore $\alpha(x_i) = \alpha(y_j)$. We conclude that N is indeed Hausdorff, and therefore a (not necessarily second-countable) manifold.

Finally, the canonical inclusion $\tilde{\phi} : \tilde{N} \to M$ trivially descends to a local diffeomorphism $\phi : N \to M$. We verify that $(N, \phi, (\alpha_i)_{i\in I})$ has the desired properties, thus proving existence.

To prove uniqueness, let $(N', \phi', (\alpha_i')_{i\in I})$ be another such triplet. Define $\tilde{\psi} : \tilde{N} \to N'$ such that, for all $x_i \in \Omega_i$,

$$\tilde{\psi}(x_i) := \alpha_i'(x_i).$$

We first show that $\tilde{\psi}$ descends to a function $\psi : N \to N'$. Indeed, let $x_i \in \Omega_i$ and $y_j \in \Omega_j$ be such that $x_i \approx y_j$. By (B),

$$\alpha_i'(\Omega_i) \cap \alpha_j'(\Omega_j) \neq \emptyset.$$

Furthermore, by (1), (C) and a connectedness argument

$$\alpha_i'|_{\Omega_i\cap\Omega_j} = \alpha_j'|_{\Omega_i\cap\Omega_j}.$$

In particular, $\tilde{\psi}(x_i) = \tilde{\psi}(y_j)$ so that $\tilde{\psi}$ indeed descends to a function $\psi : N \to N'$. By (A), ψ is surjective, by (B) and (C), it is injective. Since α_i and α_i' are local diffeomorphisms for all i, it follows that ψ is a diffeomorphism, definition, for all i, $\alpha_i' = \psi \circ \alpha_i$. This proves existence of ψ, and since uniqueness is trivial, this completes the proof. □

3.2.3 Geodesic Arcs and Convexity

We now introduce a concept of geodesics for sets of Möbius disks in a given non-elliptic Möbius surface. This in turn yields a concept of convexity for such sets which will be useful for establishing uniqueness in the constructions that follow.

To begin with, we study the geometry of the space $\mathcal{D}$ of disks in $\hat{\mathbb{C}}$. Recall that $\hat{\mathbb{C}}$ naturally identifies with the ideal boundary $\partial_\infty\mathbb{H}^3$ of $\mathbb{H}^3$. With this identification, every disk D in $\hat{\mathbb{C}}$ is the ideal boundary of a unique open half-space H in $\mathbb{H}^3$. The boundary ∂H of every open half-space in $\mathbb{H}^3$ is a totally geodesic plane which we orient so that its positively oriented normal points outwards from H. Trivially, open half-spaces in $\mathbb{H}^3$ are uniquely defined by their oriented boundaries. Consequently, any parametrisation of the space of oriented totally geodesic planes in $\mathbb{H}^3$ is also a parametrisation of $\mathcal{D}$.

The space of oriented totally geodesic planes in $\mathbb{H}^3$ is parametrised by (2, 1)-dimensional de Sitter space $\mathrm{dS}^{2,1}$ as follows. First, we identify $\mathbb{H}^3$ and $\mathrm{dS}^{2,1}$ with subsets of $\mathbb{R}^{3,1}$, namely

$$\mathbb{H}^3 := \left\{x \in \mathbb{R}^{3,1} \mid \langle x, x\rangle_{3,1} = -1,\ x_4 > 0\right\},\ \text{and} \tag{3.2.8}$$

$$\mathrm{dS}^{2,1} := \left\{x \in \mathbb{R}^{3,1} \mid \langle x, x\rangle_{3,1} = 1\right\}, \tag{3.2.9}$$

where here $\langle\cdot,\cdot\rangle_{3,1}$ denotes the *Minkowski metric* with signature (3, 1), that is

$$\langle x, x\rangle_{3,1} := x_1^2 + x_2^2 + x_3^2 - x_4^2. \tag{3.2.10}$$

With this identification, every oriented totally geodesic plane P in $\mathbb{H}^3$ is the intersection of $\mathbb{H}^3$ with a unique oriented, time-like, linear hyperplane $\hat{P}$ in $\mathbb{R}^{3,1}$. Every such hyperplane has, in turn, a well-defined positively-oriented unit normal vector N. Since N is also spacelike, it is an element of $\mathrm{dS}^{2,1}$. This yields a bijection between the space of oriented, totally geodesic planes in $\mathbb{H}^3$ and $\mathrm{dS}^{2,1}$, which is the desired parametrisation.

Recall now that a subset Γ of $\mathrm{dS}^{2,1}$ is a geodesic if and only if it is the intersection of $\mathrm{dS}^{2,1}$ with a linear plane $\hat{\Gamma}$. Furthermore, Γ is said to be *spacelike*, *lightlike* or *timelike* respectively whenever the restriction to this plane of the Minkowski metric has signature $(2, 0)$, $(1, 0)$ or $(1, 1)$. Of particular interest to us will be the spacelike geodesics. Observe first that any two distinct totally geodesic planes in $\mathbb{H}^3$ with non-trivial intersection meet along a complete geodesic.

Lemma 3.2.5 *Let P and P' be distinct, oriented totally-geodesic planes in $\mathbb{H}^3$ which are neither equal nor equal with opposing orientations. P and P' have non-trivial intersection if and only if their corresponding points in $\mathrm{dS}^{2,1}$ lie along a common spacelike geodesic Γ. Furthermore, motion at constant speed along Γ corresponds to rotation at constant angular speed around their common geodesic G.*

Proof Observe first that the orthogonal complement in $\mathbb{R}^{3,1}$ of any timelike linear plane $\hat{G}$ is a spacelike linear plane $\hat{\Gamma}$ whose intersection with $\mathrm{dS}^{2,1}$ is a circle in $\hat{\Gamma}$ and a spacelike geodesic Γ in $\mathrm{dS}^{2,1}$. Now let $P = \hat{P} \cap \mathbb{H}^3$ and $P' = \hat{P}' \cap \mathbb{H}^3$ be oriented totally-geodesic planes in $\mathbb{H}^3$ which are neither equal nor equal with opposing orientations. These planes meet along a common geodesic G if and only if $\hat{P}$ and $\hat{P}'$ contain a common timelike linear plane $\hat{G}$. This in turn holds if and only if their unit normals N and N', which are already elements of $\mathrm{dS}^{2,1}$, are also elements of the orthogonal complement $\hat{\Gamma}$ of $\hat{G}$. This proves the first assertion. Since the second assertion is straightforward, this completes the proof. □

We now return to the case of disks in $\hat{\mathbb{C}}$. We say that two distinct disks D_0 and D_1 *overlap* whenever their boundary circles meet at exactly two points. Observe that this holds if and only if their intersection is non-trivial, the intersection of their complements is non-trivial, and neither is contained within the other. With the preceding parametrisation, this is precisely the requirement for their corresponding points in $\mathrm{dS}^{2,1}$ to lie along a common spacelike geodesic. In addition, the corresponding point of a third disk D lies along the *shorter* geodesic arc between these two points if and only if

$$D_0 \cap D_1 \subseteq D \subseteq D_0 \cup D_1. \tag{3.2.11}$$

We thus define the *geodesic arc* between two overlapping disks D_0 and D_1 to be the set of all disks D in $\hat{\mathbb{C}}$ which satisfy this property. This construction is illustrated in Fig. 3.1.

This concept of geodesic arc extends to the Möbius disk decomposition of S as follows. We say that two distinct Möbius disks (D_0, α_0) and (D_1, α_1) *overlap* whenever $\alpha_0(D_0)$ and $\alpha_1(D_1)$ have non-trivial intersection and neither is contained within the other. Upon composing with ϕ, it follows that D_0 and D_1 likewise have non-trivial intersection and neither is contained within the other. In addition, since S is not elliptic, the complements of D_0 and D_1 also have non-trivial intersection, so that D_0 and D_1 also overlap. Using a connectedness argument, we show that

$$\alpha_0|_{D_0\cap D_1} = \alpha_1|_{D_0\cap D_1}, \tag{3.2.12}$$

so that these functions join to define a function $\alpha_{01} : D_0 \cup D_1 \to S$ such that

$$\phi \circ \alpha_{01} = \mathrm{Id}. \tag{3.2.13}$$

In particular, for any other disk D along the geodesic arc from D_0 to D_1, $(D, \alpha_{01}|_D)$ is also a Möbius disk in S. We thus define the *geodesic arc* from (D_0, α_0) to (D_1, α_1)

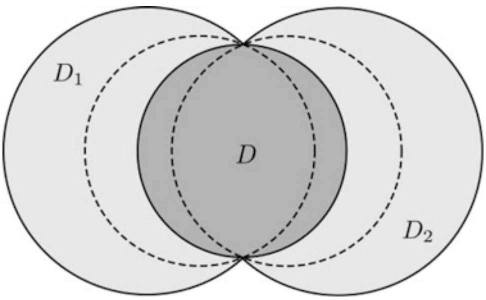


Fig. 3.1 Here the disks D_1 and D_2 overlap. The disk D is the mid-point of the geodesic arc between these two disks. Two other points along this arc are marked by dashed curves

to be the set of all Möbius disks in S of this form. We say that any subset $(D_i, \alpha_i)_{i\in J}$ of the Möbius disk decomposition of S is *convex* whenever it contains the geodesic arc between any two of its overlapping disks.

3.2.4 The Kulkarni–Pinkall Form

In [17], Kulkarni–Pinkall construct for any Möbius surface of hyperbolic type a canonical metric which encodes its global geometry in a local manner. Kulkarni–Pinkall's construction will play a central role in the C^0 estimates that we will derive in Sect. 3.4 for quasicomplete ISC immersions in $\mathbb{H}^3$. However, we will adopt here a slightly different perspective from that of [17], since we believe it to be more natural to work in terms of 2-forms rather than in terms of metrics.

Let S be a developable Möbius surface with developing map ϕ, let $(D_i, \alpha_i)_{i\in I}$ denote its Möbius disk decomposition and, for all $i \in I$, let H_i denote the open half-space in $\mathbb{H}^3$ with ideal boundary D_i. For all $x \in S$, let $I(x)$ denote the set of indices i such that $x \in \alpha_i(D_i)$. For any disk $D \in \hat{\mathbb{C}}$, let $\omega(D)$ denote the area form of its Poincaré metric. We define ω_ϕ, the *Kulkarni–Pinkall form* of S, such that, for all $x \in S$,

$$\omega_\phi(x) := \mathrm{inf}_{i\in I(x)}\phi^*\omega(D_i), \tag{3.2.14}$$

and we define g_ϕ the *Kulkarni–Pinkall metric* of S by

$$g_\phi := \omega_\phi(\cdot, J\cdot), \tag{3.2.15}$$

where J here denotes the complex structure of S.

Lemma 3.2.6 (Monotonicity) *Let S and S' be developable Möbius surfaces with respective developing maps ϕ and ϕ' and respective Kulkarni–Pinkall forms ω_ϕ and $\omega_{\phi'}$. If $\alpha : S \to S'$ is a morphism such that $\phi = \phi' \circ \alpha$, then*

$$\omega_\phi \geq \alpha^*\omega_{\phi'}. \tag{3.2.16}$$

Proof Indeed, composition with α sends the Möbius disk decomposition of S into the Möbius disk decomposition of S'. □

The following family of partial orders over I will prove useful in deriving properties of the Kulkarni–Pinkall form. For all $x \in S$, we define

$$i \geq_x j \;\Leftrightarrow\; i, j \in I(x) \text{ and } \omega(D_i)(y) \leq \omega(D_j)(y), \tag{3.2.17}$$

where $y := \phi(x)$. The geometric significance of the Kulkarni–Pinkall form as well as this partial order becomes clear once we recall the parametrisation of the space of open horoballs in $\mathbb{H}^3$ by $\Lambda^2\partial_\infty\mathbb{H}^3$ described in Sect. 3.1.2. Indeed, for all $x \in$

S, $\phi_*\omega_\phi(x)$ is simply the infimal asymptotic curvature of horoballs asymptotically centred on $\phi(x)$ and contained in H_i, as i varies over $I(x)$. Likewise, for all $x \in S$ and for all $i, j \in I(x)$, $i \geq_x j$ if and only if every open horoball asymptotically centred on $\phi(x)$ and contained in H_j is also contained in H_i.

3.2.5 Analytic Properties of the Kulkarni–Pinkall Form

We restrict our attention initially to the simpler case of Möbius surfaces of the form (Ω, z), where Ω is an open subset of $\hat{\mathbb{C}}$. At this stage, it is useful to recall that, for a disk D in the complex plane $\mathbb{C}$ of radius R with centre lying at distance $r < R$ from the origin,

$$\omega(D)(0) = \frac{4R^2 dxdy}{(R-r)^2(R+r)^2} \geqslant \frac{dxdy}{(R-r)^2}. \tag{3.2.18}$$

In particular, if $\omega(D)(0) < \lambda^2 dxdy$, then D contains a disk of radius $1/\lambda$ about the origin.

Lemma & Definition 3.2.7 *Let Ω be an open subset of $\hat{\mathbb{C}}$ and let ω denote its Kulkarni–Pinkall form.*

(1) If the complement of Ω in $\hat{\mathbb{C}}$ contains at most 1 *point then, for all x, $\omega(x) = 0$ and $I(x)$ contains no maximal element with respect to $\geq_x$.*
(2) If the complement of Ω in $\hat{\mathbb{C}}$ contains at least 2 *distinct points then, for all x, $\omega(x) > 0$ and $I(x)$ contains a unique maximal element with respect to $\geq_x$ which realizes $\omega(x)$.*

In the second case, we denote by $\max(x)$ *the unique maximal element of $I(x)$.*

Proof The first assertion is trivial. To prove the second assertion, we may suppose that Ω is a proper subset of the complex plane $\mathbb{C}$. Existence follows by compactness of the set of (possibly ideal) disks in $\mathbb{C}$ which have radius bounded below, which contain a fixed point z_0, and which avoid another fixed point w_0. Observe now that $\omega(D_i)(x)$ restricts to a strictly concave function over every geodesic arc in I(x). Uniqueness thus follows by convexity of $I(x)$. Finally, since $\omega(x)$ is realized by the unique maximal element of $I(x)$, $\omega(x) > 0$, and this completes the proof. □

Given an ideal point $x \in \partial_\infty\mathbb{H}^3$ and a closed subset $Y \subseteq \mathbb{H}^3 \cup \partial_\infty\mathbb{H}^3$, we define the *curvature of distance* $c(x, Y)$ from x to Y to be the infimal asymptotic curvature of open horoballs with asymptotic centre x which do not meet Y.

Lemma 3.2.8 *Let Ω be a proper open subset of the complex plane $\mathbb{C}$, let ω denote its Kulkarni–Pinkall form, let $(D_i, \alpha_i)_{i\in I}$ denote its Möbius disk decomposition, and, for all $i \in I$, let H_i denote the open half-space in $\mathbb{H}^3$ with ideal boundary D_i. Let K denote the convex hull in $\mathbb{H}^3 \cup \partial_\infty\mathbb{H}^3$ of the complement of Ω and let*

$\pi : \Omega \to \partial K$ denote the closest point projection. For all $x \in \Omega$,

$$\omega(x) = c(x, K), \tag{3.2.19}$$

and $H(x) := H_{\max(x)}$ is the unique supporting open half-space of K at the point $\pi(x)$ such that $\partial H(x)$ is orthogonal to the geodesic joining $\pi(x)$ to x. In particular, $\omega(x)$, $H(x)$ and $D(x) := D_{\max(x)}$ are $C^{0,1}_{\text{loc}}$ functions over Ω.

Remark 3.2.9 In fact, Kulkarni–Pinkall show in [17] that ω is a $C^{1,1}_{\text{loc}}$ function.

Proof Since the complement of K in $\mathbb{H}^3$ is the union of all open half-spaces with ideal boundary in Ω, we have

$$K^c = \cup_{i \in I} H_i,$$

from which it follows that $\omega(x) = c(x, K)$ for all $x \in \Omega$. Now choose $x \in \Omega$. Let B be the open horoball in $\mathbb{H}^3$ with asymptotic centre x and asymptotic curvature $c(x, K)$. Since $H_{\max(x)}$ is the unique open half-space in K^c containing B, the second assertion follows and this completes the proof. □

We now address the general case. Let S be a developable Möbius surface with developing map ϕ and let $(D_i, \alpha_i)_{i \in I}$ denote its Möbius disk decomposition. For all $x \in S$, with $I(x)$ defined as in Sect. 3.2.4, we define

$$\Omega_x := \cup_{i \in I(x)} D_i. \tag{3.2.20}$$

For all $i, j \in I(x)$, α_i coincides with α_j over $D_i \cap D_j$ so that the join of these functions yields a function $\alpha_x : \Omega_x \to S$ satisfying $\phi \circ \alpha_x = \text{Id}$. We call (Ω_x, α_x) the *localisation* of S at x. The following trichotomy follows immediately from Lemma 3.2.1.

Lemma 3.2.10 *Let S be a developable Möbius surface with developing map $\phi : S \to \hat{\mathbb{C}}$.*

(1) If S is elliptic, then $\Omega_x = \hat{\mathbb{C}}$ for all x.
(2) If S is parabolic, then Ω_x is the complement of a single point in $\hat{\mathbb{C}}$ for all x.
(3) If S is hyperbolic, then the complement of Ω_x contains at least two points in $\hat{\mathbb{C}}$ for all x.

For all $x \in S$, we define $\omega_{\phi,x}$, the *local Kulkarni–Pinkall form* of S at x, to be the push-forward through α_x of the Kulkarni–Pinkall form of (Ω_x, z). Since composition with α_x sends the Möbius disk decomposition of (Ω_x, z) to $I(x)$, Lemmas 3.2.7 and 3.2.10 immediately yield the following result.

Lemma & Definition 3.2.11 *Let S be a developable Möbius surface of hyperbolic type with developing map $\phi : S \to \hat{\mathbb{C}}$ and Kulkarni–Pinkall form ω_ϕ. For all $x \in S$,* $\mathrm{I}(x)$ *has a unique maximal element which realises $\omega_\phi(x)$. Furthermore*

$$\omega_\phi(x) = \omega_{\phi,x}(x), \tag{3.2.21}$$

and, for all $y \in \alpha_x(\Omega_x)$,

$$\omega_\phi(y) \leq \omega_{\phi,x}(y). \tag{3.2.22}$$

For all $x \in S$*, we denote by* $\max(x)$ *the unique maximal element of* $I(x)$.

Analytic properties of ω_ϕ analogous to those obtained in Lemma 3.2.8 for localised Möbius structures follow upon refining (3.2.21) to equality over a neighbourhood of x.

Lemma 3.2.12 *Let* S *be a developable Möbius surface of hyperbolic type with developing map* ϕ*. For all* $x \in S$*, there exists a neighbourhood* U_x *of* x *such that, for all* $y \in U_x$,

$$\max(y) \in I(x). \tag{3.2.23}$$

Proof Since S is hyperbolic, we may suppose that Ω_x is a proper subset of $\mathbb{C}$. Let $(D_i, \alpha_i)_{i\in I}$ denote the Möbius disk decomposition of S. Denote $i := \max(x)$. We may suppose that D_i is the upper half-space in $\mathbb{C}$ and that $\phi(x) = \sqrt{-1}$. Let d_h denote the hyperbolic distance in D_i and define U_x by

$$U_x := \left\{ y \in \alpha_i(D_i) \mid d_h(\phi(y), \phi(x)) < \log((1+\sqrt{5})/2) \right\}.$$

Let y be an element of U_x. Observe that $\phi(y)$ is contained in the Euclidean ball of radius $(\sqrt{5}-1)/2$ about $\phi(x)$ in $\mathbb{C}$ (see Fig. 3.2). Denote $j := \max(y)$. Since S is hyperbolic, ∂D_i intersects ∂D_j at least one point and, upon applying a suitable Möbius transformation, we may suppose that one of these points lies at infinity. In

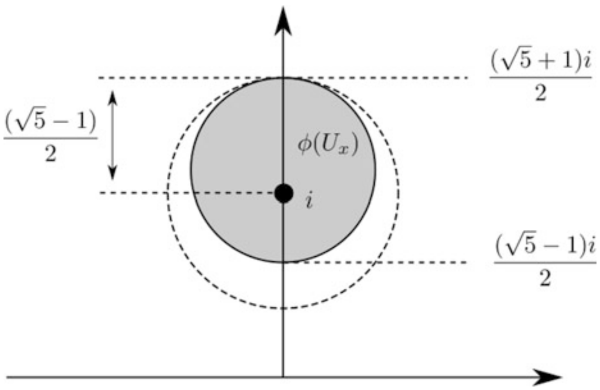


Fig. 3.2 The image of U_x is a disk in the upper half space, symmetric about the imaginary axis and passing through the points $(\sqrt{5}+1)i/2$ and $(\sqrt{5}-1)i/2$. In particular, it is contained in the Euclidean ball of radius $(\sqrt{5}-1)/2$ about i

particular D_j is a disk in $\mathbb{C}$. However, by (3.2.18) and (3.2.22),

$$\omega(D_j)(\phi(y)) = \phi_*\omega_\phi(y) \leq \phi_*\omega_{\phi,x}(y) \leq \frac{4dxdy}{(\sqrt{5}-1)^2}.$$

It follows by (3.2.18) again that D_j contains a ball of radius $(\sqrt{5}-1)/2$ about $\phi(y)$. In particular, $\phi(x) \in D_j$, so that $x \in \alpha_j(D_j)$ and $j \in I(x)$, as desired. □

Corollary 3.2.13 *Let S be a developable Möbius surface of hyperbolic type with developing map $\phi : S \to \hat{\mathbb{C}}$ and Kulkarni–Pinkall form ω_ϕ. Let x be a point of S. With U_x as in Lemma 3.2.12, for all $y \in U_x$,*

$$\omega_\phi(y) = \omega_{x,\phi}(y). \tag{3.2.24}$$

Combining the above results yields a description of the analytic properties of the Kulkarni–Pinkall form of every Möbius surface.

Theorem 3.2.14 *Let S be a developable Möbius surface with developing map $\phi : S \to \hat{\mathbb{C}}$, Kulkarni–Pinkall form ω_ϕ and Möbius disk decomposition $(D_i, \alpha_i)_{i\in I}$.*

(1) If S is of elliptic or parabolic type, then ω_ϕ vanishes identically.
(2) If S is of hyperbolic type, then ω_ϕ is a nowhere vanishing section of $\Lambda^2 S$.

Furthermore $\omega_\phi(x)$ and $D(x) := D_{\max(x)}$ are $C^{0,1}_{\text{loc}}$ functions over S.

Finally, Lemma 3.2.12 also shows that the Kulkarni–Pinkall metric of any Möbius surface of hyperbolic type is everywhere non-degenerate. In addition, we also obtain the following global information concerning this metric.

Lemma 3.2.15 *Let S be a developable Möbius surface with developing map ϕ. The Kulkarni–Pinkall metric g_ϕ of S is complete.*

Proof It suffices to show that there exists $r_0 > 0$ such that the closed ball of radius r_0 with respect to g_ϕ about any point of S is compact. Let $(D_i, \alpha_i)_{i\in I}$ denote the Möbius disk decomposition of S and for all i, let H_i denote the open half-space in $\mathbb{H}^3$ with ideal boundary D_i. We consider first the case where $(S, \phi) = (\Omega, z)$ for some connected neighbourhood Ω of 0 in $\mathbb{C}$. We identify $\mathbb{H}^3$ with the upper half-space in $\mathbb{C} \times \mathbb{R}$. Let K denote the convex hull in $\mathbb{H}^3$ of the complement of Ω in $\hat{\mathbb{C}}$. We may suppose that $D := D_{\max(0)}$ is the unit disk about the origin so that $(0, 1)^t$ is a boundary point of K. In particular, for all $j \in I$, $(0, 1)^t \notin H_j$. However, a symmetry argument shows that if ω^{Sph} denotes the standard spherical area form of $\hat{\mathbb{C}}$ then, for all $z \in \hat{\mathbb{C}}$,

$$\omega^{\text{Sph}}(z) = \inf_{j\in J(z)}\omega(D_j),$$

where $J(z)$ is the set of all indices j such that $z \in D_j$ but $(0, 1)^t \notin H_j$. It follows that

$$g_\phi \geq g^{\text{Sph}} := \frac{4}{(1+|z|^2)^2}\delta_{ij}.$$

Consider now the general case. Let x be a point of S. Let (Ω_x, α_x) denote the localisation of S at this point. As before, we may suppose that $\phi(x) = 0$ and that $D := D_{\max(x)}$ is the unit disk in $\mathbb{C}$ about the origin. Let U_x be as in Lemma 3.2.12. Recall now that the hyperbolic distance in D is given by

$$d^{\mathrm{Hyp}}(z, 0) = 2\mathrm{arctanh}(|z|).$$

From this we verify that $\phi(U_x)$ coincides with the Euclidean disk of radius $(\sqrt{5}-2)$. However, by the preceding paragraph, over this disk,

$$\phi_* g_\phi \geq g^{\mathrm{Sph}},$$

so that U_x contains the open disk of radius $\arcsin((\sqrt{5}+2)/10)$ with respect to g_ϕ about x. The result now follows with r_0 equal to this radius. □

3.3 Hyperbolic Ends

3.3.1 Hyperbolic Ends

Given a hyperbolic manifold X, we define a *height function* over X to be a strictly convex $C^{1,1}_{\mathrm{loc}}$ function $h : X \to]0, \infty[$ such that

(1) the gradient flow lines of h are unit speed geodesics; and
(2) for all $t > 0$, $h^{-1}([t, \infty[)$ is complete.

We will see in Lemma 3.3.5, below, that height functions, whenever they exist, are unique. We define a *hyperbolic end* to be a hyperbolic manifold which carries a height function. The family of hyperbolic ends forms a category whose morphisms are those functions $\psi : X \to X'$ which are local isometries. Naturally, we identify hyperbolic ends which are isometric.

We first identify various components of hyperbolic ends. Let X be a hyperbolic end with height function h. We call the gradient flow lines of h *vertical lines*. These curves form a geodesic foliation of X which we denote by $\mathcal{V}$ and which we call its *vertical line foliation*. We call the level sets of h the *levels* of X. These form another foliation of X by $C^{1,1}_{\mathrm{loc}}$ embedded surfaces which we call its *level set foliation* and which we denote by $(X_t)_{t>0}$. These two foliations are transverse to one another and every vertical line intersects every level at exactly one point. From this it follows that every level of X is naturally homeomorphic to the leaf space of $\mathcal{V}$. For all $t > 0$, we define the *vertical projection* $\pi_t : X \to X_t$ to be the function which sends each point x of X to the intersection with X_t of the vertical line on which it lies. By standard properties of convex sets in $\mathbb{H}^3$, for all $t > 0$, π_t restricts to a 1-Lipschitz function from $h^{-1}([t, \infty[)$ into X_t.

We call any local isometry $\phi : X \to \mathbb{H}^3$ a *developing map* of X. Any two developing maps $\phi, \phi' : X \to \mathbb{H}^3$ are related by

$$\phi' = \alpha \circ \phi, \tag{3.3.1}$$

for some isometry α of $\mathbb{H}^3$. We say that X is *developable* whenever it has a developing map. In particular, every simply connected hyperbolic end has this property. In the following sections, we will only be concerned with developable hyperbolic ends and we leave the reader to formulate the trivial extensions of our results to the general case. In particular, we will take the developing maps to be given, and we leave the reader to verify that our constructions are independent of the developing maps chosen.

The model examples of hyperbolic ends are the complements Ω of closed, convex subsets K of $\mathbb{H}^3$, where the height function is the distance to K. More sophisticated examples are given by quotients of such subsets by subgroups of the Möbius group $\mathrm{PSL}(2, \mathbb{C})$, such as the ends of quasi-Fuchsian manifolds studied in the introduction. We recall in addition that the complement of the Nielsen kernel of every finite geometry hyperbolic manifold is the union of finitely many hyperbolic ends (see, for example, [15]). However, we emphasize again that it is straightforward to construct hyperbolic ends that do not arise in this manner. Indeed, the developing map of the universal cover of any end of any finite geometry hyperbolic manifold with fundamental group not equal to $\mathbb{Z}$ is an embedding in $\mathbb{H}^3$. However, as we will see in Sect. 3.3.5, it is straightforward to construct simply connected hyperbolic ends with non-injective developing maps.

The key to understanding hyperbolic ends lies in the following analogue of the Hopf–Rinow Theorem.

Theorem 3.3.1 *Let* (X, h) *be a hyperbolic end. If* $\gamma : [0, a[\to X$ *is a geodesic segment such that* $\dot{\gamma}(0)$ *is not downward-pointing, then* γ *extends to a geodesic ray defined over the entire half-line* $[0, \infty[$.

Remark 3.3.2 A suitably modified version of Theorem 3.3.1 holds under the weaker condition that there exists a convex function $f : X \to]0, \infty[$ such that $f^{-1}([t, \infty[)$ is complete for all $t > 0$. In fact, using the arguments of the following sections, we may show that a hyperbolic manifold X is a hyperbolic end whenever there exists a $C^{1,1}_{\mathrm{loc}}$ convex function $f : X \to]0, \infty[$ such that $f^{-1}([t, \infty[)$ is complete for all $t > 0$ and $\|\nabla f\| \geq \epsilon > 0$. Such functions, which one may call *generalised height functions* are thus natural objects of study in the theory of hyperbolic ends (c.f. [1]).

Proof By strict convexity of h, $(h \circ \gamma)$ has strictly increasing derivative. Since, by hypothesis, $(h \circ \gamma)$ has non-negative derivative at 0, it follows that its derivative is strictly positive for all positive time, so that $(h \circ \gamma)$ is itself strictly increasing. In particular, γ remains within a complete subset of X and may thus be extended indefinitely, as desired. □

3.3.2 The Half-Space Decomposition

Let X be a developable hyperbolic end with height function h and developing map $\phi : X \to \mathbb{H}^3$. We define a *half-space* in X to be a pair (H, α) where H is an open half-space in $\mathbb{H}^3$ and $\alpha : H \to X$ satisfies

$$\phi \circ \alpha = \mathrm{Id}. \tag{3.3.2}$$

We call the set $(H_i, \alpha_i)_{i \in I}$ of all half-spaces in X its *half-space decomposition*.

Lemma 3.3.3 *Let X be a developable hyperbolic end with height function h. For all $x \in X$, there exists a unique half-space (H, α) in X such that $x \in \partial\alpha(H)$ and $\nabla h(x)$ is the inward-pointing normal to $\partial\alpha(H)$ at this point.*

Proof Let ϕ denote a developing map of X. Let x be a point of X. Define the subset E^+ of $T_x X$ by $E^+ := \{\xi \mid \langle \xi, \nabla h(x) \rangle > 0\}$. By Theorem 3.3.1, E^+ lies within the domain of the exponential map Exp_x of X at x. By Hadamard's theorem, the composition $(\phi \circ \mathrm{Exp}_x)$ restricts to a diffeomorphism from E^+ onto its image $H := (\phi \circ \mathrm{Exp}_x)(E^+)$. This image is an open half-space in $\mathbb{H}^3$ and the function $\alpha := \mathrm{Exp}_x \circ (\phi \circ \mathrm{Exp}_x)^{-1}$ is the desired right-inverse of ϕ. We readily verify that (H, α) is the desired half-space and that it is unique. This completes the proof. □

Corollary 3.3.4 *Let X be a developable hyperbolic end. The half-space decomposition of X covers X.*

Proof Indeed, for all $x \in X$, upon applying Lemma 3.3.3 to any point lying vertically below x, we obtain a half-space in X containing x. The result follows. □

We define the *join relation* $\sim$ of the half-space decomposition such that, for all $i, j \in I$,

$$i \sim j \;\Leftrightarrow\; \alpha_i(H_i) \cap \alpha_j(H_j) \neq \emptyset. \tag{3.3.3}$$

This relation is trivially reflexive and symmetric, but not transitive. Composing with ϕ, we obtain

$$i \sim j \;\Rightarrow H_i \cap H_j \neq \emptyset, \tag{3.3.4}$$

and

$$i \sim j,\; j \sim k,\; H_i \cap H_j \cap H_k \neq \emptyset \;\Rightarrow\; i \sim k. \tag{3.3.5}$$

As in Sect. 3.2.2, we call the pair $((H_i)_{i \in I}, \sim)$ the *combinatorial data* of X. By Theorem 3.2.2 and the subsequent remark, this data is sufficient to recover X up to isometry.

As a first application of the half-space decomposition, we obtain an elementary formula for the height function. Indeed, for all $x \in X$, let $I(x)$ denote the set of indices $i \in I$ such that $x \in \alpha_i(H_i)$.

Lemma 3.3.5 *Let X be a developable hyperbolic end with height function h, developing map ϕ, and half-space decomposition $(H_i, \alpha_i)_{i\in I}$. For all $x \in X$,*

$$h(x) = \text{sup}_{i\in I(x)} d(\phi(x), \partial H_i). \tag{3.3.6}$$

In particular, the height function of X is unique.

Proof Choose $x \in X$. Since the integral curves of the gradient of h are unit speed geodesics,

$$h(x) \geq \text{sup}_{i\in I(x)} d(\phi(x), \partial H_i).$$

Conversely, by completeness, there exists an integral curve $\gamma :]-h(x), \infty[\to X$ of ∇h such that $\gamma(0) = x$. By Lemma 3.3.3, for all $\epsilon > 0$, there exists $k \in I(x)$ such that $\gamma(\epsilon - h(x)) \in \partial\alpha_k(H_k)$ and $\dot{\gamma}(\epsilon - h(x))$ is the inward-pointing normal to $\partial\alpha_k(H_k)$ at this point. In particular,

$$\text{sup}_{i\in I(x)} d(\phi(x), \partial H_i) \geq d(\phi(x), \partial H_k) = h(x) - \epsilon.$$

Since $\epsilon > 0$ is arbitrary, the result follows. □

Corollary 3.3.6 (Monotonicity) *Let X and X' be developable hyperbolic ends with respective height functions h and h'. If $\psi : X \to X'$ is a morphism, then*

$$h \leq h' \circ \psi. \tag{3.3.7}$$

Proof Indeed, ψ sends the half-space decomposition of X into the half-space decomposition of X'. □

More generally, we obtain the following structure result for morphisms of hyperbolic ends.

Lemma 3.3.7 *Let X and X' be developable hyperbolic ends with respective height functions h and h'. If $\psi : X \to X'$ is a morphism then, for all $x \in X$,*

$$\langle \nabla h(x), \nabla(h' \circ \psi)(x)\rangle > 0. \tag{3.3.8}$$

Proof Let $\phi : X \to \mathbb{H}^3$ and $\phi' : X' \to \mathbb{H}^3$ be developing maps such that $\phi = \phi' \circ \psi$. Let x be a point of X. By Lemma 3.3.3, there exists a unique half-space (H, α) in X such that $x \in \partial\alpha(H)$ and ∇h is the inward-pointing normal to $\partial\alpha(H)$ at this point. Observe furthermore that the closure of $\alpha(H)$ is complete in X. Its image $(H, \psi\circ\alpha)$ is a half-space in X' such that the closure of $(\psi \circ \alpha)(H)$ is complete in X'. Denote $Y' := \partial(\psi \circ \alpha)(H)$ and let $N' : Y' \to TX'$ denote its inward-pointing unit normal vector field. At every point of Y', $\langle N', \nabla h' \rangle > 0$, for otherwise h' would vanish at some point x' of the closure of $(\psi \circ \alpha)(H)$, which is absurd. The result now follows upon pulling back this inequality through ψ and evaluating at x. □

3.3.3 Geodesic Arcs and Convexity

Geodesic arcs in the half-space decomposition are defined in a similar manner as in the Möbius case. We first consider open half-spaces H_0 and H_1 in $\mathbb{H}^3$. We say that H_0 and H_1 *overlap* whenever their boundaries meet along a geodesic. Observe that this holds if and only if their intersection is non-trivial, the intersection of their complements is non-trivial and one is not contained within the other. When H_0 and H_1 overlap, we define the *geodesic arc* between them to be the set of all open half-spaces H in $\mathbb{H}^3$ such that

$$H_0 \cap H_1 \subseteq H \subseteq H_0 \cup H_1. \tag{3.3.9}$$

This definition extends to half-spaces in developable hyperbolic ends as follows. Let X be a developable hyperbolic end with developing map ϕ. We say that two distinct open half-spaces (H_0, α_0) and (H_1, α_1) in X *overlap* whenever the sets $\alpha_0(H_0)$ and $\alpha_1(H_1)$ have non-trivial intersection and neither is contained within the other. Upon composing with ϕ, it follows that H_0 and H_1 likewise have non-trivial intersection, and neither is contained within the other. Furthermore, their complements also have non-trivial intersection, for otherwise X would be isometric to $\mathbb{H}^3$, contradicting the existence of a height function. H_0 and H_1 consequently overlap. Using a connectedness argument, we show that

$$\alpha_0|_{H_0 \cap H_1} = \alpha_1|_{H_0 \cap H_1}, \tag{3.3.10}$$

so that these functions join to define a function $\alpha_{01} : H_0 \cup H_1 \to X$ such that

$$\phi \circ \alpha_{01} = \mathrm{Id}. \tag{3.3.11}$$

In particular, for any other open half-space H along the geodesic arc from H_0 to H_1, $(H, \alpha_{01}|_H)$ is also a half-space in X. We thus define the *geodesic arc* from (H_0, α_0) to (H_1, α_1) to be the set of all half-spaces in X of this form. We say that a subset $(H_i, \alpha_i)_{i \in J}$ of the half-space decomposition of X is *convex* whenever it contains the geodesic arc between any two of its overlapping elements.

Using this concept of convexity, we obtain deeper information about the structure of the height function. Indeed, let X be a simply connected hyperbolic end with height function h, developing map ϕ and half-space decomposition $(H_i, \alpha_i)_{i \in I}$. For all $x \in X$, let $I(x)$ be as in the preceding section and observe now that this set is convex. For all $i \in I(x)$, let $r_{x,i}$ denote the supremal radius of open geodesic balls in H_i centred on $\phi(x)$. By Lemma 3.3.5, the height function h of X satisfies

$$h(x) := \sup_{i\in I(x)} r_{x,i}. \tag{3.3.12}$$

For all $x \in X$, define the partial order $\geq_x$ over I such that, for all $i, j \in I$,

$$i \geq_x j \;\Leftrightarrow i, j \in I(x) \text{ and } r_{x,i} \geq r_{x,j}. \tag{3.3.13}$$

Define also

$$\hat{\Omega}_x := \cup_{i\in I(x)} H_i. \tag{3.3.14}$$

By a connectedness argument, for all $i, j \in I(x)$,

$$\alpha_i|_{H_i\cap H_j} = \alpha_j|_{H_i\cap H_j}, \tag{3.3.15}$$

so that these functions join to define a smooth function $\alpha_x : \hat{\Omega}_x \to X$ such that

$$\phi \circ \alpha_x = \mathrm{Id}. \tag{3.3.16}$$

We call $(\hat{\Omega}_x, \alpha_x)$ the *localisation* of X about x. Let $\hat{K}_x$ denote the complement of $\hat{\Omega}_x$ in $\mathbb{H}^3$ and let $h_x : \hat{\Omega}_x \to \mathbb{R}$ denote the distance to $\hat{K}_x$. Since $\hat{K}_x$ is an intersection of closed half-spaces, it is a closed, convex subset of $\mathbb{H}^3$ so that $\hat{\Omega}_x$ is a hyperbolic end with height function h_x.

Lemma & Definition 3.3.8 *Let X be a developable hyperbolic end with height function h and developing map ϕ. Let x be a point of X, let $(\hat{\Omega}_x, \alpha_x)$ denote the localisation of X at x, and let h_x denote its height function. $I(x)$ contains a unique maximal element for $\geq_x$ which realises $h(x)$. Furthermore,*

$$h(x) = (h_x \circ \phi)(x), \tag{3.3.17}$$

and for all $y \in \alpha_x(\hat{\Omega}_x)$,

$$h(y) \geq (h_x \circ \phi)(y). \tag{3.3.18}$$

We denote by $\max(x)$ *the unique maximal element of $I(x)$.*

Proof Let x be a point of X. Since, by (3.3.12), $(r_{x,i})_{i\in I(x)}$ is bounded above by $h(x)$, $I(x)$ contains a maximal element, and existence follows. Since $I(x)$ is convex and since the restriction of $r_{x,i}$ to every geodesic arc in $I(x)$ is strictly concave, uniqueness follows. Finally, since α_x sends the half-space decomposition of $\hat{\Omega}_x$ to $I(x)$, (3.3.17) and (3.3.18) follow, and this completes the proof. □

Lemma 3.3.9 *Let X be a developable hyperbolic end with developing map ϕ. Let x be a point of X and let $(\hat{\Omega}_x, \alpha_x)$ denote the localisation of X about x. There exists* *a neighbourhood U_x of x in $\alpha_x(\hat{\Omega}_x)$ such that, for all $y \in U_x$,*

$$\max(y) \in I(x). \tag{3.3.19}$$

Proof Let h denote the height function of X and let $(H_i, \alpha_i)_{i\in I}$ denote its half-space decomposition. For $x \in X$, define

$$U_x := \left\{ y \in \alpha_x(\hat{\Omega}_x) \mid d(\phi(y), \phi(x)) < h(x)/2 \right\}.$$

For $y \in U_x$, $h(y) > h(x)/2$. It follows that if $i := \max(y)$, then H_i contains the ball of radius $h(x)/2$ about $\phi(y)$. In particular, $\phi(x)$ is an element of H_i, so that x is an element of $\alpha_i(H_i)$, and $i \in I(x)$, as desired. □

Corollary 3.3.10 *Let X be a developable hyperbolic end with height function h and developing map ϕ. Let x be a point of X, let $(\hat{\Omega}_x, \alpha_x)$ denote the localisation of X about x, and let h_x denote its height function. With U_x as in Lemma 3.3.9, for $y \in U_x$,*

$$h(y) = (h_x \circ \phi)(y). \tag{3.3.20}$$

Proof Indeed, α_x sends the half-space decomposition of $\hat{\Omega}_x$ to $I(x)$. □

We are now ready to determine more refined analytic properties of the height function. We first require the following definition of PDE theory (c.f. [3]). Given a smooth manifold Y, a point $y \in Y$, a function $f : Y \to \mathbb{R}$ and a symmetric bilinear form B over T_yY, we say that

$$\mathrm{Hess}(f)(x) \geq B \tag{3.3.21}$$

in the *weak sense* whenever there exists a neighbourhood Ω of y in Y and a smooth function $g : \Omega \to \mathbb{R}$ such that

(1) $g \leq f$;
(2) $g(y) = f(y)$; and
(3) $\mathrm{Hess}(g)(y) = B$.

We likewise say that $\mathrm{Hess}(f)(y) \leq B$ in the *weak sense* whenever $\mathrm{Hess}(-f)(y) \geq -B$ in the weak sense.

Lemma 3.3.11 *Let X be a developable hyperbolic end with height function h. For all $x \in X$, with respect to the decomposition $T_xX = \mathrm{Ker}(dh(x)) \oplus \langle\nabla h(x)\rangle$,*

$$\begin{pmatrix} \tanh(h(x))\mathrm{Id} & 0 \\ 0 & 0 \end{pmatrix} \leq \mathrm{Hess}(h)(x) \leq \begin{pmatrix} \coth(h(x))\mathrm{Id} & 0 \\ 0 & 0 \end{pmatrix}. \tag{3.3.22}$$

in the weak sense.

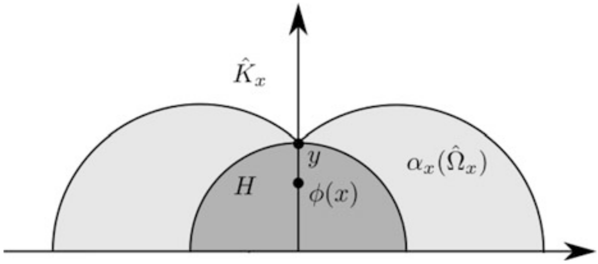


Fig. 3.3 Here, y is the closest point of $\hat{K}_x$ to $\phi(x)$, and H is the supporting half-space to $\hat{K}_x$ at this point whose inward-pointing normal points towards $\phi(x)$. For any other point z of H, ∂H is closer to z than $\hat{K}_x$, which in turn is no further from to z than y

Proof Let ϕ denote the developing map of X. Let x be a point of X. By Corollary 3.3.10, it suffices to prove the result for the localisation $(\hat{\Omega}_x, \alpha_x)$ of X at x. Let $y \in \hat{K}_x$ be the closest point to $\phi(x)$. Let H denote the supporting open half-space to $\hat{K}_x$ at y whose boundary is normal to the geodesic joining y to $\phi(x)$ (see Fig. 3.3). Let $f, g : \mathbb{H}^3 \to \mathbb{R}$ denote respectively the distance to y and the distance to ∂H. Trivially, $f(\phi(x)) = g(\phi(x)) = h_x(\phi(x))$ and, over H, $g \leq h_x \leq f$. The result now follows upon explicitly determining the hessian operators of f and g at $\phi(x)$. □

3.3.4 *Ideal Boundaries*

We now study functors which map between the categories of simply connected Möbius surfaces and simply connected hyperbolic ends. We first describe the ideal boundary functor ∂_∞ which associates a simply connected Möbius surface to every simply connected hyperbolic end. For this we require the following finer control of complete geodesic rays in hyperbolic ends.

Lemma & Definition 3.3.12 *Let X be a developed hyperbolic end with height function h. For every complete geodesic ray $\gamma : [0, \infty[\to X$,*

$$\lim_{t\to\infty}(h \circ \gamma)(t) \in \{0, \infty\}\,. \tag{3.3.23}$$

We say that γ is bounded whenever this limit is zero and unbounded otherwise.

Proof Indeed, by convexity $(h \circ \gamma)(t)$ converges to a (possibly infinite) limit as t tends to infinity. Suppose now that

$$\lim_{t\to\infty}(h \circ \gamma)(t) > 2\epsilon,$$

for some $\epsilon > 0$. Denoting $f(t) := (h \circ \gamma)(t)$, (3.3.22) yields, for sufficiently large t,

$$\ddot{f} \geqslant \tanh(\epsilon)(1 - \dot{f}^2)$$

in the weak sense. Upon solving this ordinary differential inequality, we see that $f(t)$ tends to $+\infty$ as t tends to infinity, as desired. □

Lemma 3.3.13 *Let X be a developable hyperbolic end and let $\gamma : [0, \infty[\to X$ be a complete, unbounded geodesic ray. For all $t > 0$, there exists $x \in X_t$ such that*

$$\lim_{s \to +\infty} (\pi_t \circ \gamma)(s) = x. \tag{3.3.24}$$

In particular, γ is asymptotic to the vertical line passing through x.

Proof Naturally, we may suppose that γ is parametrized by arc-length. Let h denote the height function of X. By (3.3.22), for sufficiently large t, the function $f := \langle \dot{\gamma}, \nabla h \circ \gamma \rangle$ satisfies

$$\dot{f}(t) \geq (1 - \epsilon)(1 - f(t)^2)$$

in the weak sense. Solving this ordinary differential inequality, we show that f converges exponentially fast to 1, so that the component of $\dot{\gamma}$ orthogonal to $\nabla h \circ \gamma$ converges exponentially fast to zero. Since π_t is 1-Lipschitz, and since ∇h lies in the kernel of $D\pi_t$, the curve $(\pi_t \circ \gamma)$ thus has finite length, and the result now follows by completeness. □

Let X be a developable hyperbolic end with developing map ϕ. We define $\partial_\infty X$, the *ideal boundary* of X, to be the space of equivalence classes of complete, unbounded geodesic rays in X, where two such rays are deemed equivalent whenever they are asymptotic to one another. The union $X \cup \partial_\infty X$ is topologised as follows. By Theorem 3.3.1, for all $x \in X$, for every upward-pointing unit vector $\xi \in T_x X$, for all $r > 0$, and for sufficiently small $\theta \in]0, \pi[$, the truncated cone

$$\mathrm{C}(\xi, \theta, r) := \left\{ \mathrm{Exp}_x(t\mu) \mid \mu \in T_x X,\ \|\mu\| = 1, t > r,\ \langle \mu, \xi \rangle > \cos(\theta) \right\} \tag{3.3.25}$$

is well-defined. For all such ξ, θ and r, we define the ideal boundary $\partial_\infty \mathrm{C}(\xi, \theta, r)$ to be the set of equivalence classes of unbounded geodesic rays which eventually lie in $\mathrm{C}(\xi, \theta, r)$. The collection of all sets of the form

$$\mathrm{C}(\xi, \theta, r) \cup \partial_\infty \mathrm{C}(\xi, \theta, r) \tag{3.3.26}$$

together with the open subsets of X forms a basis of a Haussdorf topology of $X \cup \partial_\infty X$ which we call the *cone topology*. In particular, with respect to this topology, $\partial_\infty X$ has the structure of a topological surface.

By Lemma 3.3.13, every complete geodesic ray in X is asymptotic to some vertical line. On the other hand, since π_t is 1-Lipschitz for all t, no two vertical lines are asymptotic to one another. It follows that $\partial_\infty X$ is homeomorphic to the leaf space of the vertical line foliation of X which, we recall, is in turn homeomorphic to every level X_t of X. In particular, since X retracts onto X_t for all t, it follows that X and $\partial_\infty X$ are homotopy equivalent.

Since the developing map $\phi : X \to \mathbb{H}^3$ sends complete geodesic rays continuously to complete geodesic rays, it defines a continuous function $\partial_\infty\phi : \partial_\infty X \to \partial_\infty\mathbb{H}^3$. By standard properties of convex subsets of hyperbolic space, this function is a local homeomorphism and thus defines a developable Möbius structure over $\partial_\infty X$ which we readily verify is of hyperbolic type. In particular, we verify that, for all t, the homeomorphism sending $\partial_\infty X$ to X_t is in fact a smooth diffeomorphism with respect to this structure.

Finally, let X' be another developable hyperbolic end with developing map $\phi' : X' \to \mathbb{H}^3$ and let $\psi : X \to X'$ be a morphism such that $\phi := \phi' \circ \psi$. Since ψ also maps complete, unbounded geodesic rays continuously to complete, unbounded geodesic rays, it defines a morphism $\partial_\infty\psi : \partial_\infty X \to \partial_\infty X'$ such that $\partial_\infty\phi' \circ \partial_\infty\psi = \partial_\infty\phi$. We verify that ∂_∞ respects identity elements and compositions, and thus defines a covariant functor from the category of simply connected hyperbolic ends into the category of simply connected Möbius surfaces.

It is useful to observe how the ideal boundary functor acts on the half-space decomposition of the hyperbolic end.

Lemma 3.3.14 *Let X be a developable hyperbolic end with developing map ϕ, let $(H_i, \alpha_i)_{i\in I}$ denote its half-space decomposition, and let $\sim$ denote its join relation. Then $(\partial_\infty H_i, \partial_\infty\alpha_i)_{i\in I}$ is a subset of the Möbius disk decomposition of $(\partial_\infty X, \partial_\infty\phi)$ which covers $\partial_\infty X$. Furthermore, the restriction to I of the join relation of the Möbius disk decomposition of $\partial_\infty X$ coincides with $\sim$.*

Remark 3.3.15 Significantly, however, $(\partial_\infty H_i, \partial_\infty\alpha_i)_{i\in I}$ rarely accounts for the entire Möbius disk decomposition of $\partial_\infty X$. Indeed, this only occurs when X is functionally maximal in the sense of Lemma and Definition 3.3.21, below.

Proof For all i, $\partial_\infty H_i$ is a disk in $\hat{\mathbb{C}} = \partial_\infty\mathbb{H}^3$ and, by functoriality, $\partial_\infty\alpha_i$ defines a function from $\partial_\infty H_i$ into $\partial_\infty X$ such that

$$\partial_\infty\phi \circ \partial_\infty\alpha_i = \mathrm{Id}.$$

It follows that $(\partial_\infty H_i, \partial_\infty\alpha_i)_{i\in I}$ is a subset of the Möbius disk decomposition of $\partial_\infty X$. We now show that $(\partial_\infty H_i, \partial_\infty\alpha_i)_{i\in I}$ covers $\partial_\infty X$. Let $\gamma : [0, \infty[\to X$ be a complete, unit speed geodesic ray. Let $t_0 > 0$ be such that $\dot\gamma(t_0)$ is upward pointing. Let i be the unique element of I such that $\gamma(t_0) \in \partial H_i$ and $\dot\gamma(t_0)$ is the inward pointing unit normal to ∂H_i at this point. The equivalence class of γ is then an element of $\partial_\infty\alpha_i(\partial_\infty H_i)$ and since γ is arbitrary, it follows that $(\partial_\infty H_i, \partial_\infty\alpha_i)_{i\in I}$ covers $\partial_\infty X$, as desired. Finally, we readily show that, for all $i, j \in I$,

$$i \sim j \;\Leftrightarrow\; \partial_\infty\alpha_i(\partial H_i) \cap \partial_\infty\alpha_j(\partial H_j) \neq \emptyset,$$

so that the restriction to I of the join relation of the Möbius disk decomposition of $\partial_\infty X$ coincides with $\sim$, as desired. □

It is also worth verifying that half-spaces in X are uniquely determined by their ideal boundaries in $\partial_\infty X$.

Lemma 3.3.16 *Let X be a developable hyperbolic end. For any two half-spaces (H_i, α_i) and (H_j, α_j) in X,*

$$(\partial_\infty H_i, \partial_\infty \alpha_i) = (\partial_\infty H_j, \partial_\infty \alpha_j) \ \Rightarrow\ (H_i, \alpha_i) = (H_j, \alpha_j). \tag{3.3.27}$$

Proof Since $\partial_\infty H_i = \partial_\infty H_j$, we have $H_i = H_j =: H$. Let ϕ denote the developing map of X. Denote $U := \alpha_i(H_i) \cap \alpha_j(H_j)$ and $V = \phi(U)$. Observe that, over V, $\alpha_i = \phi^{-1} = \alpha_j$. It thus suffices to show that $V = H$. However, since α_i and α_j are local isometries, ∂V is a totally geodesic subset of $\overline{H}$ and, since $\partial_\infty \alpha_i = \partial_\infty \alpha_j$, $\partial_\infty V = \partial_\infty H$, so that $V = H$, as desired. □

Finally, the following estimate, though elementary, will play a key role in Sect. 3.4 in the study of quasicomplete ISC immersions in $\mathbb{H}^3$. Let X be a developable hyperbolic end with developing map ϕ and let $(\partial_\infty X, \partial_\infty \phi)$ denote its ideal boundary. Let $\pi_\infty : X \to \partial_\infty X$ denote the function that sends every point $x \in X$ to the equivalence class of the vertical line on which it lies. We call π_∞ the *vertical line projection*.

Lemma 3.3.17 *Let X be a developable hyperbolic end with developing map ϕ, let $(\partial_\infty X, \partial_\infty \phi)$ denote its ideal boundary, let ω_∞ denote the Kulkarni–Pinkall form of $\partial_\infty X$ and let $\pi_\infty : X \to \partial_\infty X$ denote the vertical line projection. For all $x \in X$,*

$$\phi(x) \in B((\partial_\infty \phi)_*(\omega_\infty \circ \pi_\infty)(x)), \tag{3.3.28}$$

where B here denotes the parametrisation of the space of open horoballs in $\mathbb{H}^3$ by $\Lambda^2 \partial_\infty \mathbb{H}^3$ as described in Sect. 3.1.2.

Proof Let h denote the height function of X, let x be a point of X and denote $x_\infty := \pi_\infty(x)$. Let y be a point of X lying vertically below x. In particular, $x_\infty = \pi_\infty(y)$. Let (H, α) be the unique half-space of X such that $y \in \partial\alpha(H)$ and $\nabla h(y)$ is the inward-pointing unit normal to $\partial\alpha(H)$ at this point. Since ∂_∞ is functorial, $(\partial_\infty H, \partial_\infty \alpha)$ is a Möbius disk in $\partial_\infty X$. By definition of the Kulkarni–Pinkall form,

$$(\partial_\infty \phi)_* \omega_\infty(x_\infty) \leq \omega(\partial_\infty H)(\partial_\infty \phi(x_\infty)).$$

Thus, if B is the largest open horoball contained in H with asymptotic centre $\partial_\infty \phi(x_\infty)$, then

$$B \subseteq B((\partial_\infty \phi)_* \omega_\infty(x_\infty)).$$

Since B is an interior tangent to ∂H at $\phi(y)$, it contains $\phi(x)$, and the result follows. □

3.3.5 Extensions of Möbius Surfaces

We now construct the extension functor $\mathcal{H}$, and show that it is a right-inverse functor of ∂_∞. Let S be a developable Möbius surface of hyperbolic type with developing map ϕ. Let $(D_i, \alpha_i)_{i\in I}$ denote its Möbius disk decomposition and let $\sim$ denote its join relation. For all i, let H_i denote the open half-space in $\mathbb{H}^3$ with ideal boundary D_i. Observe that $((H_i)_{i\in I}, \sim)$ are combinatorial data of some hyperbolic manifold in the sense of Theorem 3.2.2. Let $\mathcal{H}S$, $\mathcal{H}\phi : \mathcal{H}S \to \mathbb{H}^3$ and $(\mathcal{H}\alpha_i)_{i\in I}$ denote respectively the join of $(H_i)_{i\in I}$, its canonical immersion and its canonical parametrisations. In particular, $(H_i, \mathcal{H}\alpha_i)_{i\in I}$ is a half-space decomposition of $\mathcal{H}S$.

In order to show that $\mathcal{H}S$ is a hyperbolic end, it remains only to construct its height function. Bearing in mind Lemma 3.3.5, we proceed as follows. For $x \in \mathcal{H}S$, let $I(x)$ denote the subset of I consisting of those indices for which $x \in \mathcal{H}\alpha_i(H_i)$, and observe that this set is convex. For all $i \in I(x)$, let $r_{x,i}$ denote the supremal radius of open geodesic balls in H_i centred on $\mathcal{H}\phi(x)$. We now define the function $h : \mathcal{H}S \to \mathbb{R}$ by

$$h(x) := \mathrm{sup}_{i\in I(x)} r_{x,i}. \tag{3.3.29}$$

Lemma 3.3.18 *The function h is a height function over $\mathcal{H}S$.*

Proof We first observe that, since S is of hyperbolic type, $I(x)$ contains a maximal element for all $x \in \mathcal{H}S$, and uniqueness of this maximal element follows by the convexity arguments already used earlier in this section. The construction and results of Sect. 3.3.3 now follow as before. It remains only to verify that h has the required analytic properties. Let x be a point of $\mathcal{H}S$. Let $(\hat{\Omega}_x, \mathcal{H}\alpha_x)$ denote the localisation of $\mathcal{H}S$ at this point and let h_x denote its height function. With U_x as in Lemma 3.3.9, for all $y \in U_x$, $h(y) = (h_x \circ \mathcal{H}\phi)(x)$. It thus follows by standard properties of convex sets in $\mathbb{H}^3$ that h is a locally strictly convex $C^{1,1}_{\text{loc}}$ function whose gradient flow lines are unit speed geodesics. Finally, for all $t > 0$, for all $x \in \mathcal{H}S$ such that $h(x) \geq t$, and for all $\epsilon < t$, the closed ball of radius $(t - \epsilon)$ about x in $\mathcal{H}S$ is complete. It follows that $h^{-1}([t, \infty[)$ is complete for all $t > 0$, and this completes the proof. □

It follows by Lemma 3.3.18 that the operator $\mathcal{H}$ associates a hyperbolic end $\mathcal{H}S$ to every developable Möbius surface S of hyperbolic type. Given two such Möbius surfaces S and S' and an injective morphism $\phi : S \to S'$, Theorem 3.2.2 yields a canonically defined morphism $\mathcal{H}\phi : \mathcal{H}S \to \mathcal{H}S'$. We verify that $\mathcal{H}$ respects identity elements and compositions and thus defines a covariant functor between these two categories. We call $\mathcal{H}$ the *extension functor*. It is a right inverse of the ideal boundary functor, as the following result shows.

Lemma 3.3.19 *Let S be a developable Möbius surface of hyperbolic type with developing map ϕ, Möbius disk decomposition $(D_i, \alpha_i)_{i\in I}$ and extension $(\mathcal{H}S, \mathcal{H}\phi)$.*

There exists a unique isomorphism $\psi : S \to \partial_\infty \mathcal{H}S$ such that, for all i,

$$\partial_\infty \mathcal{H}\alpha_i = \psi \circ \alpha_i. \tag{3.3.30}$$

Remark 3.3.20 Naturally, in what follows, rather than mention ψ explicitly, we identify S and $\partial_\infty \mathcal{H}S$.

Proof For all i, let H_i denote the open half-space in $\mathbb{H}^3$ with ideal boundary D_i. Since $(H_i, \mathcal{H}\alpha_i)_{i\in I}$ is a subset of the half-space decomposition of $\mathcal{H}S$ which covers $\mathcal{H}S$, by Lemma 3.3.14, $(D_i, \partial_\infty \mathcal{H}\alpha_i)_{i\in I}$ is a subset of the Möbius disk decomposition of $\partial_\infty \mathcal{H}S$ which likewise covers $\partial_\infty \mathcal{H}S$. Furthermore, the join relation of this decomposition coincides with that of $(H_i, \mathcal{H}\alpha_i)$, which in turn coincides with that of $(D_i, \alpha_i)_{i\in I}$. It follows by Theorem 3.2.2 that there exists a unique diffeomorphism $\psi : S \to \partial_\infty \mathcal{H}S$ satisfying (3.3.30), as desired. □

Finally, the height functions of hyperbolic ends obtained by extending Möbius surfaces have more structure than in the general case. Indeed, given a function $f : X \to \mathbb{R}$, a point $x \in X$, a vector $\xi \in T_x(X)$ and a real number $\lambda \in \mathbb{R}$, we say that

$$\mathrm{Hess}(f)(x)(\xi, \xi) \leq \lambda \tag{3.3.31}$$

in the *weak sense* whenever there exists a geodesic segment $\gamma :]-\epsilon, \epsilon[\to X$ such that $\gamma(0) = x$, $\dot{\gamma}(x) = \xi$, and

$$\left.\frac{\partial^2}{\partial t^2} f \circ \gamma\right|_{t=0} \leq \lambda. \tag{3.3.32}$$

in the weak sense of Sect. 3.3.3.

Lemma & Definition 3.3.21 *Let S be a developable Möbius surface of hyperbolic type, let $\mathcal{H}S$ denote its extension, and let h denote the height function of $\mathcal{H}S$. For all $x \in \mathcal{H}S$, there exists a unit vector $\xi \in T_x\mathcal{H}S$ such that*

$$\langle \xi, \nabla h(x) \rangle = 0, \ \textit{and} \tag{3.3.33}$$

$$\mathrm{Hess}(h)(x)(\xi, \xi) \leq \tanh(h(x)) \tag{3.3.34}$$

in the weak sense. We say that a hyperbolic end X is functionally maximal whenever its height function satisfies (3.3.34)*.*

Remark 3.3.22 We will see in Theorem 3.3.25, below, that a hyperbolic end is an extension of a Möbius surface if and only if it is functionally maximal.

Proof Let ϕ denote the developing map of S, let $(D_i, \alpha_i)_{i\in I}$ denote its Möbius disk decomposition, and, for all i, let H_i denote the open half-space in $\mathbb{H}^3$ with ideal boundary D_i. Let x be a point of $\mathcal{H}S$, let $(\hat{\Omega}_x, \mathcal{H}\alpha_x)$ denote the localisation of $\mathcal{H}S$ about x and let h_x denote its height function.

Denote $y := \pi_\infty(x)$. Let (Ω_y, α_y) denote the localisation of (S, ϕ) about y and denote

$$\mathcal{H}\Omega_y := \cup_{i\in I(y)} H_i.$$

Let $\hat{K}_y$ denote the complement of $\mathcal{H}\Omega_y$ in $\mathbb{H}^3$. Let $h_y : \mathcal{H}\Omega_y \to]0,\infty[$ denote the distance to $\hat{K}_y$ and observe that h_y is a height function over $\mathcal{H}\Omega_y$ so that $\mathcal{H}\Omega_y$ is a hyperbolic end. Indeed, it is none other than the extension of Ω_y. By functoriality, the extension of α_y is a morphism $\mathcal{H}\alpha_y : \mathcal{H}\Omega_y \to \mathcal{H}S$ such that

$$\mathcal{H}\phi \circ \mathcal{H}\alpha_y = \mathrm{Id}.$$

In particular, $\mathcal{H}\alpha_y$ embeds $\mathcal{H}\Omega_y$ into $\mathcal{H}S$.

For all $i \in I$,

$$x \in \mathcal{H}\alpha_i(H_i) \Rightarrow y \in \alpha_i(D_i),$$

so that every half-space in $(\hat{\Omega}_x, \mathcal{H}\alpha_x)$ is also a half-space in $(\mathcal{H}\Omega_y, \mathcal{H}\alpha_y)$. Consequently,

$$\mathcal{H}\alpha_x(\hat{\Omega}_x) \subseteq \mathcal{H}\alpha_y(\mathcal{H}\Omega_y) \subseteq \mathcal{H}S.$$

It follows by Corollaries 3.3.6 and 3.3.10 that, over U_x,

$$h_x \circ \mathcal{H}\phi = h_y \circ \mathcal{H}\phi = h.$$

Now let z denote the closest point in $\hat{K}_y$ to $\hat{x} := \mathcal{H}\phi(x)$. Since $\hat{K}_y$ is the convex hull in $\mathbb{H}^3$ of the complement of Ω_y in $\partial_\infty\mathbb{H}^3$, there exists an open geodesic segment $\gamma :]-\epsilon, \epsilon[\to \hat{K}_y$ such that $\gamma(0) = z$ (see, for example, Section 4.5 of [27]). Let $P \subseteq \mathbb{H}^3$ be the totally geodesic plane containing γ and $\hat{x}$. Let $\mu :]-\delta, \delta[\to P$ be a curve segment in P lying at constant distance from γ such that $\mu(0) = \hat{x}$ (see Fig. 3.4). Since μ has constant geodesic curvature equal to $\tanh(h(x))$, upon denoting $\xi := \dot{\mu}(0)$, we obtain

$$\mathrm{Hess}(h_y)(\mathcal{H}\phi(x))(\xi, \xi) \leq \tanh(h(x))$$

in the weak sense, and this completes the proof. □

3.3.6 *Left Inverses and Applications*

We study the extent to which $\mathcal{H}$ is also a left inverse of ∂_∞.

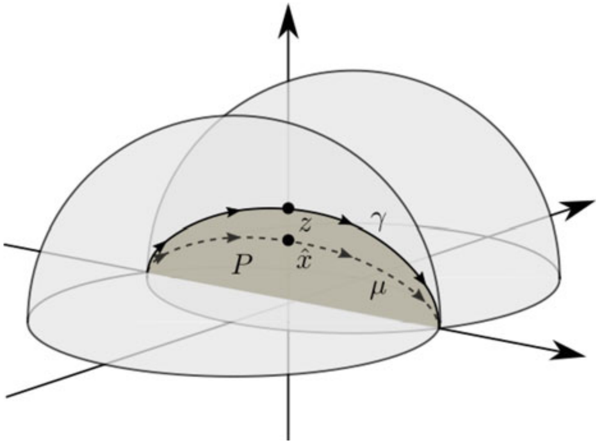


Fig. 3.4 The totally geodesic plane P contains both γ and $\hat{x}$. The curve μ lies at constant distance from γ and is contained in P

Lemma 3.3.23 *Let S be a developable Möbius surface with developing map ϕ. Let X be a developable hyperbolic end with developing map ψ. Let $f : \partial_\infty X \to S$ be an injective morphism such that*

$$\phi \circ f = \partial_\infty \psi. \tag{3.3.35}$$

There exists a unique injective morphism $\hat{f} : X \to \mathcal{H}S$ such that

$$\mathcal{H}\phi \circ \hat{f} = \psi, \text{ and} \tag{3.3.36}$$

$$\partial_\infty \hat{f} = f. \tag{3.3.37}$$

Proof Let $(D_i, \alpha_i)_{i\in I}$ denote the Möbius disk decomposition of S with join relation $\sim_\alpha$, let $(H_j, \beta_j)_{j\in J}$ denote the half-space decomposition of X with join relation $\sim_\beta$, and, for all $j \in J$, denote $D_j := \partial_\infty H_j$. By Lemma 3.3.14, $(D_j, \partial_\infty\beta_j)_{j\in J}$ is a subset of the Möbius disk decomposition of $\partial_\infty X$ which covers $\partial_\infty X$. By (3.3.35), $(D_j, f \circ \partial_\infty\beta_j)_{j\in J}$ is a subset of the Möbius disk decomposition of S which covers $\mathrm{Im}(f)$. We thus identify J with a subset of I in such a manner that, for all j,

$$f \circ \partial_\infty\beta_j = \alpha_j.$$

X identifies with the join of $((H_j)_{j\in J}, \sim_\beta)$ whilst the join Y of $((H_j)_{j\in J}, \sim_\alpha)$ identifies with an open subset of $\mathcal{H}S$. However, by injectivity of f, the join relations $\sim_\alpha$ and $\sim_\beta$ coincide over J, so that, by Theorem 3.2.2, there exists a unique isomorphism $\hat{f} : X \to Y$ such that, for all $j \in J$,

$$\hat{f} \circ \beta_j = \mathcal{H}\alpha_j.$$

Consequently, for all j,

$$\mathcal{H}\phi \circ \hat{f} \circ \beta_j = \mathcal{H}\phi \circ \mathcal{H}\alpha_j = \mathrm{Id} = \psi \circ \beta_j.$$

Likewise, by functoriality,

$$\partial_\infty \hat{f} \circ \partial_\infty \beta_j = \partial_\infty(\hat{f} \circ \beta_j) = \partial_\infty \mathcal{H}\alpha_j = \partial_\infty \mathcal{H}(f \circ \partial_\infty \beta_j).$$

Since $(\beta_j)_{j \in J}$ and $(\partial_\infty \beta_j)_{j \in J}$ cover X and $\partial_\infty X$ respectively, it follows that

$$\mathcal{H}\phi \circ \hat{f} = \psi, \text{ and}$$
$$\partial_\infty \hat{f} = \partial_\infty \mathcal{H} f.$$

Identifying $(\partial_\infty \mathcal{H}) f$ with f as in Lemma 3.3.19, we obtain (3.3.37), and existence follows. To prove uniqueness, let $\hat{f}' : X \to \mathcal{H}S$ be another function satisfying (3.3.36) & (3.3.37). Let j be an element of J. Since

$$\mathcal{H}\phi \circ \hat{f}' \circ \beta_j = \psi \circ \beta_j = \mathrm{Id} = \mathcal{H}\phi \circ \hat{f} \circ \beta_j,$$

it follows that $(H_j, \hat{f}' \circ \beta_j)$ and $(H_j, \hat{f} \circ \beta_j)$ are half-spaces in $\mathcal{H}S$. Furthermore,

$$\partial_\infty(\hat{f}' \circ \beta_j) = \partial_\infty \hat{f}' \circ \partial_\infty \beta_j = (\partial_\infty \mathcal{H} f) \circ \partial_\infty \beta_j = \partial_\infty \hat{f} \circ \partial_\infty \beta_j = \partial_\infty(\hat{f} \circ \beta_j),$$

so that, by Lemma 3.3.16,

$$\hat{f}' \circ \beta_j = \hat{f} \circ \beta_j.$$

Since $(H_j, \beta_j)_{j \in J}$ covers X, it follows that $\hat{f}' = \hat{f}$, and uniqueness follows. □

With Lemma 3.3.23 in mind, we now study the relationship between two hyperbolic ends when one is contained within the other. Thus, let X be a developable hyperbolic end. Let $\mathcal{V}$ denote its vertical line foliation whose leaf space we recall is naturally homeomorphic to $\partial_\infty X$. Let S be a C^1 embedded surface in X. We say that S is a *graph* over an open subset Ω of $\partial_\infty X$ whenever it is transverse to $\mathcal{V}$ and the vertical line projection restricts to a homeomorphism from ∂S onto Ω.

Lemma 3.3.24 *Let X and X' be developable hyperbolic ends. If $\psi : X \to X'$ is an injective morphism, then the image $\psi(X_t)$ of every level of X is a graph over $\partial_\infty \psi(\partial_\infty X)$ in X'.*

Proof Indeed, choose $t > 0$. Let $\pi'_\infty : X' \to \partial_\infty X'$ denote the vertical line projection of X'. By Lemma 3.3.7, $Y_t := \psi(X_t)$ is everywhere transverse to the vertical foliation of X' so that the restriction of π'_∞ to this surface is everywhere a local homeomorphism. By Theorem 3.3.1, any vertical line which enters $\psi(h^{-1}[t,\infty[)$ remains within this set, so that no vertical line of X' can cross Y_t more than once. It follows that the restriction of π'_∞ to this surface is injective.

It only remains to prove surjectivity. By connectedness, it suffices to show that $\pi'_\infty(Y_t)$ is a closed subset of $\partial_\infty\psi(\partial_\infty X)$. Thus, let $(x'_m)_{m\in\mathbb{N}}$ be a sequence of points of $\pi'_\infty(Y_t)$ converging to the limit $x'_\infty \in \partial_\infty\psi(\partial_\infty X)$. For all $m \in \mathbb{N}\cup\{\infty\}$, let $\gamma'_m :]0,\infty[\rightarrow X'$ be the height parametrisation of the vertical line of X' terminating at x'_m and, for all finite m, let $T_m > 0$ be such that $\gamma'_m(T_m) \in Y_t$. Since x'_∞ is an element of $\partial_\infty\psi(\partial_\infty X)$, there exists $T > 0$ such that $\gamma'_\infty(T) \in \psi(X)$. Since $\psi(X)$ is open, we may therefore suppose that $\gamma'_m(T) \in \psi(X)$ for all m. In particular, $T_m < T$ for all m, and we may therefore suppose that $(T_m)_{m\in\mathbb{N}}$ converges to some value T_∞, say. For all $m \in \mathbb{N}\cup\{\infty\}$, let $\gamma_m : [T_m - T,\infty[\rightarrow X$ denote the unit speed parametrisation of the preimage of γ'_m under ψ, normalised such that $\psi(\gamma_m(0)) = \gamma'_m(T)$. For all finite m, denote $y_m := \mu(T_m - T)$ so that $y_m \in X_t$ and $x'_m = \psi(y_m)$. Since the projection along vertical lines in X to X_t is distance decreasing, it follows that

$$\text{limsup}_{m,n\rightarrow\infty} d(y_m, y_n) \leq 2T,$$

so that, by completeness, there exists a point y_∞, say, of X_t towards which $(y_m)_{m\in\mathbb{N}}$ subconverges. We verify that $\pi'_\infty(\psi(y_\infty)) = x'_\infty$ so that $\pi'_\infty(Y_t)$ is indeed a closed subset of $\partial_\infty\psi(\partial_\infty X)$. Surjectivity follows, and this completes the proof. □

Theorem 3.3.25 *Let S be a developable Möbius surface of hyperbolic type with developing map ϕ. Let $(\mathcal{H}S, \mathcal{H}\phi)$ denote its extension. $\mathcal{H}S$ is the only functionally maximal developable hyperbolic end with ideal boundary S.*

Proof Let X be another developable hyperbolic end with height function h and developing map ψ such that $\partial_\infty X = S$. Let $\hat{f} : X \rightarrow \mathcal{H}S$ denote the unique injective morphism such that $\mathcal{H}\phi \circ \hat{f} = \psi$ and $\partial_\infty \hat{f} = \text{Id}$. We identify X with its image $\hat{f}(X)$ in $\mathcal{H}S$. Let $\hat{h}$ denote the height function of $\mathcal{H}S$. By Corollary 3.3.6, $h \leq \hat{h}$. We now claim that $h = \hat{h}$. Indeed, suppose the contrary. Choose $x \in \mathcal{H}S$ such that $\hat{h}(x) > h(x)$. By completeness, for sufficiently small $\epsilon > 0$, there exists a geodesic ray $\gamma : [-h(x)-\epsilon,\infty[\rightarrow \mathcal{H}S$ such that $\gamma(0) = x$ and $\dot{\gamma}(0) = \nabla h(x)$. Let (H,α) be the unique half-space in $\mathcal{H}S$ such that $\gamma(-h(x)-\epsilon) \in \partial\alpha(H)$ and $\dot{\gamma}(-h(x)-\epsilon)$ is the inward-pointing unit normal to $\partial\alpha(h)$ at this point, let $h_\epsilon : \alpha(H) \rightarrow]0,\infty[$ denote distance to $\partial\alpha(H)$, and let Y_ϵ denote the level set of this function at height $h(x)+\epsilon$. For sufficiently small ϵ, Y_ϵ is wholly contained in X and h restricts to a proper function over this set. Let $y \in Y_\epsilon$ be the point at which h is minimised. Since $x \in Y_\epsilon$, $h(y) \leq h(x)$. However, at this point, with respect to the decomposition $T_yX = \text{Ker}(dh(y)) \oplus \langle\nabla h(y)\rangle$,

$$\text{Hess}(h)(y) \geq \text{Hess}(h_\epsilon)(y) = \begin{pmatrix}\tanh(h(x)+\epsilon) & 0\\ 0 & 0\end{pmatrix},$$

which contradicts (3.3.34). It follows that $h = \hat{h}$ as asserted. Finally, since every level of X is a graph over S, it follows that $X = \mathcal{H}S$, as desired. □

We conclude this section by addressing the case of non-developable Möbius surfaces and proving Theorem 3.1.8. Observe first that the definition of the ideal boundary functor ∂_∞ given in Sect. 3.3.4 readily extends to the non-developable case. We now examine the extension functor. Let S be a Möbius surface with fundamental group Π. Let $\tilde{S}$ denote its universal cover, let ϕ be a developing map of $\tilde{S}$ and let θ denote its holonomy. Let $\text{Deck} : \Pi \to \text{Isom}(\tilde{S})$ denote the action of Π on $\tilde{S}$ by deck transformations. By definition, for all $\gamma \in \Pi$,

$$\theta(\gamma) \circ \phi = \phi \circ \text{Deck}(\gamma). \tag{3.3.38}$$

By Lemma 3.3.23, Deck extends to a unique homeomorphism $\mathcal{H}\text{Deck} : \Pi \to \text{Isom}(\mathcal{H}\tilde{S})$ such that

$$\theta(\gamma) \circ \mathcal{H}\phi = \mathcal{H}\phi \circ \mathcal{H}\text{Deck}(\gamma), \text{ and} \tag{3.3.39}$$

$$\partial_\infty(\mathcal{H}\phi \circ \mathcal{H}\gamma) = \phi \circ \text{Deck}(\gamma). \tag{3.3.40}$$

In addition, for all γ, $\mathcal{H}\text{Deck}(\gamma)$ preserves every level of $\mathcal{H}\tilde{S}$, and its action on each level is conjugate to its action on S. It follows that $\mathcal{H}\text{Deck}$ acts discretely on $\mathcal{H}\tilde{S}$, and we define

$$\mathcal{H}S := \mathcal{H}\tilde{S}/\mathcal{H}\text{Deck}(\Pi). \tag{3.3.41}$$

We verify that $\mathcal{H}S$ is a hyperbolic end with ideal boundary canonically isomorphic to S and, in the case where S is developable, this hyperbolic end is canonically isomorphic to the extension of S constructed above. This completes the construction of the extension functor in the non-developable case. We now prove Theorem 3.1.8.

Proof of Theorem 3.1.8 Suppose that $\mathcal{H}S$ is not maximal. There exists a hyperbolic end X and an injective morphism $f : \mathcal{H}S \to X$ such that $\partial_\infty f : S \to \partial_\infty X$ is an isomorphism. In particular, $\partial_\infty X$, and therefore also X, has fundamental group Π. since $\partial_\infty X$ is diffeomorphic to S, Π is also the fundamental group of X. Lifting to the universal covers, f defines a Π-equivariant injective morphism $\tilde{f} : \mathcal{H}\tilde{S} \to \tilde{X}$ such that $\partial_\infty \tilde{f}$ is an isomorphism. Now let ϕ and ψ be respectively developing maps of $\tilde{S}$ and $\tilde{X}$ such that

$$\psi \circ \tilde{f} = \mathcal{H}\phi.$$

Since $\partial_\infty \tilde{f}$ is an isomorphism, by Lemma 3.3.23, there exists a unique injective morphism $g : \tilde{X} \to \mathcal{H}\tilde{S}$ such that $\mathcal{H}\phi \circ g = \psi$ and $\partial_\infty g \circ \partial_\infty \tilde{f} = \text{Id}$. In particular

$$\mathcal{H}\phi \circ (g \circ \tilde{f}) = \mathcal{H}\phi, \text{ and}$$

$$\partial_\infty (g \circ \tilde{f}) = \text{Id},$$

so that, by uniqueness,

$$g \circ \tilde{f} = \mathrm{Id}.$$

Since g is injective, it is also a right-inverse of $\tilde{f}$, so that $\tilde{f}$ is an isomorphism, and therefore so too is f. This proves maximality. Uniqueness is proven in a similar manner, and this completes the proof. □

3.4 Infinitesimally Strictly Convex Immersions

3.4.1 Infinitesimally Strictly Convex Immersions

We define an *immersed surface* in $\mathbb{H}^3$ to be a pair (S, e) where S is an *oriented* surface and $e : S \to \mathbb{H}^3$ is a smooth immersion. In what follows, we denote the immersed surface sometimes by S and sometimes by e, depending on which is more appropriate to the context. The family of immersed surfaces forms a category where the morphisms between two immersed surfaces (S, e) and (S, e') are those functions $\phi : S \to S'$ such that $e = e' \circ \phi$. Naturally, we identify two immersed surfaces which are isomorphic.

Let (S, e) be an immersed surface. In what follows, we will use the terminology of classical surface theory already described in Sect. 3.1.2. Recall that S is said to be *infinitesimally strictly convex* (ISC) whenever its second fundamental form is everywhere positive-definite. When this holds, every point x of S has a neighbourhood Ω over which e takes values on the boundary of some strictly convex subset X and N_e points *outwards* from this set. Recall also that S is said to be *quasicomplete* whenever the metric $\mathrm{I}_e + \mathrm{III}_e$ is complete. We now show that this is a natural requirement for studying ISC immersions in $\mathbb{H}^3$ in terms of hyperbolic ends. Indeed, denote $\mathcal{E}S := S \times]0, \infty[$ and define the function $\mathcal{E}e : \mathcal{E}S \to \mathbb{H}^3$ by

$$\mathcal{E}e(x, t) = \mathrm{Exp}(tN_e(x)), \qquad (3.4.1)$$

where Exp here denotes the exponential map of $\mathbb{H}^3$. By standard properties of convex surfaces in $\mathbb{H}^3$, $\mathcal{E}e$ is an immersion and we therefore furnish $\mathcal{E}S$ with the unique hyperbolic metric that makes it into a local isometry.

Lemma & Definition 3.4.1 *Let (S, e) be an ISC immersed surface in $\mathbb{H}^3$. $\mathcal{E}S$ is a hyperbolic end if and only if S is quasicomplete. When this holds, we call $\mathcal{E}S$ the end of S, its developing map is $\mathcal{E}e$ and its height function is*

$$h(x, t) := t. \qquad (3.4.2)$$

Proof It suffices to verify that h defines a height function over $\mathcal{E}S$ if and only if S is quasicomplete. By definition, h is smooth, its gradient flow lines are unit speed geodesics and, by standard properties of convex subsets of hyperbolic space, it is strictly convex. It thus remains only to study completeness. Choose $t > 0$ and let e_t denote the restriction of $\mathcal{E}e$ to $S \times \{t\}$. By classical hyperbolic geometry, the first fundamental form of this immersion is

$$\mathrm{I}_t := \cosh^2(t)\mathrm{I}_e + 2\cosh(t)\sinh(t)\mathrm{II}_e + \sinh^2(t)\mathrm{III}_e,$$

so that, by infinitesimal strict convexity,

$$\sinh^2(t)(\mathrm{I}_e + \mathrm{III}_e) \leq \mathrm{I}_t \leq 2\cosh^2(t)(\mathrm{I}_e + \mathrm{III}_e).$$

It follows that $\mathrm{I}_e + \mathrm{III}_e$ is complete if and only if I_t is complete. However, by convexity, $h^{-1}([t,\infty[)$ is complete if and only if I_t is complete. Since $t > 0$ is arbitrary, it follows that $h^{-1}([t,\infty[)$ is complete for all t if and only if S is quasicomplete, as desired. □

The operation $\mathcal{E}$ trivially sends morphisms between quasicomplete ISC immersed surfaces to morphisms between hyperbolic ends. Since $\mathcal{E}$ respects identity elements and compositions, it therefore defines a covariant functor between these two categories which we call the *end functor*.

There is also a natural way to associate a Möbius surface to every ISC immersed surface. Indeed, let (S, e) be an ISC immersed surface and let ϕ_e denote its asymptotic Gauss map. By infinitesimal strict convexity, ϕ_e is a local diffeomorphism from S into $\partial_\infty\mathbb{H}^3 = \hat{\mathbb{C}}$ and is thus the developing map of a Möbius structure over S. We denote $\mathcal{M}S := S$ and $\mathcal{M}e := \phi_e$, and we verify that $\mathcal{M}$ defines a covariant functor from the category of ISC immersed surfaces into the category of Möbius surfaces. However, this level of precision will be of little use to us and we will not use this terminology in other sections.

We have now reached a pivotal point of our construction. Indeed, we have associated two distinct hyperbolic ends to each quasicomplete ISC immersed surface, namely the end $\mathcal{E}S$ of S constructed above, and the hyperbolic end $\mathcal{HM}S$ obtained by applying the extension functor of Sect. 3.3.5 to the Möbius surface $\mathcal{M}S$. Furthermore, by Lemmas 3.3.23 and 3.3.24, $\mathcal{E}S$ naturally embeds into $\mathcal{HM}S$ in such a manner that the levels of $\mathcal{E}S$ are mapped to graphs in $\mathcal{HM}S$. Since e is smooth and ISC, a small modification of the proofs of these results extends this embedding to the boundary of $\mathcal{E}S$. In this manner, we obtain an embedding $\tilde{e}$ of S into $\mathcal{HM}S$ which factors the immersion e through the developing map $\mathcal{HM}\phi_e$.

This construction allows us to apply the machinery of Sects. 3.2 and 3.3 to the study of quasicomplete ISC immersions. Given its utility, we first extend it as follows. Let (S,ϕ) be a simply connected Möbius surface and let $(\mathcal{H}S, \mathcal{H}\phi)$ denote its extension. Let $\Omega \subseteq S$ be an open subset and let $e : \Omega \to \mathbb{H}^3$ be a quasicomplete ISC immersion whose Gauss map ϕ_e is equal to the restriction of ϕ to Ω. Let $(\mathcal{E}\Omega, \mathcal{E}e)$ denote the end of (Ω, e) and let $\psi : \mathcal{E}\Omega \to \mathcal{H}S$ denote the unique injective morphism such that

$$\mathcal{H}\phi \circ \psi = \mathcal{E}e, \text{ and} \tag{3.4.3}$$

$$\partial_\infty \psi = \text{Id}. \tag{3.4.4}$$

As before, a small modification of the proof of Lemma 3.3.23 shows that ψ extends to a smooth embedding from the whole of $\overline{\mathcal{E}\Omega} = \Omega \times [0, \infty[$ into $\mathcal{H}S$. We define the embedding $\tilde{e} : S \to \mathcal{H}S$ by

$$\tilde{e}(x) := \psi(x, 0), \tag{3.4.5}$$

and we call it the *canonical lift* of e. By (3.4.3), $\tilde{e}$ factors e through $\mathcal{H}\phi$ in the sense that

$$e = \mathcal{H}\phi \circ \tilde{e}. \tag{3.4.6}$$

Furthermore, an equally small modification of the proof of Lemma 3.3.24 then shows that the image of $\tilde{e}$ is also a graph over Ω in $\mathcal{H}S$, and we denote by $\text{Ext}(\tilde{e})$ the open subset of $\mathcal{H}S$ lying above this graph.

Let $\text{U}^+\mathcal{H}S$ denote the bundle of upward-pointing unit vectors over $\mathcal{H}S$. As in Sect. 3.1.2, we define the *horizon map* $\text{Hor} : \text{U}^+\mathcal{H}S \to \partial_\infty \mathcal{H}S = S$ such that, for every unit speed geodesic ray $\gamma : [0, \infty[\to \mathcal{H}S$ with $\dot{\gamma}(0) = \text{U}^+\mathcal{H}S$,

$$\text{Hor}(\dot{\gamma}(0)) := \lim_{t \to +\infty} \gamma(t). \tag{3.4.7}$$

This function is well-defined by Theorem 3.3.1. Let $N_{\tilde{e}} : S \to \text{U}^+\mathcal{H}S$ denote the positively-oriented unit normal vector field over $\tilde{e}$. We define the *asymptotic Gauss map* of $\tilde{e}$ by

$$\phi_{\tilde{e}} := \text{Hor} \circ N_{\tilde{e}}, \tag{3.4.8}$$

so that $\phi_{\tilde{e}}$ maps S into $\partial_\infty \mathcal{H}S$.

Lemma 3.4.2 *Let S be a developable Möbius surface of hyperbolic type with developing map ϕ and let $(\mathcal{H}S, \mathcal{H}\phi)$ denote its extension. Let $\Omega \subseteq S$ be an open subset and let $e : \Omega \to \mathbb{H}^3$ be a quasicomplete ISC immersion whose asymptotic Gauss map ϕ_e is equal to the restriction of ϕ to Ω. If $\tilde{e} : \Omega \to \mathcal{H}S$ denotes the canonical lift of e, then its asymptotic Gauss map $\phi_{\tilde{e}}$ satisfies*

$$\phi_{\tilde{e}} = \text{Id}. \tag{3.4.9}$$

Proof Indeed, let $(\mathcal{E}\Omega, \mathcal{E}e)$ denote the end of (Ω, e). Let $\psi : \overline{\mathcal{E}\Omega} \to \mathcal{H}S$ denote the unique injective morphism such that $\mathcal{H}\phi \circ \psi = \mathcal{E}e$ and $\partial_\infty \psi = \text{Id}$. For all $x \in S$,

$$\phi_{\tilde{e}}(x) = \lim_{t \to \infty} \psi(x, t) = \partial_\infty \psi(x) = x,$$

as desired. □

As a byproduct of the preceding construction, we obtain the following estimate for the Kulkarni–Pinkall metric of S in terms of the geometry of e.

Lemma 3.4.3 *Let S be a developable Möbius surface with developing map ϕ and Kulkarni–Pinkall metric g_ϕ. Let $\Omega \subseteq S$ be an open subset and let $e : \Omega \to \mathbb{H}^3$ be a quasicomplete ISC immersion whose asymptotic Gauss map ϕ_e is equal to the restriction of ϕ to Ω. If I_e, II_e and III_e denote respectively the first, second and third fundamental forms of e then, over Ω,*

$$g_\phi \leq \mathrm{I}_e + 2\mathrm{II}_e + \mathrm{III}_e. \tag{3.4.10}$$

Remark 3.4.4 The right hand side of (3.4.10) is the none other than the horospherical metric studied by Schlenker in [22].

Proof Let $(\mathcal{H}S, \mathcal{H}\phi)$ denote the extension of (S, ϕ). Let $(\mathcal{E}\Omega, \mathcal{E}e)$ denote the end of (Ω, e) and let h denote its height function. Let $\psi : \overline{\mathcal{E}\Omega} \to \mathcal{H}S$ denote the unique injective morphism such that $\mathcal{H}\phi \circ \psi = \mathcal{E}e$ and $\partial_\infty \psi = \mathrm{Id}$. Let x be a point of Ω. Let (H, α) denote the half space in $\overline{\mathcal{E}\Omega}$ such that $(x, 0) \in \partial\alpha(H)$ and $\partial_t = \nabla h(x)$ is the inward-pointing normal to this surface at this point. Denote $P := \partial H$ and $D := \partial_\infty H$. Let g' denote the Poincaré metric of D and let $\phi_0 : P \to D$ denote the asymptotic Gauss map of P. Observe that ϕ_0 is an isometry from P into (D, g').

Let Hor denote the horizon map of $\mathrm{U}\mathbb{H}^3$. Let $N_e : S \to \mathrm{U}\mathbb{H}^3$ denote the positively-oriented unit normal vector field of S. Denote $\nu := N_e(x)$ and observe that, by the chain rule,

$$D\phi(x) = D\mathrm{Hor}(\nu) \circ DN_e(x).$$

Recall now that $\mathrm{T}_\nu\mathrm{U}\mathbb{H}^3$ decomposes as

$$\mathrm{T}_\nu\mathrm{U}\mathbb{H}^3 = H_\nu \oplus V_\nu, \tag{3.4.11}$$

where V_ν is the vertical subspace and H_ν is the horizontal subspace of the Levi–Civita covariant derivative. Recall furthermore that $D\pi(\nu)$ maps H_ν isomorphically onto $T_y\mathbb{H}^3$ and that there exists a natural projection $p_\nu : V_\nu \to \langle\nu\rangle^\perp$. We henceforth identify vectors in H_ν and V_ν with their respective images under $D\pi(\nu)$ and p_ν. With respect to the decomposition (3.4.11), for all $\xi \in T_xS$,

$$DN_e(x) \cdot \xi = (De(x) \cdot \xi, De(x) \cdot A_e(x) \cdot \xi),$$

where A_e here denotes the shape operator of e.

Since P is totally geodesic, for all $\xi \in \langle\nu\rangle^\perp$,

$$D\mathrm{Hor}(\nu) \cdot (\xi, 0) = D\phi_0(y) \cdot \xi.$$

Using the fact that Hor restricts to a conformal diffeomorphism from $\mathrm{U}_y\mathbb{H}^3$ into $\partial_\infty\mathbb{H}^3$, we likewise show that, for all $\xi \in \langle \nu\rangle^\perp$,

$$D\mathrm{Hor}(\nu)\cdot(0,\xi) = D\phi_0(y)\cdot\xi.$$

Combining the above relations, it follows that, for all $\xi \in T_xS$,

$$D\phi(x)\cdot\xi = D\phi_0(y)\cdot De(x)\cdot(\xi + A_e(x)\cdot\xi).$$

Since ϕ_0 is an isometry, it follows that

$$(\phi^*g')(x) = \mathrm{I}_e + 2\mathrm{II}_e + \mathrm{III}_e. \tag{3.4.12}$$

However, by definition of the Kulkarni–Pinkall metric,

$$g(x) \leq (\phi^*g')(x),$$

and the result follows. □

The proof of Lemma 3.4.3 also yields the following useful result.

Lemma 3.4.5 *Let (S, e) be an immersed surface in $\mathbb{H}^3$ and let I_e, II_e and III_e denote respectively its first, second and third fundamental forms. The asymptotic Gauss map ϕ_e of e is conformal with respect to the non-negative semi-definite bilinear form $\mathrm{I}_e + 2\mathrm{II}_e + \mathrm{III}_e$.*

Proof Indeed, this follows from (3.4.12) since, with the notation of the proof of Lemma 3.4.3, g' is a conformal metric over D. □

3.4.2 A Priori Estimates

We are now ready to derive our main a priori estimates for quasicomplete ISC immersed surfaces in $\mathbb{H}^3$. First, for every open half-space H in $\mathbb{H}^3$, denote

$$H_r := \{x \in H \mid d(x, \partial H) \geq r\}. \tag{3.4.13}$$

Lemma 3.4.6 *Let S be a developable Möbius surface with developing map ϕ. Let $(\mathcal{H}S, \mathcal{H}\phi)$ denote its extension, let $(D_i, \alpha_i)_{i\in I}$ denote its half-space decomposition and, for all i, let H_i denote the open half-space in $\mathbb{H}^3$ with ideal boundary D_i. Let Ω be an open subset of S and let $e : \Omega \to \mathbb{H}^3$ be a quasicomplete ISC immersion whose asymptotic Gauss map ϕ_e is equal to the restriction of ϕ to Ω. Let $\tilde{e} : \Omega \to \mathcal{H}S$ denote the canonical lift of e and let $\mathrm{Ext}(\tilde{e})$ denote the subset of $\mathcal{H}S$ lying above $\tilde{e}(\Omega)$. For all $r > 0$, if the extrinsic curvature of e satisfies*

$$K_e \leq \tanh(r)^2, \tag{3.4.14}$$

then, for all i such that $\alpha_i(D_i) \subseteq \Omega$,

$$\mathcal{H}\alpha_i(H_{i,r}) \subseteq \mathrm{Ext}(\tilde{e}). \tag{3.4.15}$$

Proof Let i be an element of I such that $D_i \subseteq \Omega$. Let $j \in I$ be another element such that $\overline{D}_j \subseteq D_i$ and $\alpha_j = \alpha_i|_{D_j}$. Since $\tilde{e}(\Omega)$ is a graph over Ω, for sufficiently large s,

$$\mathcal{H}\alpha_j(H_{j,s}) \subseteq \mathrm{Ext}(\tilde{e}).$$

We claim that this holds for all $s > \tanh(r)$. Indeed, suppose the contrary, and let $s_0 > \tanh(r)$ be the infimal value of s for which this relation holds. In particular, the surface $\partial\mathcal{H}\alpha_j(H_{j,s_0})$ is an exterior tangent to $\tilde{e}(\Omega)$ at some point. However, since $\partial\mathcal{H}\alpha_j(H_{j,s_0})$ has constant extrinsic curvature equal to $\tanh(s_0)^2$, this yields a contradiction by the geometric maximal principle. The result follows upon letting D_j tend to D_i. □

Theorem 3.4.7 *Let S be a developable Möbius surface with developing map ϕ. Let $(D_i, \alpha_i)_{i\in I}$ denote its Möbius disk decomposition and, for all i, let H_i denote the open half-space in $\mathbb{H}^3$ with ideal boundary D_i. Let Ω be an open subset of S and let $e : \Omega \to \mathbb{H}^3$ be a quasicomplete ISC immersion whose asymptotic Gauss map ϕ_e is equal to the restriction of ϕ to Ω. For all $r > 0$, if the extrinsic curvature of e satisfies*

$$K_e \leq \tanh(r)^2, \tag{3.4.16}$$

then, for all $x \in \Omega$ and for all $i \in I$,

$$x \in \alpha_i(D_i) \subseteq \Omega \ \Rightarrow e(x) \notin H_{i,r}. \tag{3.4.17}$$

Proof Let $(\mathcal{H}S, \mathcal{H}\phi)$ denote the extension of S, let $(\mathcal{E}\Omega, \mathcal{E}e)$ denote the end of (Ω, e), and let $\psi : \overline{\mathcal{E}\Omega} \to \mathcal{H}S$ denote the unique injective morphism such that $\mathcal{H}\phi \circ \psi = \mathcal{E}e$ and $\partial_\infty\psi = \mathrm{Id}$. Let x be a point of Ω, and let i be an element of I such that $x \in \alpha_i(D_i) \subseteq \Omega$. Define $\gamma(t) := \mathcal{E}e(x, t)$ and $\mu(t) := \psi(x, t)$, so that $\gamma = \mathcal{H}\phi \circ \mu$. Observe that μ is the geodesic ray in $\mathcal{H}S$ normal to $\tilde{e}(\Omega)$ at $\tilde{e}(x)$. In particular,

$$\lim_{t\to+\infty}\mu(t) = \partial_\infty\psi(x) = x \in \alpha_i(D_i).$$

Since, by Lemma 3.4.6, $\tilde{e}(\Omega)$ lies outside $\mathcal{H}\alpha_i(H_{i,r})$, it follows by the intermediate value theorem that μ crosses $\partial\mathcal{H}\alpha_i(H_{i,r})$ at some point. Furthermore, by convexity, μ crosses this surface transversally from the outside to the inside. Composing with $\mathcal{H}\phi$, it follows that γ crosses $\partial H_{i,r}$ transversally from the outside to the inside at some point. However, since $\partial H_{i,r}$ is strictly convex, γ can meet this surface no more than once, so that $e(x) = \gamma(0)$ lies outside $H_{i,r}$, as desired. □

Theorem 3.4.8 *Let S be a developable Möbius surface with developing map ϕ and Kulkarni–Pinkall form ω_ϕ. Let Ω be an open subset of S and let $e : \Omega \to \mathbb{H}^3$ be a quasicomplete ISC immersion whose asymptotic Gauss map ϕ_e is equal to the restriction of ϕ to Ω. For all $x \in S$,*

$$e(x) \in \overline{B}(\phi_*\omega_\phi(x)), \tag{3.4.18}$$

where B here denotes the parametrisation of the space of open horoballs in $\mathbb{H}^3$ by $\Lambda^2\partial_\infty\mathbb{H}^3$ described in Sect. 3.1.2.

Proof Let $(\mathcal{E}\Omega, \mathcal{E}e)$ denote the end of (Ω, e). Observe that the level set foliation of $(\mathcal{E}\Omega, \mathcal{E}e)$ is $(\Omega\times\{t\})_{t>0}$ so that every level of this hyperbolic end as well as its ideal boundary $\partial_\infty\mathcal{E}\Omega$ naturally identifies with Ω. With respect to these identifications, $\partial_\infty\mathcal{E}e = \phi$ and the vertical line projection $\pi_\infty : \mathcal{E}\Omega \to \partial_\infty\mathcal{E}\Omega$ is given by $\pi_\infty(x, t) = x$. Thus, if ω denotes the Kulkarni–Pinkall form of (Ω, ϕ) then, by Lemma 3.3.17, for all $x \in \Omega$ and for all $t > 0$,

$$\mathcal{E}e(x, t) \in B(\phi_*\omega(x)).$$

However, by Lemma 3.2.6, $\omega \geq \omega_\phi$, so that, for all such x and t,

$$\mathcal{E}e(x, t) \in B(\phi_*\omega_\phi(x)),$$

and the result now follows upon letting t tend to zero. □

3.4.3 Cheeger–Gromov Convergence

In order for the text to be as self-contained as possible, we now recall the basic theory of Cheeger–Gromov convergence. A *pointed Riemannian manifold* is a triplet (X, g, x), where X is a smooth manifold, g is a Riemannian metric and x is a point of X. We say that a sequence $(X_m, g_m, x_m)_{m\in\mathbb{N}}$ of complete pointed Riemannian manifolds converges to the complete pointed Riemannian manifold $(X_\infty, g_\infty, x_\infty)$ in the *Cheeger–Gromov sense* whenever there exists a sequence $(\Phi_m)_{m\in\mathbb{N}}$ of functions such that

(1) for all m, $\Phi_m : X_\infty \to X_m$ and $\Phi_m(x_\infty) = x_m$; and
    for every relatively compact open subset Ω of X_∞, there exists M such that
(2) for all $m \geq M$, the restriction of Φ_m to Ω defines a smooth diffeomorphism onto its image; and
(3) the sequence $((\Phi_m|_\Omega)^* g_m)_{m\geq M}$ converges to $g_\infty|_\Omega$ in the C^∞_{loc} sense.

We call $(\Phi_m)_{m\in\mathbb{N}}$ a sequence of *convergence maps* of $(X_m, g_m, x_m)_{m\in\mathbb{N}}$ with respect to $(X_\infty, g_\infty, x_\infty)$.

At first sight, the concept of Cheeger–Gromov convergence can appear rather daunting and, indeed, its correct usage can be sometimes counterintuitive. However, it is reassuring to observe that it defines a Hausdorff topology over the space of isometry equivalence classes of complete pointed Riemannian manifolds.[4] Furthermore, although the convergence maps are trivially non-unique, any two sequences $(\Phi_m)_{m\in\mathbb{N}}$ and $(\Phi'_m)_{m\in\mathbb{N}}$ of convergence maps are equivalent in the sense that there exists an isometry $\Psi : X_\infty \to X_\infty$ preserving x_∞ such that, for any two relatively compact open subsets $U \subseteq \overline{U} \subseteq V$ of X_∞, there exists M such that

(1) for all $m \geq M$, the respective restrictions of Φ_m and $\Phi'_m \circ \Psi$ to U and V define smooth diffeomorphisms onto their images;
(2) for all $m \geq M$, $(\Phi'_m \circ \Psi)(U) \subseteq \Phi_m(V)$; and
(3) the sequence $((\Phi_m|_V)^{-1} \circ \Phi'_m \circ \Psi)_{m\in\mathbb{N}}$ converges in the C^∞ sense to the identity map over U.

The concept of Cheeger–Gromov convergence applies to sequences of immersed submanifolds as follows. We say that a sequence $(S_m, x_m, \phi_m)_{m\in\mathbb{N}}$ of complete pointed immersed submanifolds in a complete Riemannian manifold (X, g) converges to the complete pointed immersed submanifold $(S_\infty, x_\infty, \phi_\infty)$ in the *Cheeger–Gromov sense* whenever $(S_m, x_m, \phi_m^* g)_{m\in\mathbb{N}}$ converges to $(S_\infty, x_\infty, \phi_\infty^* g)$ in the Cheeger–Gromov sense and, for one, and therefore for any, sequence $(\Phi_m)_{m\in\mathbb{N}}$ of convergence maps, the sequence $(\phi_m \circ \Phi_m)_{m\in\mathbb{N}}$ converges to ϕ_∞ in the C^∞_{loc} sense.

Cheeger–Gromov convergence of immersed submanifolds can also be described in terms of graphs. Indeed, let NS_∞ denote the normal bundle of (S_∞, ϕ_∞) in $\phi_\infty^* TX$. Recall that the exponential map of X defines a smooth function $\text{Exp} : NS_\infty \to X$. In particular, given a sufficiently small smooth section $f : \Omega \to NS_\infty$ defined over an open subset Ω of S_∞, the composition $\text{Exp} \circ f$ defines a smooth immersion of Ω in X which we call the *graph* of f. It is straightforward to show that if the sequence $(S_m, x_m, \phi_m)_{m\in\mathbb{N}}$ converges to $(S_\infty, x_\infty, \phi_\infty)$ in the Cheeger–Gromov sense, then there exists a sequence $(x'_m)_{m\in\mathbb{N}}$ of points in S_∞ and sequences of functions $(f_m)_{m\in\mathbb{N}}$ and $(\alpha_m)_{m\in\mathbb{N}}$ such that

(1) $(x'_m)_{m\in\mathbb{N}}$ converges to x_∞;
(2) for all m, f_m maps S_∞ into NS_∞, α_m maps S_∞ into S_m and $\alpha_m(x'_m) = x_m$; and for every relatively compact open subset Ω of S_∞, there exists M such that
(3) for all $m \geq M$, f_m restricts to a smooth section of NS_∞ over Ω, α_m restricts to a smooth diffeomorphism of Ω onto its image, and

$$\text{Exp} \circ f_m|_\Omega = \phi_m \circ \alpha_m|_\Omega; \text{ and}$$

(4) the sequence $(f_m)_{m\geq M}$ tends to zero in the C^∞_{loc} sense.

[4] Strictly speaking, of course, the family of all complete pointed Riemannian manifolds is not a set. However, by Whitney's theorem, we restrict attention to the family of submanifolds of $\mathbb{R}^m$, for some m, with smooth complete metrics defined over them, which is a set.

Finally, these definitions are readily extended in a number of ways. For example, in the case of a sequence $(X_m, g_m, x_m)_{m\in\mathbb{N}}$ of pointed Riemannian manifolds, the hypothesis of completeness is unnecessary. Instead, it is sufficient to assume that for all $R > 0$, there exists M such that, for all $m \geq M$, the closed ball of radius R about x_m in (X_m, g_m) is compact. Likewise, in the case of immersed submanifolds, the target space can be replaced with a sequence $(X_m, g_m, x_m)_{m\in\mathbb{N}}$ of pointed Riemannian manifolds converging in the Cheeger–Gromov sense to some complete pointed Riemannian manifold. Furthermore, it is not necessary to suppose that the Riemannian manifolds in this sequence are complete, and so on.

3.4.4 Labourie's Theorems and Their Applications

For $k > 0$, we define a *k-surface* to be a quasicomplete ISC immersed surface in $\mathbb{H}^3$ of constant extrinsic curvature equal to k. Let (S, e) be a k-surface in $\mathbb{H}^3$. Let N_e denote its positively-oriented unit normal vector field. Observe that, if I_e and III_e denote respectively the first and third fundamental forms of e, then $\mathrm{I}_e + \mathrm{III}_e$ is uniformly equivalent to the pull-back through N_e of the Sasaki metric of $\mathrm{U}\mathbb{H}^3$ so that, by quasicompleteness, N_e is actually a complete immersion of S in $\mathrm{U}\mathbb{H}^3$. In order to emphasise our interest in this function as an *immersion* rather than as a *vector field*, following Labourie, we denote $\hat{e} := N_e$ and we call $\hat{e}$ the *Gauss lift* of e. In [18], Labourie proves the following result (see also [28]).

Theorem 3.4.9 (Labourie's Compactness Theorem) *Choose $k > 0$. Let $(S_m, e_m, x_m)_{m\in\mathbb{N}}$ be a sequence of pointed k-surfaces in $\mathbb{H}^3$. For all m, let $\hat{e}_m : S_m \to \mathrm{U}\mathbb{H}^3$ denote the Gauss lift of e_m. If $(\hat{e}_m(x_m))_{m\in\mathbb{N}}$ remains within a compact subset of $\mathrm{U}\mathbb{H}^3$ then there exists a complete, pointed immersed surface $(S_\infty, \hat{e}_\infty, x_\infty)$ in $\mathrm{U}\mathbb{H}^3$ towards which $(S_m, \hat{e}_m, x_m)_{m\in\mathbb{N}}$ subconverges in the Cheeger–Gromov sense.*

Significantly, Theorem 3.4.9 does not affirm that the limit is a lift of some k-surface. In order to address this, Labourie introduces what he calls curtain surfaces, which are defined as follows. Given a complete geodesic Γ in $\mathbb{H}^3$, we denote by $\mathrm{N}\Gamma \subseteq U\mathbb{H}^3$ the set of unit normal vectors over Γ. Observe that $\mathrm{N}\Gamma$ is an immersed surface conformally equivalent to $\mathbb{S}^1 \times \mathbb{R}$ with respect to the Sasaki metric of $\mathrm{U}\mathbb{H}^3$. We define a *curtain surface* to be any immersed surface $(S, \hat{e})$ in $\mathrm{U}\mathbb{H}^3$ which is a cover of $\mathrm{N}\Gamma$, for some complete geodesic Γ.

Theorem 3.4.10 (Labourie's Dichotomy) *Choose $k > 0$. Let $(S_\infty, \hat{e}_\infty)$ be a limit of a sequence of lifts of k-surfaces, as in Theorem 3.4.9. If $(S_\infty, \hat{e}_\infty)$ is not a curtain surface, then $(S_\infty, \pi \circ \hat{e}_\infty)$ is a k-surface.*

The phenomenon described in Theorem 3.4.10 is best illustrated by the case where $k = 1$. Indeed, by a theorem of Volkov–Vladimirova and Sasaki (see Theorem 46 of [29]), the only 1-surfaces in $\mathbb{H}^3$ are the horospheres and covers of equidistant cylinders around complete geodesics.[5] Let Γ be a complete geodesic in $\mathbb{H}^3$. For all $r > 0$, let C_r denote the cylinder of radius r about Γ. For all m, let $e_m : \mathbb{R}^2 \to C_{1/m}$ be a covering map which is isometric with respect to the sum of its first and third fundamental forms. The sequence $(\mathbb{R}^2, e_m, 0)_{m\in\mathbb{N}}$ subconverges in the C^∞_{loc} sense to a smooth function $e_\infty : \mathbb{R}^2 \to \Gamma$. Consequently, viewed as a sequence of pointed immersed surfaces, this sequence degenerates. However, the sequence $(\mathbb{R}^2, \hat{e}_m, 0)_{m\in\mathbb{N}}$ of Gauss lifts converges in the C^∞_{loc} sense to a cover of $\text{N}\Gamma$, that is, a curtain surface. Labourie's dichotomy affirms that, even in the general case, this is the only mode of degeneration that can occur.

We now prove one of the main results of this chapter.

Theorem 3.4.11 (Monotone Convergence) *Let S be a developable Möbius surface of hyperbolic type with universal cover not isomorphic to* $(\mathbb{C}, \text{Exp}(z))$. *Let ϕ be a developing map of S. Let $(\Omega_m)_{m\in\mathbb{N}}$ be a nested sequence of open subsets of S which exhausts S. For $k > 0$ and for all m, let $e_m : \Omega_m \to \mathbb{H}^3$ be a k-surface whose Gauss map ϕ_m is equal to the restriction of ϕ to Ω. There exists a k-surface $e_\infty : S \to \mathbb{H}^3$ towards which $(e_m)_{m\in\mathbb{N}}$ subconverges in the C^∞_{loc} sense over S.*

Proof For all m, let $\hat{e}_m$ denote the Gauss lift of e_m. Let ω_ϕ denote the Kulkarni–Pinkall form of S. Let x be a point of S. We claim that there exists a Möbius disk (D, α) in S such that $x \in \alpha(D)$ and $(\hat{e}_m \circ \alpha)_{m\in\mathbb{N}}$ subconverges in the C^∞_{loc} sense. Indeed, let (D', α') be a Möbius disk in S such that $x \in \alpha'(D')$ and the closure of $\alpha'(D')$ in S is compact. By Theorems 3.4.7 and 3.4.8, for all sufficiently large m,

$$e_m(x) \in K := \overline{B}(\phi_*\omega_\phi(x)) \setminus \mathcal{H}\alpha'(H'_r),$$

where H' here denotes the open half-space in $\mathbb{H}^3$ with ideal boundary D' and $r := \text{arctanh}(\sqrt{k})$. Since K is compact, it follows by Theorem 3.4.9 that there exists a complete pointed immersed surface $(S_\infty, \hat{e}_\infty, x_\infty)$ in $\text{U}\mathbb{H}^3$ towards which $(\Omega_m, \hat{e}_m, x)_{m\in\mathbb{N}}$ subconverges in the Cheeger–Gromov sense. Denote

$$\phi_\infty := \text{Hor} \circ \hat{e}_\infty,$$

where Hor here denotes the horizon map of $\text{U}\mathbb{H}^3$. Since ϕ_∞ is a local diffeomorphism from S into $\partial_\infty\mathbb{H}^3$, it defines a developable Möbius structure over S_∞. Let $(\Phi_m)_{m\in\mathbb{N}}$ be a sequence of convergence maps of $(\Omega_m, \hat{e}_m, x)_{m\in\mathbb{N}}$ with respect to $(S_\infty, \hat{e}_\infty, x_\infty)$. Let (D'', α'') be a Möbius disk about x_∞ in (S_∞, ϕ_∞) such that $\alpha''(D'')$ is relatively compact in S_∞. Let M be such that, for all $m \geq M$, Φ_m restricts to a smooth diffeomorphism from $\alpha''(D'')$ onto an open subset U_m of S. Since $(\phi_m \circ \Phi_m \circ \alpha'')_{m\geq M}$ converges in the C^∞_{loc} sense over D'' to $(\phi_\infty \circ \alpha'') = \text{Id}$, upon increasing M and reducing D'' if necessary, we may suppose that, for all $m \geq M$, $(\phi_m \circ \Phi_m \circ \alpha'')$ is a diffeomorphism onto its image Ω_m whose inverse we denote by β_m. Let D be another disk with closure contained in D'' such that

[5] We refer the reader to Footnote 4 of Sect. 3.1.3.

$x_\infty \in \alpha''(D)$. For sufficiently large m, Ω_m contains D and we therefore define $\alpha_m : D \to S$ by $\alpha_m := (\Phi_m \circ \alpha'' \circ \beta_m)$. For all such m, (D, α_m) is in fact a Möbius disk in (S, ϕ) and, upon increasing m further if necessary, we may suppose in addition that $\alpha_m(D_m)$ contains x. It then follows that α_m is independent of m, and we therefore denote $\alpha := \alpha_m$. By construction $(\hat{e}_m \circ \alpha)_{m\geq M}$ subconverges to $(\hat{e}_\infty \circ \alpha'')$, and (D, α) is therefore the desired Möbius disk.

A diagonal argument now shows that $(\hat{e}_m)_{m\in\mathbb{N}}$ subconverges in the C^∞_{loc} sense to a smooth immersion $\hat{e}_\infty : S \to \mathrm{U}\mathbb{H}^3$ satisfying

$$\mathrm{Hor} \circ \hat{e}_\infty = \phi_\infty.$$

We now claim that $\hat{e}_\infty$ is complete. Indeed, for all $m \in \mathbb{N}$, if I_m, II_m and III_m denote respectively the first, second and third fundamental forms of e_m, then, by Lemma 3.4.3,

$$\mathrm{I}_m + \mathrm{III}_m \geq \frac{1}{2}(\mathrm{I}_m + 2\mathrm{II}_m + \mathrm{III}_m) \geq g_\phi,$$

where g_ϕ here denotes the Kulkarni–Pinkall metric of (S, ϕ). However, for all m, $\mathrm{I}_m + \mathrm{III}_m$ is also the metric of $\hat{e}_m$. It follows upon taking limits that the metric of $\hat{e}_\infty$ is bounded below by g_ϕ. However, by Lemma 3.2.15, g_ϕ is complete, and therefore so too is $\hat{e}_\infty$, as asserted. Finally, by Labourie's dichotomy, either $e_\infty := \pi \circ \hat{e}_\infty$ is a k-surface, or $\hat{e}_\infty$ is a curtain surface. Since the latter can only occur when (S, ϕ) is isomorphic to a cover of $(\mathbb{C}^*, z)$, that is, when its universal cover is isomorphic to $(\mathbb{C}, \mathrm{Exp}(z))$, it follows that e_∞ is a k-surface, and this completes the proof. □

3.4.5 Uniqueness and Existence

We are now ready to prove the main result of this chapter. Before proceeding, we require the following technical lemma.

Lemma 3.4.12 *Let (S, e) be an ISC immersion in $\mathbb{H}^3$ and let $\hat{e} : S \to \mathrm{U}\mathbb{H}^3$ denote its Gauss lift. If, for $0 < k < 1$, e has constant extrinsic curvature equal to k then, for all $t > 0$, the immersion $e_t(x) := \mathrm{Exp}(t\hat{e}(x))$ has curvature at every point strictly greater than k and strictly less than 1.*

Proof Indeed, by the tube formula (see [12]), the shape operator A_t of e_t solves

$$\dot{A}_t = \mathrm{Id} - \mathrm{A}_t^2.$$

Thus, denoting $H_t := \mathrm{Tr}(A_t)$ and $K_t := \mathrm{Det}(A_t)$, we have

$$\frac{\partial}{\partial t} K_t = \mathrm{Tr}(A_t^{-1} - A_t) = \frac{1}{K_t}(1 - K_t)H_t.$$

Solving this ordinary differential equation with $K_0 = k < 1$ yields, for all $t > 0$,

$$k < K_t < 1,$$

as desired. □

The following result is proven by Labourie in [19]. Since its proof fits into the framework developed in this chapter, we include it for completeness.

Theorem 3.4.13 (Monotonicity) *Let S be a developable Möbius surface with developing map ϕ and let $(\mathcal{H}S, \mathcal{H}\phi)$ denote its extension. For $0 < k < 1$, and for $i \in \{1, 2\}$, let $\Omega_i \subseteq S$ be an open subset of S, let $e_i : \Omega_i \to \mathbb{H}^3$ be a k-surface whose asymptotic Gauss map ϕ_i is equal to the restriction of ϕ to Ω_i, and for each i, let $\tilde{e}_i : \Omega_i \to \mathcal{H}S$ denote the canonical lift of e_i. If $\Omega_1 \subseteq \Omega_2$ then* $\mathrm{Ext}(\tilde{e}_1) \subseteq \mathrm{Ext}(\tilde{e}_2)$.

Proof Suppose the contrary. Let U denote the set of all points in Ω_1 whose image under $\tilde{e}_1$ lies in the complement of the closure of $\mathrm{Ext}(\tilde{e}_2)$. For each i, let $(\mathcal{E}\Omega_i, \mathcal{E}e_i)$ denote the end of (Ω_i, e_i) and let $\psi_i : \overline{\mathcal{E}\Omega_i} \to \mathcal{H}S$ denote the unique injective morphism such that $\mathcal{H}\phi \circ \psi_i = \mathcal{E}e_i$ and $\partial_\infty \psi_i = \mathrm{Id}$. Denote $r := \arctan(\sqrt{k})$. We claim that there exists a unique smooth function $f : U \to [0, r]$ and a unique function $\alpha : U \to \Omega_2$ such that α is a diffeomorphism onto its image and, for all $x \in U$,

$$(\tilde{e}_2 \circ \alpha)(x) = \psi_i(x, f(x)).$$

Indeed, let x be a point of U. Let (H, α) denote the unique half space in $\overline{\mathcal{E}\Omega_1}$ such that $(x, 0) \in \partial\alpha(H)$ and let D denote the ideal boundary of H. $(H, \psi_1 \circ \alpha)$ is the unique half-space in $\mathcal{H}S$ which is tangent to $\tilde{e}_1(\Omega)$ at $\tilde{e}_1(x)$. Since

$$\partial_\infty(\psi_1 \circ \alpha)(D) = (\partial_\infty \psi_1 \circ \partial_\infty \alpha)(D) = \partial_\infty \alpha(D) \subseteq \Omega_1 \subseteq \Omega_2,$$

it follows by Lemma 3.4.6 that

$$(\psi_1 \circ \alpha)(H_r) \subseteq \mathcal{H}(\tilde{e}_2).$$

In particular, for all $t > r$,

$$\psi_1(x, t) \in \mathcal{H}(\tilde{e}_2).$$

We define $f(x)$ to be the infimal value of t such that $\psi_1(x, t) \in \tilde{e}_2(\Omega_2)$. By Theorem 3.3.1, this is the only value of t such that $\psi_1(x, t) \in \tilde{e}_2(\Omega_2)$. Since, by convexity, the geodesic $t \mapsto \psi_1(x, t)$ is transverse to $\tilde{e}_2(\Omega_2)$ at this point, it follows that f is smooth. The existence of α now follows from the fact that $\tilde{e}_2$ is an embedding, and this proves the assertion.

We now apply a maximum principle at infinity to obtain a contradiction. Let $(x_m)_{m\in\mathbb{N}}$ be a sequence in U such that

$$\lim_{m\to\infty} f(x_m) = f_0 := \sup_{x\in U} f(x) \leq r,$$

and, for all m, denote $y_m := \alpha(x_m)$. Let p be a fixed point of $\mathbb{H}^3$. For all m, let β_m be an isometry of $\mathbb{H}^3$ such that $\beta_m(e_1(x_m)) = p$ and, for each i, denote $e_{i,m} := \alpha_m \circ e_i$ and let $\hat{e}_{i,m}$ denote its Gauss lift. By Theorem 3.4.9, we may suppose that there exist pointed immersions $(S_{1,\infty}, \hat{e}_{1,\infty}, x_\infty)$ and $(S_{2,\infty}, \hat{e}_{2,\infty}, y_\infty)$ towards which $(\Omega_1, \hat{e}_{1,m}, x_m)_{m\in\mathbb{N}}$ and $(\Omega_2, \hat{e}_{2,m}, y_m)_{m\in\mathbb{N}}$ subconverge in the Cheeger–Gromov sense. Observe that there exist neighbourhoods U of x_∞ in $S_{1,\infty}$, V of y_∞ in $S_{2,\infty}$, a smooth diffeomorphism $\alpha : U \to V$ and a smooth function $f : U \to [0, r]$ such that $\alpha(x_\infty) = y_\infty$, f attains its maximum value of f_0 at x_∞ and, for all $z \in U$,

$$(\hat{e}_{2,\infty} \circ \alpha)(z) = \mathrm{Exp}(f(z)\hat{e}_{1,\infty}(z)). \tag{3.4.19}$$

We now show that this is absurd. Define $e_{1,\infty} : S_{2,\infty} \to \mathbb{H}^3$ by $e_{1,\infty}(z) := \mathrm{Exp}(f_0\hat{e}_{1,\infty})$. We claim that the extrinsic curvature of this immersion is at every point strictly greater than k. Indeed, there are two cases to consider. If $\hat{e}_{1,\infty}$ is a curtain surface, then e_1' is a cylinder of radius f_0 about a complete geodesic in $\mathbb{H}^3$ and thus has constant extrinsic curvature equal to 1. On the other hand, if $e_{1,\infty}$ is the lift of a k-surface, then, by Lemma 3.4.12, $\hat{e}_1'$ also has extrinsic curvature at every point strictly greater than k, and the assertion follows. We now examine the function $e_{2,\infty} := \pi \circ \hat{e}_{2,\infty}$. Once again, there are two cases to consider. If $\hat{e}_{2,\infty}$ is a curtain surface, then $e_{2,\infty}(S_{2,\infty})$ is a complete geodesic Γ in $\mathbb{H}^3$ which, by (3.4.19), is an interior tangent to $e_{1,\infty}$ at $e_{1,\infty}(x_\infty)$. By convexity, this is absurd. Otherwise, $e_{2,\infty}$ is a k-surface which is an interior tangent to $e_{1,\infty}$ at $e_{1,\infty}(x_\infty)$, which is also absurd by the geometric maximum principle. We thus obtain a contradiction in all cases, and this completes the proof. □

Theorem 3.4.13 is useful for studying the geometry of k-surfaces in $\mathbb{H}^3$. In the present section, when $\Omega_1 = \Omega_2$, it yields uniqueness.

Theorem 3.4.14 *For all $0 < k < 1$ and for every developable Möbius surface S with developing map ϕ, there exists at most one k-surface $e : S \to \mathbb{H}^3$ such that $\phi_e = \phi$.*

Proof Indeed, let $e, e' : S \to \mathbb{H}^3$ be k-surfaces such that $\phi_e = \phi_{e'} = \phi$. Let $(\mathcal{H}S, \mathcal{H}\phi)$ denote the extension of (S, ϕ). Let $\tilde{e}, \tilde{e}' : S \to \mathcal{H}S$ denote the respective canonical lifts of e and e'. By Theorem 3.4.13, $\tilde{e}(S) = \tilde{e}'(S)$. From this it follows that $\tilde{e} = \tilde{e}'$ and so $e = e'$, as desired. □

The following result is proven by Labourie in [19].

Theorem 3.4.15 *Let S be a developable Möbius surface with developing map ϕ. Let Ω be a relatively compact open subset of S with smooth boundary. For all $0 < k < 1$, there exists a k-surface $e : \Omega \to \mathbb{H}^3$ such that $\phi_e = \phi|_\Omega$.*

Sketch of Proof Labourie's result holds for asymptotic Plateau problems in Cartan-Hadamard manifolds of bounded geometry. For the reader's convenience, we sketch a simpler proof valid for the hyperbolic case studied here. First, using the Beltrami–Klein model we identify $\mathbb{H}^3$ with the unit ball in $\mathbb{R}^3$. For all $r \in]0,1[$, let $\overline{B}_r$ denote the closed ball of (Euclidean) radius r about the origin, let $S_r := \partial\overline{B}_r$ denote its boundary, let $\pi_r : \partial_\infty\mathbb{H}^3 \to S_r$ denote the canonical projection, and denote $\phi_r := \pi_r \circ \phi$. For all r, the argument used in [13] and [32] applies equally well in the hyperbolic case to prove the existence of an ISC immersion $e_r : \Omega \to B_r$, isotopic through ISC immersions to ϕ_r, of constant extrinsic curvature equal to k and whose restriction to $\partial\Omega$ coincides with ϕ_r (c.f. Theorem 1.2 of [25]). Reasoning as in [32], we then show that $(e_r)_{r>0}$ subconverges to a complete, locally Lipschitz immersion $e_\infty : \Omega \to \mathbb{H}^3$, of constant extrinsic curvature equal to k in the viscosity sense, and solving the asymptotic Plateau problem (Ω, ϕ). Finally, since (Ω, e_∞) is not a tube, the arguments of Sect. 3.4.4, allow us to show that this sequence in fact converges in the C^∞_{loc} sense, so that e_∞ is smooth, and the result follows. □

We now obtain our main existence result.

Theorem 3.4.16 *Let S be a developable Möbius surface of hyperbolic type with developing map ϕ. If the universal cover of S is not isomorphic to $(\mathbb{C}, \text{Exp}(z))$, then for all $0 < k < 1$, there exists a unique k-surface $e : S \to \mathbb{H}^3$ such that $\phi_e = \phi$.*

Remark 3.4.17 Recall that the hypothesis that S be of hyperbolic type is equivalent to excluding the possibility that the universal cover of S be equivalent to $(\hat{\mathbb{C}}, z)$ or $(\mathbb{C}, z)$.

Proof Let $(\Omega_m)_{m\in\mathbb{N}}$ be a nested sequence of relatively compact open subsets of S with smooth boundary which exhausts S. By Theorem 3.4.15, for all m, there exists a k-surface $e_m : \Omega_m \to \mathbb{H}^3$ such that $\phi_{e_m} = \phi|_{\Omega_m}$. By Theorem 3.4.11, there exists a k-surface $e : S \to \mathbb{H}^3$ towards which $(e_m)_{m\in\mathbb{N}}$ subconverges in the C^∞_{loc} sense. Uniqueness follows by Theorem 3.4.14, and this completes the proof. □

We conclude by proving the results of Sect. 3.1.3.

Theorem 3.4.18 *For all $0 < k < 1$ and for all $f \in \text{Hol}(\mathbb{D})$, there exists a unique element $e \in \text{Imm}_k(\mathbb{D})$ such that*

$$\Sigma[e] = f. \tag{3.4.20}$$

Furthermore, e depends continuously on f. In other words, Σ defines a homeomorphism from $\text{Imm}_k(\mathbb{D})$ into $\text{Hol}(\mathbb{D})$.

Proof It suffices to construct a continuous inverse of Σ. Let $\widetilde{\text{Conf}}(\mathbb{D})$ denote the space of locally conformal functions from $\mathbb{D}$ into $\hat{\mathbb{C}}$ furnished with the C^0_{loc} topology. Let $\text{Conf}(\mathbb{D})$ denote the quotient of this space under the action of post-composition by Möbius maps. Now choose $f \in \text{Hol}(\mathbb{D})$. By Theorem 1.1 of Section 2 of [20], there exists an element $\phi := \tilde{A}f \in \widetilde{\text{Conf}}(\mathbb{D})$ with Schwarzian derivative equal to f. Furthermore, ϕ is unique up to post-composition by Möbius maps, and its equivalence class in Conf($\mathbb{D}$) varies continuously with f. $\tilde{A}$ thus defines a continuous map from Hol($\mathbb{D}$) into Conf($\mathbb{D}$) which we denote by A. Now let ϕ be an element of $\widetilde{\text{Conf}}(\mathbb{D})$. Since the developed Möbius surface $(\mathbb{D}, \phi)$ is not of any of the exceptional types, it follows by Theorem 3.4.16, that there exists a unique k-surface $e := B\phi : \mathbb{D} \to \mathbb{H}^3$ such that $\phi_e = \phi$. Trivially, for any element $\alpha \in \text{PSO}(3, 1)$,

$$B(\alpha \circ \phi) = \alpha \circ (B\phi),$$

so that B descends to a map from Conf($\mathbb{D}$) into $\text{Imm}_k(\mathbb{D})$. We readily verify that BA inverts Σ. Finally, by Theorem 1.5 of [26], B is continuous, and therefore so too is BA. This completes the proof. □

Theorem 3.4.19 *For all* $0 < k < 1$ *and for all* $f \in \text{Hol}(\mathbb{C}) \setminus \mathbb{C}$*, there exists a unique element* $e \in \text{Imm}_k(\mathbb{C})$ *such that*

$$\Sigma[e] = f. \tag{3.4.21}$$

Proof The proof is identical to that of Theorem 3.4.18 with two modifications. First, the developed Möbius surface $(\mathbb{C}, \phi)$ is equivalent to $(\mathbb{C}, e^z)$ if and only if the Schwarzian derivative of ϕ is a non-zero constant, and it is equivalent to $(\mathbb{C}, z)$ if and only its Schwarzian derivative vanishes. Second, since Theorem 1.5 of [30] does not apply in this case, continuity of the inverse of Σ remains unproven. □

Appendix A: A Non-complete *k*-Surface

In this appendix, we describe a non-complete k-surface. We leave the reader to provide the complete proofs of the statements made in what follows. Consider the holomorphic function

$$f(z) := -\text{Exp}(z)\cosh(z). \tag{A.1}$$

This is the Schwarzian derivative of the function

$$\tilde{\phi}(z) := \text{Exp}(\text{Exp}(z)). \tag{A.2}$$

For $0 < k < 1$, let $\tilde{e}_k : \mathbb{C} \to \mathbb{H}^3$ denote the unique k-surface solving the asymptotic Plateau problem $(\mathbb{C}, \tilde{\phi})$. By uniqueness, for all $k \in \mathbb{Z}$,

$$\tilde{\phi}(z + 2\pi ik) = \tilde{\phi}(z). \tag{A.3}$$

so that $\tilde{e}_k$ descends to a unique k-surface $e_k : \mathbb{C}^* \to \mathbb{H}^3$ such that, for all $z \in \mathbb{C}$,

$$e_k(\mathrm{Exp}(z)) = \tilde{e}_k(z). \tag{A.4}$$

This k-surface solves the Plateau problem $(\mathbb{C}^*, \mathrm{Exp}(z))$.

We now identify $\hat{\mathbb{C}}$ with $\partial_\infty \mathbb{H}^3$. In [23], we show that e_k has a cusp at 0 whose end point in $\partial_\infty \mathbb{H}^3 = \hat{\mathbb{C}}$ is 1. We now study the asymptotic geometry of $e_k(z)$ as z tends to infinity in $\mathbb{C}^*$. Let Γ denote the geodesic in $\mathbb{H}^3$ joining 0 and ∞. For all $y \in \mathbb{R}$, denote

$$L_y := \{x + iy \mid x \in \mathbb{R}\}, \tag{A.5}$$

The image of L_y under e_k converges exponentially fast to a constant speed parametrisation of Γ as y tends to $\pm\infty$. On the other hand, the image of L_y under Exp is a complete radial line rotating at constant speed as y varies. Since, by definition, Exp is the asymptotic Gauss map of e_k, we see that e_k wraps around Γ, ever tighter, infinitely many times as y tends to $\pm\infty$.

We now use a heuristic argument to show that e_k is not complete. By Theorem 3.4.11, e_k is the limit as m tends to infinity of the solution $e_{m,k}$ of the asymptotic Plateau problem $(\mathbb{C} \setminus 2m\pi i\mathbb{Z}, \mathrm{Exp}(z))$. However, by uniqueness, for all m, $e_{m,k}$ is $2m\pi$ periodic in the y direction with fundamental domain

$$\Omega_m := \{x + iy \mid x \in \mathbb{R},\ y \in]-\pi m, \pi m[\} \setminus \{0\}. \tag{A.6}$$

In particular, for all m, the resulting quotient surface has the conformal type of $\hat{\mathbb{C}} \setminus \{0, 1, \infty\}$ and, by the Gauss–Bonnet Theorem, its metric has area $2\pi/(1-k)$. Since this area is independent of m, upon letting m tend to infinity, it is reasonable to expect that the area induced over $\mathbb{C}^*$ by e_k is also equal to $2\pi/(1-k)$. In particular, since e_k has the topology of a pointed disk, it cannot be complete, for there is no hyperbolic surface with finite area, vanishing genus and two cusps. In fact, we expect the metric induced by e_k over $\mathbb{C}^*$ to be, up to rescaling, isometric to the surface

$$S := \{z := x + iy \in \mathbb{C} \mid y > 0,\ d(z, 2m\mathbb{Z}) > 1\}/4m\mathbb{Z}, \tag{A.7}$$

whose fundamental domain is shown in Fig. 3.5.

Appendix B: Category Theory

Our presentation has been structured around the framework of category theory. For didactic purposes, we provide here an elementary and low-level discussion of its basic definitions.

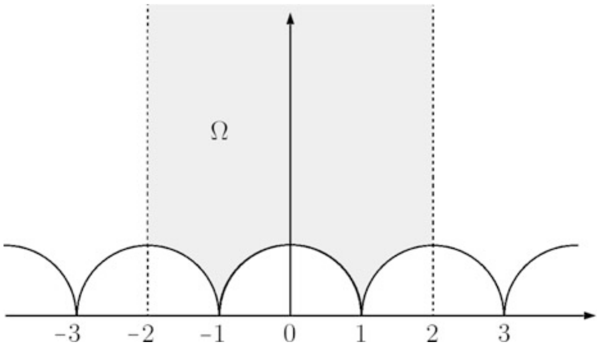


Fig. 3.5 The fundamental domain of e_k

A *category* consists of

(1) a family $\mathcal{A}$ of mathematical objects;
(2) for any two objects X and Y of $\mathcal{A}$, a set $\mathrm{Mor}(X, Y)$, which we call the *morphisms* from X to Y; and
(3) for any three objects X, Y and Z in $\mathcal{A}$, a function

$$\circ : \mathrm{Mor}(X, Y) \times \mathrm{Mor}(Y, Z) \to \mathrm{Mor}(X, Z), \tag{B.1}$$

which we call *composition*,

such that

(4) for any object X of $\mathcal{A}$, there exists a unique element $e \in \mathrm{Mor}(X, X)$, which we call the *identity*, such that for any other object Y of $\mathcal{A}$, and for all $f \in \mathrm{Mor}(X, Y)$,

$$e \circ f = f, \tag{B.2}$$

whilst, for all $f \in \mathrm{Mor}(Y, X)$,

$$f \circ e = f; \textit{ and} \tag{B.3}$$

(5) for any four objects X, Y, Z and W of $\mathcal{A}$, for all $\alpha \in \mathrm{Mor}(X, Y)$, $\beta \in \mathrm{Mor}(Y, Z)$ and $\gamma \in \mathrm{Mor}(Z, W)$,

$$\alpha \circ (\beta \circ \gamma) = (\alpha \circ \beta) \circ \gamma. \tag{B.4}$$

It is crucial at this stage to pay close attention to the semantics of these definitions. A family is *not* a set. In fact, there is an implicit abuse of language in the concept of family: a family is a list of axioms which can be written down. Likewise, an object of a family is *not* an element of a set: it is a mathematical object which satisfies the axioms of the family. Thus, the family of groups is given by the axioms of group theory; the family of vector spaces is given by the axioms of linear algebra; and so on.

Most familiar mathematical constructs lie within this framework. For example, the category of vector spaces is the category whose objects are vector spaces and whose morphisms are linear maps; the category of Banach spaces is a category whose morphisms are bounded linear maps; the category of smooth manifolds is a category whose morphisms are smooth maps; and so on. It should hopefully become clear that in defining new mathematical objects, it is indeed often desirable to identify their morphisms and to verify whether these morphisms include identity elements and compose associatively. It is in this sense that the above axioms constitute a check-list of properties that families of mathematical objects ought to possess.

A *(covariant) functor* $\mathcal{F}$ between two categories $\mathcal{A}$ and $\mathcal{B}$ consists of

(1) a mathematical operation that associates to every object X of $\mathcal{A}$ an object $\mathcal{F}(X)$ of $\mathcal{B}$; and
(2) another mathematical operation which associates to every pair X and Y of objects of $\mathcal{A}$ and to every morphism α in $\mathrm{Mor}(X, Y)$ a morphism $\mathcal{F}(\alpha)$ in $\mathrm{Mor}(\alpha(X), \alpha(Y))$,

such that

(3) for any object X of $\mathcal{A}$,

$$\mathcal{F}(e) = e; \text{ and} \tag{B.5}$$

(4) for any three objects X, Y and Z of $\mathcal{A}$, for all $\alpha \in \mathrm{Mor}(X, Y)$ and for all $\beta \in \mathrm{Mor}(Y, Z)$,

$$\mathcal{F}(\beta \circ \alpha) = \mathcal{F}(\beta) \circ \mathcal{F}(\alpha). \tag{B.6}$$

Condition (4) can also be replaced with the condition that

$$\mathcal{F}(\beta \circ \alpha) = \mathcal{F}(\alpha) \circ \mathcal{F}(\beta), \tag{B.7}$$

in which case the functor is said to be *contravariant*. However, although the simplest examples of functors are often contravariant, only covariant functors will be used in this chapter.

As before, it is crucial to pay close attention to the semantics of these definitions. A functor is *not* a function: it is a list of mathematical operations which can be written down. For example, the dual operation, which associates to every vector space its dual vector space is a contravariant functor from the category of vector spaces to itself; the C^∞ operation, which associates to every smooth manifold the vector space of smooth functions defined over that manifold, is a contravariant functor from the category of smooth manifolds to the category of vector spaces; and so on. Once again, it should hopefully become clear that in defining new mathematical operations between families of objects, it is often desirable to know their effects on morphisms so that the above axioms again provide a check-list of properties that such operations ought to possess.

For those of us trained to express our ideas in terms of sets and functions, this formalism can appear at first quite unsettling. However, the concepts of category theory are, ironically, less abstract than those of set theory and closer to what we have in mind when mathematical operations are performed. To see this, recall that sets are actually abstract mathematical objects which are not necessarily constructible in any sense that we would normally understand, which is precisely what gives the mystery to such results as the Banach–Tarski paradox. Families, on the other hand, are clearly defined by fixed lists of axioms which can be written down. Likewise, functions are abstract objects of set theory which are also not necessarily constructible in any sense that we would normally understand, whilst functors are fixed lists of mathematical operations which can again be written down. In fact, whenever we carry out explicit calculations, we never work with functions, but rather with the sequences of operations used to define them. Such sequences, which we regularly encounter in our day-to-day mathematical life, are, in fact, closer in kind to the functors of category theory than they are to the functions of set theory.

References

1. L. Andersson, T. Barbot, F. Béguin, A. Zeghib, Cosmological time versus CMC time in spacetimes of constant curvature. Asian J. Math. **16**(2), 37–87 (2012)
2. W. Ballmann, M. Gromov, V. Schroeder, *Manifolds of Nonpositive Curvature*. Progress in Mathematics, vol. 61 (Birkhäuser, Boston, 1985)
3. E. Calabi, I. An extension of E. Hopf's maximum principle with an application to Riemannian geometry. Duke Math. J. **25**(1), 45–56 (1958)
4. G. Calsamiglia, B. Deroin, S. Francaviglia, Branched projective structures with Fuchsian holonomy. Geom. Topo. **18**, 379–446 (2014)
5. R.D. Canary, D. Epstein, A. Marden, *Fundamentals of Hyperbolic Geometry* (Cambridge University Press, Cambridge, 2006)
6. I. Chavel, *Riemannian Geometry, A Modern Introduction* (Cambridge University Press, Cambridge, 1994)
7. D. Dumas, Complex projective structures, in *Handbook of Teichmüller theory. Vol. II*, ed. by A. Papadopoulos (Eur. Math. Soc., Zürich, 2009), pp. 455–508
8. J.M. Espinar, J.A. Gálvez, P. Mira, Hypersurfaces in $\mathbb{H}^{n+1}$ and conformally invariant equations: the generalized Christoffel and Nirenberg problems. J. Eur. Math. Soc. **11**, 903–939 (2009)
9. F. Fillastre, G. Smith, Group actions and scattering problems in Teichmüller theory, in *Handbook of Group Actions IV*, ed. by L. Ji, A. Papadopoulos, S.-T. Yau. Advanced Lectures in Mathematics, vol. 40 (International Press, Boston, 2018), pp. 359–417
10. D. Gallo, M. Kapovich, A. Marden, The monodromy groups of Schwarzian equations on closed Riemann surfaces. Ann. Math. **151**(2), 625–704 (2000)
11. W.M. Goldman, Projective structures with Fuchsian holonomy. J. Differ. Geom. **25**(3), 297–326 (1987)
12. M. Gromov, Sign and geometric meaning of curvature. Rend. Sem. Mat. Fis. Milano **61**, 9–123 (1991)
13. B. Guan, J. Spruck, The existence of hypersurfaces of constant Gauss curvature with prescribed boundary. J. Differ. Geom. **62**(2), 259–287 (2002)

14. J.H. Hubbard, *Teichmüller Theory and Applications to Geometry, Topology and Dynamics, Vol. I, Teichmüller Theory* (Matrix Editions, Ithaca, 2006)
15. M. Kapovich, *Hyperbolic Manifolds and Discrete Groups* (Birkhaüser, Boston, 2010)
16. R.S. Kulkarni, Conformal structures and Möbius structures, in *Conformal Geometry*. Aspects of Mathematics (MPIM, Bonn, 1988)
17. R.S. Kulkarni, U. Pinkall, A canonical metric for Möbius structures and its applications. Math. Zeit. **216**(1), 89–129 (1994)
18. F. Labourie, Problèmes de Monge-Ampère, courbes holomorphes et laminations. Geom. Funct. Anal. **7**(3), 496–534 (1997)
19. F. Labourie, Un lemme de Morse pour les surfaces convexes. Invent. Math. **141**, 239–297 (2000)
20. O. Lehto, *Univalent Functions and Teichmüller Spaces* (Springer, Berlin, 1990)
21. K.P. Scannell, Flat conformal structures and the classification of de Sitter manifolds. Comm. Anal. Geom. **7**(2), (1999), 325–345
22. J.M. Schlenker, Hypersurfaces in $\mathbb{H}^n$ and the space of its horospheres. Geom. Funct. Anal. **12**, 395–435 (2002)
23. G. Smith, Pointed k-surfaces. Bull. Soc. Math. France **134**(4), 509–557 (2006)
24. G. Smith, Moduli of flat conformal structures of hyperbolic type. Geom. Dedicata **154**(1), 47–80 (2011)
25. G. Smith, Compactness results for immersions of prescribed Gaussian curvature I - analytic aspects. Adv. Math. **229**(2), 731–769 (2012)
26. G. Smith, Hyperbolic Plateau problems. Geom. Dedicata **176**(1), 31–44 (2014)
27. G. Smith, *Global Singularity Theory for the Gauss Curvature Equation*. Ensaios Matemáticos, vol. 28 (Sociedade Brasileira De Matemática, Rio de Janeiro, 2015), pp. 1–114
28. G. Smith, k-surfaces and quaternionic structures (in preparation)
29. M. Spivak, *A Comprehensive Introduction to Differential Geometry*, Vol IV (Publish or Perish Inc., Berkeley, 1979)
30. W.P. Thurston, Three dimensional manifolds, Kleinian groups and hyperbolic geometry. Bull. Am. Mat. Soc. **6**(3), 357–381 (1982)
31. W.P. Thurston, The geometry and topology of three-manifolds (MSRI, Berkeley, 2002)
32. N.S. Trudinger, X.J. Wang, On locally convex hypersurfaces with boundary. J. Reine Angew. Math. **551**, 11–32 (2002)

Chapter 4
Cone 3-Manifolds

Joan Porti

Abstract This is an overview on hyperbolic cone 3-manifolds, their deformation theory and their role in Thurston's orbifold theorem. We also describe the phenomena that may occur when deforming the cone angles, like cusp opening or collapses, under the assumption that the cone angles are less than π.

Keywords Cone manifolds · Deformations · Rigidity · Orbifold

MSC 57M50

4.1 Introduction

Cone 3-manifolds are manifolds equipped with metrics of constant curvature that are singular at an embedded graph, and the singularity follows a specific conical structure. They can be obtained from 3-dimensional polyhedra of constant curvature by identifying their faces along isometries, thus the singular locus is contained in the 1-skeleton.

Cone 3-manifolds were considered by Thurston in his proof of the orbifold theorem. The underlying space of an orientable orbifold of constant curvature has a natural metric of cone manifold. The starting point in the proof of the orbifold theorem is another well known theorem of Thurston: the hyperbolic Dehn filling theorem. The proof of the hyperbolic Dehn filling theorem provides cone manifolds with small cone angles; then the main strategy of the orbifold theorem is to increase those cone angles (until the angles determined by the topology of the orbifold) and to analyze the possible phenomena that may occur. This motivates the study of geometric properties of cone 3-manifolds, like their deformation theory or the convergence of sequences of cone 3-manifolds.

J. Porti (✉)
Departament de Matemàtiques UAB and Centre de Recerca Matemàtica, Cerdanyola del Vallès, Spain
e-mail: porti@mat.uab.cat

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_4

The first sketch, or program, of the proof of the orbifold theorem was the content of a preprint by Thurston in 1982 [48], see also [49]. Then the proof was completed by several contributors [3, 4, 14, 25, 45, 54]. There are some later results that give a more natural argument in some parts of the proof, like the local rigidity theorems surveyed here. The goal of this paper is not to provide a proof of the orbifold theorem, but to give an overview of some properties of cone 3-manifolds.

Kleiner and Lott have proved the orbifold theorem with Ricci flow on orbifolds [30], without using cone manifolds. Cone manifolds remain however an interesting geometric object, that may have an intuitive visualization. Besides, cone manifolds have applications other than the orbifold theorem: Hodgson and Kerckhoff use them in [27] to find a uniform upper bound on the number of non-hyperbolic Dehn fillings. The deformation theory of cone manifolds is used by Bromberg in the proof of the Bers density conjecture [10], by Brock and Bromberg in in a generalization of this conjecture [7], by Brock, Bromberg, Evans and Souto in the tameness conjecture [8], as well as by Bonahon and Otal [5] to study bending measured laminations.

It is worth mentioning that there are a lot of contributions on cone 3-manifolds that are not overviewed here. For instance, the many examples of deformations and volume computations of the Siberian school around Alexander D. Mednykh, as well as the pioneering examples from the long term collaboration between Mike Hilden, José Maria Montesinos Amilibia and Maite Lozano Imízcoz. Here we just mention a few examples from these authors.

This paper is organized as follows: Sect. 4.2 reviews the definition, basic constructions, and elementary properties of cone manifolds, focusing in dimensions 2 and 3. Section 4.3 is devoted to Thurston's hyperbolic Dehn filling theorem, that explains how cone 3-manifolds with small cone angles occur, and the natural questions that arise. Then Sect. 4.4 reviews local rigidity results, in particular the results that allow to deform cone angles. Section 4.5 is devoted to sequences of cone 3-manifolds, more precisely to the notions of convergence, a compactness theorem, a description of thin parts and their applications (eg. global rigidity), all for cone angles strictly less than π. Finally Sect. 4.6 is devoted to some examples, that illustrate previous sections and give examples of different phenomena that occur to cone manifolds, including some examples with cone angles larger than π.

4.2 Cone Manifolds

In this section we give the definition and basic constructions of cone manifolds, focusing on dimensions two and three.

We start with the definition in dimension 2, with curvature $\kappa \in \mathbb{R}$. To describe the metric in constant curvature κ, consider the function

$$\mathrm{s}_k(r) = \begin{cases} \frac{\sin(r\sqrt{k})}{\sqrt{k}} & \text{if } \kappa > 0 \\ r & \text{if } \kappa = 0 \\ \frac{\sinh(r\sqrt{-k})}{\sqrt{-k}} & \text{if } \kappa < 0 \end{cases}$$

This function is the unique solution to the differential equation $s_k'' + \kappa s_k = 0$ with initial conditions $s_k(0) = 0$, $s_k'(0) = 1$. In the next definition, the local description of the metric in polar coordinates for a cone surface (4.1) is a modification of the metric of the plane of constant curvature, $ds^2 = dr^2 + s_k^2(r)d\theta^2$, with $r \in (0, r_0), \theta \in [0, 2\pi]$.

Definition 4.1 A *cone surface* of constant curvature $\kappa \in \mathbb{R}$ is a surface equipped with a length distance, where the metric is locally described, in polar coordinates, by

$$ds^2 = dr^2 + \left(\frac{\alpha}{2\pi}\right)^2 s_k^2(r)d\theta^2, \qquad r \in (0, r_0),\ \theta \in [0, 2\pi], \tag{4.1}$$

where $\theta = 2\pi$ is identified to $\theta = 0$. The parameter $\alpha > 0$ is called the *cone angle* at the point with coordinate $r = 0$.

When $\alpha \neq 2\pi$, we say that the point is *singular*, or a *cone point*.

For $\alpha = 2\pi$ the metric is locally a Riemannian metric of constant curvature κ and the point is called *regular*.

In Eq. (4.1), $r \in (0, r_0)$ is the radial coordinate and $\theta \in [0, 2\pi]$ is the angle parameter. Furthermore, when $\kappa > 0$ we require $r_0 \leq \frac{\pi}{\sqrt{\kappa}}$.

Notice that the metric in (4.1) can be changed to the standard metric by re-parameterizing and changing the domain of the coordinate θ:

$$ds^2 = dr^2 + s_k^2(r)d\theta^2, \qquad r \in (0, r_0),\ \theta \in [0, \alpha], \tag{4.2}$$

where $\theta = \alpha$ is identified to $\theta = 0$. Namely, we consider a sector of angle α in the space of constant curvature κ and we identify its sides by a rotation, Fig. 4.1.

Example 4.1 Consider a triangle with angles $0 < \frac{\alpha_1}{2}, \frac{\alpha_2}{2}, \frac{\alpha_3}{2} < \pi$. It lies in a plane of constant curvature κ, with

$$\begin{cases} \kappa < 0 & \text{if } \frac{\alpha_1}{2} + \frac{\alpha_2}{2} + \frac{\alpha_3}{2} < \pi, \\ \kappa = 0 & \text{if } \frac{\alpha_1}{2} + \frac{\alpha_2}{2} + \frac{\alpha_3}{2} = \pi, \\ \kappa > 0 & \text{if } \frac{\alpha_1}{2} + \frac{\alpha_2}{2} + \frac{\alpha_3}{2} > \pi \quad (\text{and } \frac{\alpha_i}{2} + \frac{\alpha_j}{2} < \frac{\alpha_k}{2} + \pi, \quad \text{for } i \neq j \neq k \neq i). \end{cases}$$

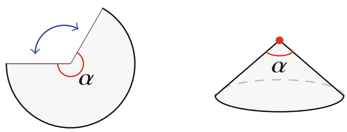


Fig. 4.1 A singular point of cone angle $\alpha < 2\pi$ is viewed as a cone (though the definition allows cone angle $\alpha > 2\pi$)

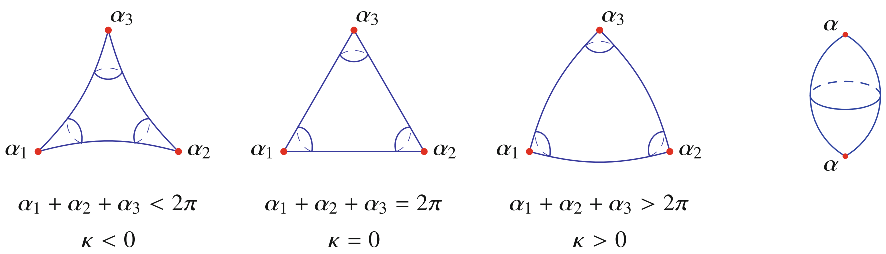


Fig. 4.2 On the left, three turnovers: cone surfaces $S^2(\alpha_1, \alpha_2, \alpha_3)$ of curvature κ (subject to $\alpha_i + \alpha_j < 2\pi + \alpha_k$, for $i \neq j \neq k \neq i$). On the right, cone surfaces $S^2(\alpha, \alpha)$ of curvature > 0

We double the triangle with angles $\frac{\alpha_1}{2}$, $\frac{\alpha_2}{2}$, and $\frac{\alpha_3}{2}$ along its boundary: in this way we obtain a Riemannian metric on S^2 of constant curvature everywhere except at the vertices. Namely we obtain a cone surface with three cone points, of respective cone angles α_1, α_2 and α_3, subject to $\alpha_i + \alpha_j < \alpha_k + 2\pi$, for $i \neq j \neq k \neq i$. This example is called a *turnover* and it is denoted by $S^2(\alpha_1, \alpha_2, \alpha_3)$, see Fig. 4.2, left.

Example 4.2 Consider a spherical bigon of angle $0 < \alpha < 2\pi$, namely the region of S^2 bounded by two meridians at angle α. In spherical coordinates it is the region with parameters $(\rho, \theta) \in (0, \pi) \times [0, \alpha]$ where ρ is the distance to the north pole and θ the longitude (and $\pi/2 - \rho$ the latitude). We identify the sides by a rotation. The result is a cone manifold with two cone points of angle α, that we denote by $S^2(\alpha, \alpha)$. See Fig. 4.2 right. It is the spherical suspension of a circle of length α, namely with the metric

$$ds^2 = d\rho^2 + \sin^2(\rho) d\theta^2, \qquad \text{for } \rho \in (0, \pi) \text{ and } \theta \in [0, \alpha]/\alpha \sim 0.$$

The cone manifold $S^2(\alpha, \alpha)$ can be seen as the limit of $S^2(\alpha_1, \alpha_2, \alpha_3)$ when $\alpha_3 \to 2\pi$, because $|\alpha_1 - \alpha_2| \leq 2\pi - \alpha_3$.

The definition of cone manifold is inductive on the dimension. The construction uses the metric cone. Start with the topological cone: for a compact topological space X, consider the product $X \times [0, R)$ for some $R > 0$ and collapse $X \times \{0\}$ to a point (the *tip* of the cone), and denote the quotient by $X \times [0, R)/\sim$.

Definition 4.2 Let (X, d_X) be a metric space of diameter $\leq \pi$. The *metric cone* of constant curvature κ on X is the topological cone $X \times [0, R)/\sim$ (we require that $R < 2\pi/\sqrt{\kappa}$ when $\kappa > 0$) equipped with the distance d so that $(x_1, r_1), (x_2, r_2) \in X \times (0, R]$ and the tip of the cone $(*, 0)$ form a triangle isometric to the triangle in the plane of constant curvature κ with edge lengths $r_1, r_2, d((x_1, r_1), (x_2, r_2))$, and angle $d_X(x_1, x_2)$ at the tip $(*, 0)$. It is denoted by

$$\mathrm{Cone}_{R,\kappa}(X) = (X \times [0, R)/\sim, d).$$

The space X is called the *link* of the *tip* of the cone.

When the distance on X is provided by a Riemannian metric ds_X^2, then the metric on $\mathrm{Cone}_{R,\kappa}(X)$ is described by

$$ds^2 = dr^2 + \mathrm{s}_\kappa^2(r) ds_X^2.$$

Notice that in Definition 4.1 the local description of a cone surface is the metric cone of constant curvature over a circle.

Definition 4.3 A d-dimensional *cone manifold* of constant curvature κ is a metric length space C satisfying the following local property. For each $x \in C$ there exists a cone manifold $\mathrm{Link}(x)$ of curvature 1 homeomorphic to S^{d-1} such that a neighborhood of x is isometric to the metric cone of constant curvature κ on $\mathrm{Link}(x)$, $\mathrm{Cone}_{\varepsilon,\kappa}(\mathrm{Link}(x))$.

When the curvature κ is equal to 1 the cone manifold is called *spherical*, when $\kappa = 0$, *Euclidean*, and when $\kappa = -1$, *hyperbolic*.

Remark 4.1 We require that $\mathrm{Link}(x)$ is homeomorphic (not isometric) to a sphere S^{d-1}, so that C is topologically a manifold.

If we do not require $\mathrm{Link}(x)$ to be homeomorphic to a sphere, then we talk about *conifolds*, but we will not consider them here. The easiest example of conifold that is not a cone manifold is the cone on the projective plane.

Proposition 4.1 *The underlying space of an orientable orbifold of constant curvature and dimension* 2 *or* 3 *inherits naturally the structure of a cone manifold.*

Proof The underlying space of an orbifold of constant curvature is locally modeled on $\mathbb{X}_\kappa^n/\Gamma$, where $\mathbb{X}_\kappa^n$ is the space of constant curvature κ, and $\Gamma \subset \mathrm{SO}(n)$ is a finite group of isometries fixing a point.

By construction, there exists $\varepsilon > 0$ such that a neighborhood of a point x in the underlying space is isometric to $B(\tilde{x}, \varepsilon)/\Gamma$, where $B(\tilde{x}, \varepsilon)$ is a metric ball of radius $\varepsilon > 0$ in $\mathbb{X}_\kappa^n$. Notice that $B(\tilde{x}, \varepsilon)/\Gamma$ is the metric cone of radius $\varepsilon > 0$ on its link S^{n-1}/Γ. Since we assume orientability, for $n = 2$, S^1/Γ is homeomorphic to a circle and for $n = 3$, S^2/Γ is homeomorphic to a 2-sphere. □

The previous proposition holds in any dimension if we allow conifolds instead of cone manifolds, i.e. if we do not require the link to be homeomorphic to a sphere.

Proposition 4.2 (Gluing Polygons in Dimension 2) *Let* $P_1, \dots, P_k \subset \mathbb{X}_\kappa^2$ *be polygons of constant curvature* κ. *Assume that their edges* $(P_i)_j$ *are paired by isometries* s_{ij}. *Then the metric space obtained by identification along the isometries*

$$(P_1 \sqcup \cdots \sqcup P_k)/\sim_{s_{ij}}$$

is a cone surface.

The cone structure is easily constructed from matching the cones on the polygons. The key point is to prove that the link of each point is a circle; this follows from the classification of 1-dimensional manifolds (see also Theorem 6.7.6 in [42]).

Remark 4.2 Proposition 4.2 generalizes to dimension 3 if we can guarantee that links of equivalence classes of vertices are homeomorphic to spheres. This holds true for instance if cone angles of edges are $\leq 2\pi$, by Gauss-Bonnet (Proposition 4.3).

Proposition 4.2 is illustrated in Examples 4.1 and 4.2. By means of the Dirichlet polyhedron (below in Definition 4.5 and Proposition 4.5) we show that all cone manifolds can be constructed from Proposition 4.2.

Definition 4.4 On a cone d-manifold C, a point $x \in C$ is *singular* if its link is not *isometric* to the standard $(d-1)$-sphere S^{d-1}, and regular otherwise. The singular locus of C is denoted by Σ.

Remark 4.3 The singular locus Σ is a stratified subspace of codimension ≥ 2. In particular, for a cone surface, Σ is a discrete subset.

For a 2-dimensional cone manifold, we have a Gauss-Bonnet formula, see for instance [34, 50]:

Proposition 4.3 (Gauss-Bonnet Formula for Cone Surfaces) *Let C^2 be a cone surface of constant curvature κ, with finite area and n cone points of respective cone angles $\alpha_1, \ldots, \alpha_n$. Then*

$$\kappa \operatorname{area}(C^2) + \sum_i (2\pi - \alpha_i) = 2\pi \chi(C^2),$$

where $\chi(C^2)$ denotes the Euler characteristic of the underlying surface.

It follows from the Gauss-Bonnet formula that if $\kappa = 1$ and the cone angles are $\leq \pi$, then there are at most three cone points. With some extra work, one can determine geometrically those cone manifolds:

Proposition 4.4 *Let C^2 be a spherical cone surface with cone angles $\leq \pi$. If C^2 is orientable, then it is one of the following:*

1. *A smooth sphere S^2.*
2. *$S^2(\alpha, \alpha)$, the spherical suspension of a circle.*
3. *$S^2(\alpha, \beta, \gamma)$, a turnover with $\alpha + \beta + \gamma > 2\pi$.*

If C^2 is not orientable, then it is the quotient of S^2 or $S^2(\alpha, \alpha)$ by the antipodal map, i.e. the projective plane with possibly a cone point, P^2 or $P^2(\alpha)$.

Furthermore, the isometry class of C^2 is determined by the cone angles (namely they are rigid).

This rigidity does not hold anymore for spherical cone manifolds with cone angles larger than π; consider for instance the double of a spherical quadrilateral. See [34] for a description of the moduli space of spherical cone surfaces.

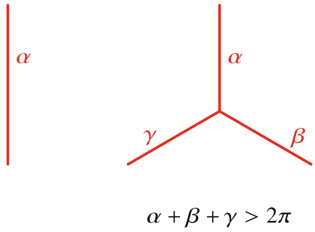


Fig. 4.3 The models of the singular locus Σ when cone angles are ≤ π

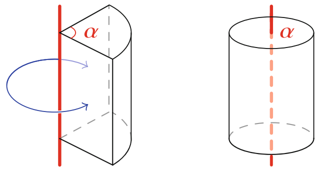


Fig. 4.4 Locally, a singular edge is the result of identifying the sides of a sector in the space of constant curvature by a rotation

From Proposition 4.4, as the link of a point is a spherical cone manifold, we have:

Corollary 4.1 *A 3-dimensional cone manifold with cone angles* $\leq \pi$ *is locally isometric to one of the following:*

1. *A smooth point (the cone of a smooth sphere).*
2. *A singular edge (the cone of* $S^2(\alpha, \alpha)$*).*
3. *A trivalent vertex of a singular graph (the cone of* $S^2(\alpha, \beta, \gamma)$*).*

In particular, the singular locus Σ *is a disjoint union of circles and trivalent graphs, see Fig. 4.3.*

Furthermore, the isometry class of a neighborhood is determined by the cone angles.

For a singular edge, the cone angle of the link at every point is also called the cone angle of the edge, see Fig. 4.4.

Definition 4.5 Let C be a cone 3-manifold and $x \in C \setminus \Sigma$ a regular point. The *cut locus* of C centered at x is

$$\mathrm{Cut}_x = \{y \in C \mid y \in \Sigma \text{ or } y \text{ has at least 2 minimizing segments to } x\}.$$

The complement of the cut locus is the *Dirichlet domain* centered at $x \in C \setminus \Sigma$, $D_x = C \setminus \mathrm{Cut}_x$:

$$D_x = \{y \in C \setminus \Sigma \mid \text{there is a unique minimizing segment from } x \text{ to } y\}.$$

Proposition 4.5 *The Dirichlet domain embeds as a star-shaped domain in* $\mathbb{X}^3_\kappa$*, the space of constant curvature* $\kappa \in \mathbb{R}$*, and for* $\kappa \leq 0$ *its closure is a polyhedron.*

Furthermore, when the cone angles are $\leq \pi$*, this Dirichlet polyhedron is convex.*

This proposition helps to explain why the hypothesis on cone angles $\leq \pi$ is relevant for cone manifolds. The fact that the Dirichlet polyhedron is convex allows to reproduce arguments in Riemannian geometry in this context. We see examples in Sect. 4.5.

Before finishing this section, we state a result related to the following section.

Proposition 4.6 *Let C be a closed hyperbolic cone 3-manifold without singular vertices (i.e. Σ is a link). Then $|C| \setminus \Sigma$ is a hyperbolic manifold (namely, it admits a complete hyperbolic metric).*

Proof Deform the non-complete metric on $|C| \setminus \Sigma$ to a complete metric of variable negative curvature. Then one can show that it has the topological properties required for being hyperbolic (irreducible, atoroidal, and $\pi(|C| \setminus \Sigma)$ has no center) and apply geometrization for Haken manifolds. □

The complete structure on this proposition can be seen as a cone manifold structure of angle zero. This is better explained in the next section, by Thurston's hyperbolic Dehn filling.

Remark 4.4 Let C be a closed hyperbolic cone 3-manifold without singular vertices as in Proposition 4.6. Then the volume of the complete hyperbolic structure on $|C| \setminus \Sigma$ is larger than the volume of the cone 3-manifold C. The maximality of the volume is due to Gromov–Thurston–Goldman, and written by Dunfield [20]. More precisely, as explained in [20], Goldman notices in [21] that the proof of Mostow rigidity in Thurston's notes [47] applies to representations, a proof that Thurston attributes to Gromov.

4.3 Hyperbolic Dehn Filling

Thurston's hyperbolic Dehn filling provides examples of hyperbolic cone three-manifolds, with small cone angles, and it is the starting point of the proof of the orbifold theorem. Those cone 3-manifolds are obtained by deforming cusped manifolds and then taking the metric completion.

We first consider a two-dimensional example:

Example 4.3 Start with a hyperbolic triangle with ideal vertices and double it along its boundary. This yields a planar hyperbolic surface with three cusps, that we call $S^2(0, 0, 0)$. Next consider triangles with finite vertex and angle $\alpha/2 > 0$ at every vertex, for α in a neighborhood of 0. By taking the double of the triangles along the boundary, we get a family of turnovers $S^2(\alpha, \alpha, \alpha)$ as in Example 4.1 and Fig. 4.2. As triangles with small angles are deformations of ideal triangles, the turnovers $S^2(\alpha, \alpha, \alpha)$ are deformations of the cusped surface $S^2(0, 0, 0)$ (Fig. 4.5).

In dimension three, we first recall the topological description of filling. Consider a compact 3-manifold M^3 with boundary a 2-torus $\partial M^3 \cong T^2 \cong S^1 \times S^1$. Attach a

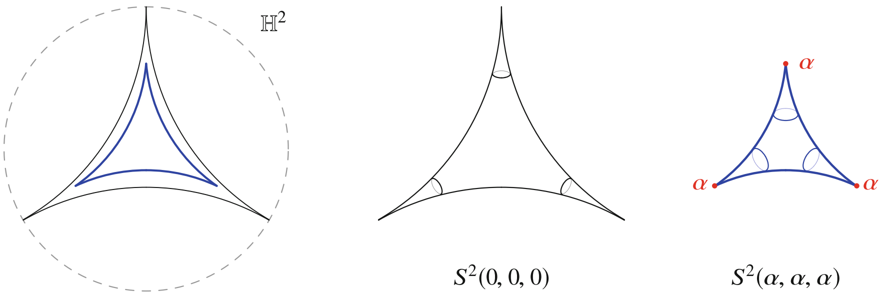


Fig. 4.5 Triangle of small angles as perturbation of the ideal triangle, with angles 0 (left). The double of the ideal triangle is the cusped surface $S^2(0, 0, 0)$ (center) and the double of the compact triangle is the turnover $S^2(\alpha, \alpha, \alpha)$ (right)

solid torus $D^2 \times S^1$ (a product of a disc and a circle) to its boundary:

$$M^3 \cup_\varphi D^2 \times S^1 = \left(M^3 \sqcup D^2 \times S^1\right)/x \sim \varphi(x)$$

where $\varphi\colon \partial D^2 \times S^1 \to \partial M^3$ is a homeomorphism. The solid torus $D^2 \times S^1$ is called the *filling torus* and the curve $\{0\} \times S^1$ its *soul*.

The homeomorphism type of the Dehn filling depends only on the unoriented isotopy class of the curve $\varphi(\partial D^2 \times \{*\})$ in $\partial M^3 \cong T^2$, the filling meridian. In its turn, this unoriented isotopy class is determined by its homology class in $H_1(T^2, \mathbb{Z})$ up to sign, hence it may be described by a rational slope, an element of $\mathbb{Q} \cup \{\infty\}$, as follows. Fix a basis for the first cohomology group, namely an isomorphism $H_1(T^2, \mathbb{Z}) \cong \mathbb{Z}^2$; the filling meridian with homology class $\pm(p, q)$ via this isomorphism is described by the slope $p/q \in \mathbb{Q} \cup \{\infty\}$.

When the 3-manifold is a knot exterior in S^3, a Dehn filling on its exterior is called Dehn surgery on the knot.

Next we state the well known Thurston hyperbolic Dehn filling theorem in terms of cone manifolds. To simplify, we state it for only one cusp.

Theorem 4.1 (Thurston's Generalized Hyperbolic Dehn Filling) *Let M^3 be a compact orientable 3-manifold with boundary a 2-torus. Assume its interior is hyperbolic.*

For every slope $q \in \mathbb{Q} \cup \{\infty\}$ there exists $\Theta_q > 0$, depending on the slope q and the manifold M^3, so that there is a family of cone manifold structures on the Dehn filling with slope q, with singular locus the soul of the filling torus, and with cone angles in the interval $(0, \Theta_q)$.

Furthermore, the number of slopes $q \in \mathbb{Q} \cup \{\infty\}$ such that $\Theta_q \leq 2\pi$ is finite.

Notice that when $\Theta_q > 2\pi$, Thurston's hyperbolic Dehn filling provides a honest hyperbolic three-manifold (e.g. with a metric with no singularities), and we recover the usual statement of Thurston's hyperbolic Dehn filling theorem. The last statement in Theorem 4.1 guarantees that almost all Dehn fillings are hyperbolic manifolds. In fact the statement is even more general. Thurston's proof provides a deformation space with a complex parameter. In this deformation space, the metric is non complete and its metric completion may yield a topological manifold (with singular metric or not) or a singular space, a so-called singularity of "generalized Dehn type". In this deformation space, the manifold Dehn filings are a countable set of points, joined by lines to the initial point, corresponding to the angle deformation of the cone manifolds.

Remark 4.5 Cone manifolds in Theorem 4.1 are constructed by deforming the complete metric structure and taking the metric completion. As in Example 4.3, we may view the metric at angle zero as the complete metric on the interior of M^3, hence the cone angle varies in $[0, \Theta_q)$.

Remark 4.6 In Theorem 4.1 one can replace 2π in the last sentence by any positive constant $C > 0$; the conclusion is that the number $\#\{q \in \mathbb{Q} \cup \{\infty\} \mid \Theta_q \leq C\}$ is finite. Of course this number depends on C, and a priori it depends also on M^3.

Theorem 4.2 *Let M^3 be a compact orientable 3-manifold with boundary a 2-torus and with hyperbolic interior. Then, for* every *slope $q \in \mathbb{Q} \cup \{\infty\}$, $\Theta_q \geq 2\pi/3$.*

This theorem is part of the proof of Thurston's orbifold theorem, and can be found in the different approaches to the proof [3, 4, 14, 25, 45, 54]. We discuss it later in Sect. 4.5. When $\Theta_q > 2\pi/n$ Theorem 4.1 yields a hyperbolic orbifold, with branching locus the soul of the filling torus, and branching index n. Thus, as corollary of Theorem 4.2:

Corollary 4.2 *Let M^3 be a compact orientable 3-manifold with boundary a 2-torus and with hyperbolic interior. For* every *slope $q \in \mathbb{Q} \cup \{\infty\}$, the orbifold with underlying space the q-Dehn filling, branching locus the soul of the filling torus, and branching index $n \geq 4$ is hyperbolic.*

The bound $n = 4$ is optimal: for instance the orbifold with underlying space the three-sphere, branching locus the figure eight knot and ramification 3 is Euclidean. Equivalently, there exists a Euclidean cone manifold structure on S^3, with singular locus the figure eight knot and cone angle $2\pi/3$.

If we focus on nonsingular Dehn fillings, then a natural question is to find a uniform bound on the number of Dehn fillings that are not hyperbolic. This has been found by Hodgson and Kerckhoff in [27]:

Theorem 4.3 (Hodgson–Kerckhoff) *Let M^3 be a compact orientable 3-manifold with boundary a 2-torus and with hyperbolic interior. Then $\Theta_q \leq 2\pi$ for at most 60 slopes $q \in \mathbb{Q} \cup \{\infty\}$, independently of M^3.*

The statement in [27] involves the so-called normalized length of a slope in the hororspherical torus. This torus has a natural Euclidean structure up to homotety, and one normalizes it so that it has area 1. Hodgson and Kerckhoff prove that for slopes q so that its normalized length in the horospherical torus is at least 7.515, we have $\Theta_q > 2\pi$. Besides the tools we describe here, one of the main innovations of Hodgson and Kerckhoff are infinitesimal harmonic deformations. They succeed in controlling the radius of a metric tube around the singular geodesic when deforming. We recall more results of [27] in Sect. 4.5.

In the proof of Theorem 4.2 there are two basic ingredients: deforming the structures by changing the cone angles and studying the limits of sequences. In Sect. 4.4 we describe how cone manifolds are deformed, and in Sect. 4.5 we analyze sequences of cone manifolds.

4.4 Local Rigidity

In this section we overview results that allow to deform the cone angles of cone manifolds. Those are *local rigidity* results because they show that the multiangles are local parameters of the deformation space.

Given a cone manifold C, the *underlying manifold* is denoted by $|C|$. The topological pair formed by $(|C|, \Sigma)$ is called the *topological type*, where $\Sigma \subset |C|$ is the singular locus. The *meridians* are (conjugacy classes of) elements in the fundamental group $\pi_1(|C| \setminus \Sigma)$ represented by loops around the arcs and circles of Σ (that in $|C|$ bound a disc that intersects Σ in its center), Fig. 4.6.

We are interested in deformations that preserve the topological type. The complement $|C| \setminus \Sigma$ inherits a non-singular hyperbolic metric that is not complete, whose metric completion is C. The incomplete structure on $|C| \setminus \Sigma$ has a holonomy representation

$$\mathrm{hol}_C : \pi_1(|C| \setminus \Sigma) \to \mathrm{Isom}^+(\mathbb{H}^3) \cong \mathrm{PSL}(2, \mathbb{C})$$

that is unique up to conjugation. We consider the topology in the deformation space of C induced by the variety of representations up to conjugation

$$\hom(\pi_1(|C| \setminus \Sigma), \mathrm{PSL}(2, \mathbb{C}))/\mathrm{PSL}(2, \mathbb{C}).$$

Here we are using Ehresmann principle to say that deformations of structures are described by conjugacy classes of representations, cf. [13].

Notice that not all representations in $\hom(\pi_1(|C| \setminus \Sigma), \mathrm{PSL}(2, \mathbb{C}))/\mathrm{PSL}(2, \mathbb{C})$ close to the holonomy of the initial cone manifold correspond to the holonomy of a cone manifold structure: we must require that the holonomy of the meridians of Σ are rotations.

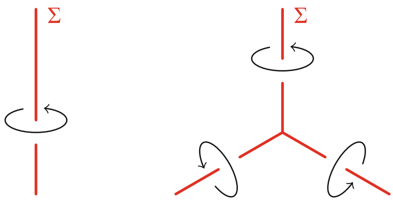


Fig. 4.6 Loops representing meridians of the singular locus in $\pi_1(|C| \setminus \Sigma)$

Theorem 4.4 (Hyperbolic Local Rigidity [26, 33, 51, 53]) *Let C be a compact orientable hyperbolic 3-manifold with topological type* $(|C|, \Sigma)$. *Then the deformation space with fixed topological type is locally parameterized by the cone angles (in particular it cannot be deformed without changing the cone angles).*

This theorem was first proved by Hodgson and Kerckhoff [26] when the singular locus Σ is a link. For arbitrary Σ but cone angles $\leq \pi$ (hence the singular locus is a trivalent graph) it was proved by Weiss [51], and the general case was proved independently by Montcouquiol-Mazzeo [33] and Weiss [53]. The approach of Hodgson–Kerckhoff and Weiss uses infinitesimal deformations as differential forms valued on the Lie algebra and their cohomology theory, though Mazzeo and Montcouquiol use the deformation theory of Einstein metrics.

The local rigidity theorem requires a fixed topological type $(|C|, \Sigma)$. This hypothesis is satisfied when the singular locus Σ is a manifold (there are no singular vertices) or when all cone angles are at most π. In general there are deformations that may change the singular locus: for instance a 4-valent vertex of Σ may split into two 3-valent vertices joined by a graph (this does not change the topology of $|C| \setminus \Sigma$). See [36].

Infinitesimal rigidity has been generalized by Bromberg to noncompact geometrically finite manifolds (without rank one cusps nor singular vertices):

Theorem 4.5 ([9]) *If C^3 is a geometrically finite cone-manifold without rank one cusps and if all cone angles are $\leq 2\pi$, then M is locally rigid relative to the cone angles and the conformal boundary.*

Remark 4.7 There is a stronger notion, *infinitesimal rigidity*, that implies local rigidity. In fact Theorems 4.4 and 4.5 are proved by establishing infinitesimal rigidity first.

When the cone angles are larger than 2π, infinitesimal rigidity does not hold. In a talk at the Third MSJ regional workshop in Tokyo in 1998 (devoted to the orbifold theorem), Casson gave an example of infinitesimally non-rigid hyperbolic cone 3-manifolds with singular vertices. Izmestiev has given further examples of infinitesimally non-rigid hyperbolic cone 3-manifolds, including examples without singular vertices. Furthermore, Izmestiev has provided examples that are not locally rigid in [28].

We conclude the section discussing spherical and Euclidean geometry.

Theorem 4.6 (Spherical Local Rigidity [51]) *Let C be a spherical cone 3-manifold with cone angles $\leq \pi$ and such that the topological pair* $(|C|, \Sigma)$ *is not Seifert fibered; then it is locally rigid.*

We say that the pair $(|C|, \Sigma)$ is Seifert fibered when $|C|$ is Seifert fibered and Σ consists of fibres. The Seifert fibered case has been discussed by Kolpakov in [32]. Essentially, it corresponds to the deformation space of the basis.

By means of polyhedra, Schlenker constructed examples of spherical cone 3-manifolds that are not locally rigid, with singular vertices and cone angles $\leq 2\pi$

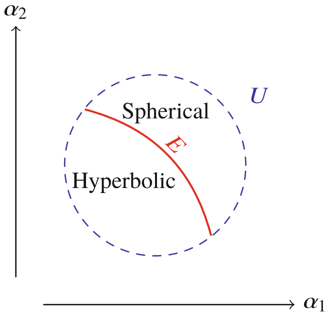


Fig. 4.7 The open set U in the space of multiangles and the hypersurface $E \subset U$ as in Theorem 4.7

[43]. Without singular vertices and allowing cone angles $\geq 2\pi$, non locally rigid spherical cone manifolds are found in [38].

Theorem 4.7 (Euclidean Local Rigidity [41]) *Let C be a closed orientable Euclidean cone 3-manifold with cone angles $\leq \pi$. If C is not an almost product, then in a neighborhood U of the space of multiangles there is a cone manifold structure with topological type $(|C|, \Sigma)$ with these angles. To determine the type of structure, there exists a smooth, properly embedded hypersurface $E \subset U$ consisting of multiangles of Euclidean cone structures that splits U into 2-connected components corresponding to multiangles of spherical and hyperbolic cone structures respectively, Fig. 4.7.*

Furthermore, for each $\bar{\alpha} \in E$ the tangent space of E at $\bar{\alpha}$ is orthogonal to the vector of singular lengths $\bar{l}$.

Almost product means that it can be realized as a product $C^2 \times S^1$ divided by a finite group of isometries. For instance the cone manifold structure on S^3 with singular locus the Borromean rings and cone angle π is an almost product. In Sect. 4.6 we describe the deformation space of the Borromean rings, as well as an example that illustrates Theorem 4.7.

The last result we review in this section is Schläfli's formula. It is named so because it can be established from the classical formula for the volume variation in a family of polyhedra of constant curvature (due to Schläfli for spherical tetrahedra). See for instance [24, 37].

Proposition 4.7 (Schläfli's Formula) *Let C_t be a deformation of cone manifolds of constant curvature κ, for $t \in I$. Assume that it has fixed topological type $(|C|, \Sigma_C)$ and that it is of class C^1 (in the variety of representations of $|C| \setminus \Sigma_C$). Then the volume is differentiable and*

$$\kappa \frac{d\,\mathrm{Vol}(C_t)}{dt} = \frac{1}{2} \sum_e l_e \frac{d\alpha_e}{dt},$$

where the sum runs on the singular edges and circles e of Σ, l_e denotes the length and α_e, the cone angle at e.

A consequence of this formula is that, when cone angles increase, then the volume decreases for hyperbolic cone manifolds, but the volume increases for spherical cone manifolds. It also explains why the space of multiangles of Euclidean structures is perpendicular to the vector of singular lengths in Theorem 4.7.

4.5 Sequences of Cone Manifolds

After reviewing results that allow us to deform cone angles, we look for applications by considering sequences of cone manifolds with fixed topological type. We start with the notion of convergence and a compactness result in Sect. 4.5.1. Then, in Sect. 4.5.2 we analyze the thin part, in order to describe the possible limiting cone manifolds. Finally, applications are described in Sects. 4.5.3 and 4.5.4, by decreasing and increasing respectively the cone angles.

4.5.1 Compactness Theorem

Let C be a compact cone 3-manifold of constant curvature κ. By definition, for every $x \in C$ a metric ball $B(x, \varepsilon)$ centered at x of radius $\varepsilon > 0$ is isometric to the cone (of curvature κ) of its link $\mathrm{Link}(x)$, see Definition 4.2, which is a spherical cone surface:

$$B(x, \varepsilon) \cong \mathrm{Cone}_{\kappa,\varepsilon}(\mathrm{Link}(x)).$$

This is called a *standard ball*.

Definition 4.6 The *injectivity radius* of C at x is

$$\mathrm{inj}(x) = \sup\{\delta > 0 \text{ such that } B(x, \delta) \text{ is standard ball in } C\}.$$

The *cone-injectivity radius* of C at x is

$$\mathrm{cinj}(x) = \sup\{\delta > 0 \text{ such that } B(x, \delta) \text{ is contained in a standard ball in } C\}.$$

Notice that in a compact cone manifold, a point x can be arbitrarily close to the singular locus, therefore its injectivity radius can be arbitrarily small, this is why Thurston defined the cone injectivity radius. The standard ball in the definition of cone injectivity radius does not need to be centered at x, in this way regular points arbitrarily close to the singular locus may have cone-injectivity radius away from zero. The definition of injectivity radius $\mathrm{inj}(x)$ can also be given in terms of the exponential map.

Let X and Y be metric spaces and $\varepsilon > 0$. We call a map $\phi : X \to Y$ a $(1+\varepsilon)$-*bi-Lipschitz embedding* if

$$\frac{1}{1+\varepsilon} < \frac{d(\phi(x_1), \phi(x_2))}{d(x_1, x_2)} < (1+\varepsilon)$$

holds for all $x_1 \neq x_2 \in X$.

Definition 4.7 (Geometric Convergence) Let $(C_n, x_n)_{n\in\mathbb{N}}$ be a sequence of pointed cone-3-manifolds. We say that the sequence (C_n, x_n) converges *geometrically* to a pointed cone-3-manifold (C_∞, x_∞) if for every $R > 0$ and $\varepsilon > 0$ there exists $N = N(R, \varepsilon) \in \mathbb{N}$ such that for all $n \geq N$ there is a $(1+\varepsilon)$-bi-Lipschitz embedding $\phi_n : B_R(x_\infty) \to C_n$ satisfying:

1. $d(\phi_n(x_\infty), x_n) < \varepsilon$,
2. $B(x_n, (1-\varepsilon)R) \subset \phi_n(B(x_\infty, R))$, and
3. $\phi_n(B(x_\infty, R) \cap \Sigma_\infty) = \phi_n(B(x_\infty, R)) \cap \Sigma_n$.

If the C_n have curvature $\kappa_n \in \mathbb{R}$, then C_∞ has curvature $\kappa_\infty = \lim_{n\to\infty} \kappa_n$. The cone-angle along an edge of Σ_∞ is the limit of the cone-angles along the corresponding edge in Σ_n. Notice also that part of the singular locus of the approximating cone-3-manifolds may disappear at the limit by going to infinity.

Theorem 4.8 (Compactness) *Let* $(C_n, x_n)_{n\in\mathbb{N}}$ *be a sequence of pointed cone-3-manifolds with curvature* $\kappa_n \in [-1, 1]$ *and cone-angles* $\leq \pi$. *Suppose that for some* $\rho > 0$, $\operatorname{inj}(x_n) > \rho$. *Then (possibly after passing to a subsequence) the sequence* (C_n, x_n) *converges geometrically to a pointed cone-3-manifold* (C_∞, x_∞) *with curvature* $\kappa_\infty = \lim_{n\to\infty} \kappa_n$.

There are two remarks to be made:

- Firstly, we fix a lower bound $\rho > 0$ on the injectivity radius of the base point x_n, not the cone-injectivity radius. We can use the cone injectivity radius if we fix a lower bound away from zero for the cone angles.
- Secondly, Theorem 4.8 is analogous to a well known compactness theorem for sequences of pointed Riemannian manifolds with bounded sectional curvature and injectivity radius at the base point bounded away from zero. One of the main steps is to establish a uniform lower bound on the cone-injectivity radius at every point in balls $B(x_n, R)$, depending only on ρ and R. This uses the hypothesis on cone angles $\leq \pi$, see Proposition 4.5.

In view of applications we consider sequences of cone manifolds with fixed topological type $(|C|, \Sigma)$ and with bounded volume. To analyze the limits, we need to understand non compact hyperbolic cone manifolds with finite volume. In particular their thin or cone-thin parts.

4.5.2 Cone-Thin Part

For a non-singular hyperbolic 3-manifold, Margulis theorem yields a description of the set of points with injectivity radius less than a uniform constant μ_3, called the Margulis constant. Those are either cusps or tubular neighborhoods of short geodesics. Besides the cone manifold version of cusps and tubes, we still need another model to describe regions with small injectivity radius, called *necks*.

Let $S^2(\alpha, \beta, \gamma)$ be a turnover, with constant curvature -1, 0 or $+1$ according to the sign of $\alpha + \beta + \gamma - 2\pi$, Example 4.1 and Fig. 4.2. View it as the double of a triangle $T = T(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$ and consider the following constructions:

- When $\alpha + \beta + \gamma < 2\pi$, view the triangle $T = T(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$ in a totally geodesic plane $\mathbb{H}^2 \subset \mathbb{H}^3$. Consider

  $$N_R(T) = \{x \in \mathbb{H}^3 \mid d(x, \mathbb{H}^2) \leq R \text{ and } \mathrm{pr}(x) \in T\}$$

  where $\mathrm{pr} : \mathbb{H}^3 \to \mathbb{H}^2$ denotes the orthogonal projection, see Fig. 4.8.

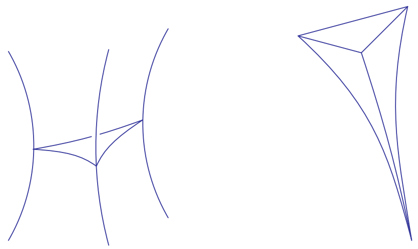
  A *neck* of radius R over $S^2(\alpha, \beta, \gamma)$ is the double of $N_R(T)$ along $\partial T \times [-R, R]$. In the smooth part, the metric is written locally as

  $$ds^2 = dt^2 + \cosh^2(t)\left(dr^2 + \sinh^2(r)d\theta^2\right),$$

  where $\left(dr^2 + \sinh^2(r)d\theta^2\right)$ is the hyperbolic metric on the smooth part of $S^2(\alpha, \beta, \gamma)$ and $t \in [-R, R]$ is the signed distance to the turnover.

  The boundary of a neck consists of two umbilical turnovers of curvature $-\cosh^{-2}(R)$.
- When $\alpha + \beta + \gamma = 2\pi$, view the triangle $T = T(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$ in a horosphere $\mathcal{H}$ centered at an ideal point $\mathrm{center}(\mathcal{H}) \in \partial_\infty \mathbb{H}^3$. Consider

  $$N_\infty(T) = \{x \in \mathbb{H}^3 \mid x \text{ lies in a geodesic from } T \text{ to } \mathrm{center}(\mathcal{H})\},$$



$N_R(T(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2}))$ $\alpha + \beta + \gamma < 2\pi$

$N_\infty(T(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2}))$ $\alpha + \beta + \gamma = 2\pi$

Fig. 4.8 The models whose double are a neck (left) and a cusp (right) see Fig. 4.8. A *cusp* with horospherical cross-section $S^2(\alpha, \beta, \gamma)$ is the double of $N_\infty(T)$ along $\partial T \times [0, +\infty)$. In the smooth part, the metric is locally written as

$$ds^2 = dt^2 + e^{-2t}\left(dr^2 + r^2 d\theta^2\right),$$

where $\left(dr^2 + r^2 d\theta^2\right)$ is the Euclidean metric on the smooth part of $S^2(\alpha, \beta, \gamma)$ and $t \in (0, +\infty)$ is the distance to the turnover.

The boundary of a cusp is an umbilical Euclidean turnover.

Those necks and cusps are naturally foliated by umbilical cone surfaces, that are turnovers. Here we have considered hyperbolic and Euclidean turnovers. For spherical turnovers the corresponding region foliated by umbilical turnovers is a standard ball.

Cusps have always small cone injectivity radii. Hyperbolic necks may have small cone injectivity radius. If we fix a lower bond on the cone angle, small cone injectivity radius at necks only occurs when $\alpha + \beta + \gamma$ approaches π.

Theorem 4.9 (Cone-Thin Part [4]) *For* $D > 0$ *and* $0 < \alpha \leq \beta < \pi$ *there exists* $\rho = \rho(D, \alpha, \beta) > 0$ *such that the following holds: If* C *is an orientable cone-3-manifold (without boundary) of constant curvature* $\kappa \in [-1, 0)$ *with cone-angles* $\in [\alpha, \beta]$ *and* $\operatorname{diam}(X) \geq D$*, then the set of points* $\{x \in C \mid \operatorname{cinj}(x) < \rho\}$ *is contained in the disjoint union of:*

1. *Tubular neighborhoods of (perhaps singular) closed geodesics.*
2. *Cusps with horospherical cross-section a 2-torus or a Euclidean turnover*
3. *Necks on a hyperbolic turnover.*

Here are some remarks about Theorem 4.9:

- The theorem assumes that the diameter is larger than a positive constant $D > 0$. In fact there are hyperbolic cone manifolds with small cone-injectivity radius everywhere, but they have small diameter: they correspond to sequences of hyperbolic cone manifolds that collapse to a point.
- The theorem does not hold when we allow cone angles close to π: we show in Sect. 4.6 sequences of hyperbolic cone manifolds that Hausdorff converge to a two-dimensional cone manifold, hence with a positive lower bound of the diameter.
- Notice that the necks describe the only way two singular edges can approach, under the assumptions that cone angles are bounded above away from π and that the diameter is bounded below away from zero.

Remark 4.8 In [4] a stronger version of this theorem is proved, with the description of points with injectivity radius less than some constant, i.e. including regular points close to the singularity. This includes cones over turnovers, that have a large cone injectivity radius but small injectivity radius at all regular points. One of the conclusions is that the boundary of the components of the thin part includes a point with large injectivity radius.

Theorem 4.9 and the stronger statement in Remark 4.8 need a careful analysis to construct, from short loops, foliations by umbilical surfaces. Theorem 4.9 can also be proved from the classification of non-compact Euclidean cone manifolds with cone angles less than π.

Next we give two applications of Theorem 4.9. Notice that the boundary of the neighborhoods of small cone-injectivity radius contains always a point with large injectivity radius. Thus we have:

Corollary 4.3 (Thickness) *There exists $r = r(D, \alpha, \beta) > 0$ such that if C is as in Theorem 4.9, then C contains an embedded smooth standard ball of radius r.*

Corollary 4.4 (Finiteness) *Let C be as in Theorem 4.9 and suppose in addition that $vol(C) < \infty$. Then C has finitely many ends and all of them are (smooth or singular) cusps with compact horospherical cross-sections.*

4.5.3 Decreasing Cone Angles: Global Rigidity

Definition 4.8 We say that a hyperbolic cone 3-manifold C is *globally rigid* if, when C' is a hyperbolic cone manifold with the same topological type, $(|C'|, \Sigma_{|C'|}) \cong (|C|, \Sigma_{|C|})$, and the same cone angles, then C' is isometric to C.

Theorem 4.10 (Hyperbolic Global Rigidity [31, 52]) *Hyperbolic cone manifolds with cone angles less than π are globally rigid.*

This theorem was first proved by Kojima in [31] when there are no singular vertices, using the local rigidity theorem of Hodgson and Kerckhoff, available at that time, and the case with vertices was proved by Weiss in [52], after he had proved local rigidity when there are vertices.

Here is a sketch of the proof. Assume first that C has no singular vertices, i.e. that Σ is a link. The proof consists in decreasing the cone angles, until one reaches a hyperbolic orbifold. The angles can be decreased by local rigidity, and one has to analyze the limits to prove that the space of angles realized by a hyperbolic cone structure on $(|C|, \Sigma)$ is not only open but closed. We consider sequences of cone manifold structures with decreasing cone angles. The volume of these sequences increases (by Schläfli's formula), in particular the diameter is bounded below by $D > 0$. Furthermore the volume of C is bounded above by the volume of the complete structure on $|C| \setminus \Sigma$, Remark 4.4. As the diameter is bounded below by $D > 0$, by the compactness theorem (Theorem 4.8) the sequence of cone manifolds converges to a finite volume hyperbolic manifold C_∞. If C_∞ is compact, then it has the same topological type as C_n, which means that we can continue decreasing the angles. If C_∞ is non compact, then one uses the finiteness theorem (Corollary 4.4) and a topological argument to get a contradiction with the opening of cusps.

When C has singular vertices, then one has to take into account that some of the singular vertices can go to infinity, i.e. the cone on a spherical turnover becomes a cusp with horospherical cross-section a turnover.

Once one reaches cone angles that are $2\pi/n$, the argument concludes from Mostow–Prasad rigidity on orbifolds: the structure on the orbifold is unique, and, by local rigidity, the path to reach it is also unique.

In the spherical case, Weiss establishes also global rigidity by increasing cone angles; here Mostow–Prasad is replaced by a rigidity theorem in the spherical case due to de Rham:

Theorem 4.11 (Spherical Global Rigidity [52]) *Non Seifert fibered spherical cone manifolds with cone angles less than π are globally rigid.*

From Theorems 4.7, 4.10 and 4.11, we get:

Theorem 4.12 (Euclidean Global Rigidity [41]) *Let C be a closed orientable Euclidean cone 3-manifold with cone angles $\leq \pi$. If C is not an almost product, then C is globally rigid (up to homoteties).*

Furthermore, for every multiangle $\bar{\alpha} \in (0, \pi)^q$ there exists a unique cone manifold structure of constant curvature in $\{-1, 0, 1\}$ on C with those cone angles:

- *If all cone angles of C are π, then every point in $(0, \pi)^q$ is the multiangle of a hyperbolic cone structure on C.*
- *If some of the cone angles is $< \pi$, then the subset $E \subseteq (0, \pi)^q$ of multiangles of Euclidean cone structures is a smooth, properly embedded hypersurface that splits $(0, \pi)^q$ into 2 connected components, corresponding to multiangles of spherical and hyperbolic cone structures respectively.*

Sequences of cone manifolds without singular vertices can also be analyzed by controlling the radius of a metric tube around the singular locus, so the singular locus does not crosses itself. This is the technique of Hodgson and Kerckhoff to prove Theorem 4.3, and by decreasing the cone angle they also prove the following theorem for short geodesics (cf. Proposition 4.6):

Theorem 4.13 ([27]) *Let M^3 be a closed hyperbolic 3-manifold and γ a geodesic in M^3 of length less than 0.111; then there exists a family of hyperbolic cone structures on M^3 with singular locus γ and cone angle in $[0, 2\pi]$ (the cone angles decrease from 2π, the non-singular metric, to 0, the complete structure on $M^3 \setminus \gamma$).*

This has applications in Kleinian groups [7, 8, 10].

4.5.4 Increasing Cone Angles

Next we discuss sequences of cone manifolds with increasing cone angles. We assume that the cone angles are bounded above away from π.

Theorem 4.14 *Let C_n be a sequence of compact hyperbolic cone manifolds with fixed topological type and increasing cone angles that are bounded above by $\eta < \pi$. Then, up to a subsequence, there are three possibilities:*

- *It converges geometrically (Definition 4.7) to a compact hyperbolic cone manifold with the same topological type.*
- *It converges geometrically to a hyperbolic cone manifold C_∞ of finite volume with cusps, each cusp with horospherical cross-section a turnover (singular cusps opening).*
- *It collapses to a point and, after rescaling, it converges geometrically to a Euclidean cone manifold.*

The idea of the proof is to apply the compactness theorem (Theorem 4.8) and the finiteness theorem (Corollary 4.4). More precisely, if the diameter of C_n stays bounded below away from zero, then we apply the compactness theorem, and the limit C_∞ is a manifold of finite volume (the deformation decreases the volume by Schläfli's formula). Furthermore we can get rid of the case where C_∞ has some nonsingular cusp by a topological argument on Dehn fillings. Hence all cusps of C_∞ are singular, and they have horospherical cross-section a turnover. This yields the first two items of the conclusion of the theorem. The remaining case occurs when the diameter of C_n converges to zero: then the cone manifold collapses to a point. In this case we rescale by the diameter, so that the curvature converges to zero. We apply again the compactness theorem and we get convergence to a compact Euclidean cone manifold (of diameter one).

Recall that Theorem 4.2 says that, for an orientable hyperbolic manifold with a single cusp M^3 and for any slope q, we have $\Theta_q \geq \frac{2\pi}{3}$, i.e. the cone manifold is hyperbolic for cone angles $\alpha \in (0, \frac{2\pi}{3})$. With all the results we have reviewed, we can sketch its proof.

Sketch of the Proof of Theorem 4.2 in the Introduction
Assume that for some slope q, $\Theta_q < \frac{2\pi}{3}$ and, seeking a contradiction, consider a sequence of angles $\alpha_n < \Theta_q$ converging to Θ_q. Apply Theorem 4.14; then there are three possibilities. The first one is that the sequence converges to a compact hyperbolic manifold with the same topological type. In this case the cone angle can be increased by the local rigidity theorem and we get a contradiction by the definition of Θ_q. The second case of Theorem 4.14 is that a singular cusp opens, with horospherical cross-section a Euclidean turnover. But a Euclidean turnover has at least one cone angle $\geq \frac{2\pi}{3}$. Therefore this case does not occur because $\Theta_q < \frac{2\pi}{3}$. The third case is that the sequence of cone manifolds collapses to a Euclidean cone manifold with cone angle Θ_q. Since $\Theta_q < \frac{2\pi}{3}$, the Euclidean cone manifold is not an almost product. By Theorem 4.7 the cone angle can be increased to be spherical. Then, by Weiss's theorem (Theorem 4.11 and its proof), or by Theorem 4.12, the cone manifold with cone angle $\frac{2\pi}{3}$ is spherical. As the cone angle is $\frac{2\pi}{3}$ and M^3 (the smooth part) is hyperbolic, then the spherical orbifold is not Seifert fibered. Finally, we look at the classification of Dunbar of spherical orbifolds that are not Seifert fibered [19], and we reach a contradiction.

This finishes the sketch of the proof of Theorem 4.2.

4.6 Examples

In this section we discuss a few examples of deformations of cone manifolds, possibly with cone angles π or larger.

4.6.1 Hyperbolic Two-Bridge Knots and Links

A two-bridge knot or link is the result of gluing two tangles (a tangle is the pair formed by a ball with two unknotted arcs), Figs. 4.9 and 4.10. Such a link is either a torus link or hyperbolic. See [11] for details.

We next discuss the *canonical tunnels*. The arcs in the tangle may be joined by a third arc, called tunnel, so that they form a letter H shape, Fig. 4.9. These arcs are indeed tunnels: the exterior of the union of the link and any of the tunnels is a handlebody. These tunnels are geodesic in the complete hyperbolic structure [1], see also [2] for hyperbolic cone manifold structures singular along the tunnels. Here we show that the tunnels play a role in the limit of spherical cone structures.

The double covering of S^3 branched along a two-bridge link L is a lens space (it is the union of two solid tori, the double covering of the balls branched along the tangles). Hence the orbifold on S^3 with ramification locus L and branching index 2 is spherical.

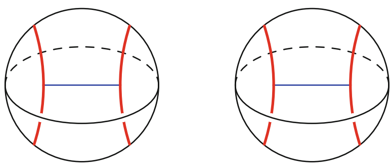


Fig. 4.9 Two tangles, with the arcs and the canonical tunnels. The union along the boundaries yields a 2-bridge link of one or two components

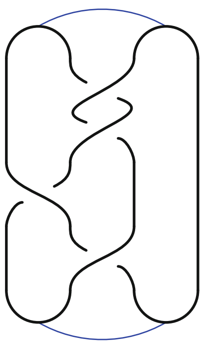


Fig. 4.10 The figure eight knot as a two bridge knot. The canonical tunnels are represented by a blue thin line

Proposition 4.8 ([39]) *Let L be a hyperbolic 2-bridge knot or link. There exists $\alpha_{\mathrm{Euc}} \in [\frac{2\pi}{3}, \pi)$ such that S^3 has a cone manifold structure with singular locus L and cone angle α (the same cone angle in both components if it is a link) of the following type:*

- *hyperbolic for $\alpha \in (0, \alpha_{\mathrm{Euc}})$,*
- *Euclidean for $\alpha = \alpha_{\mathrm{Euc}}$,*
- *spherical for $\alpha \in (\alpha_{\mathrm{Euc}}, 2\pi - \alpha_{\mathrm{Euc}})$.*

Furthermore, when $\alpha \to 2\pi - \alpha_{\mathrm{Euc}}$ the singular locus intersects itself transversely along two points (the length of the canonical tunnels converges to zero) and the cone manifold converges to the spherical suspension of a sphere with four cone points of cone angle $2\pi - \alpha_{\mathrm{Euc}}$.

From Theorem 4.1 the cone manifold is hyperbolic for angles in the interval $(0, \frac{2\pi}{3})$. Furthermore, as it is spherical for angle $\alpha = \pi$, it has to become Euclidean at some angle $\alpha_{\mathrm{Euc}} \in [\frac{2\pi}{3}, \pi)$, by Boileau and Porti [3, Appendix A]. By Theorem 4.12 it is spherical for $\alpha \in (\alpha_{\mathrm{Euc}}, \pi]$. The spherical structures with cone angles $(\pi, 2\pi - \alpha_{\mathrm{Euc}})$ are constructed in [39], using the symmetry of the variety of representations of $\pi_1(S^3 \setminus K)$ in $\mathrm{SU}(2)$, as $\mathrm{SU}(2) \times \mathrm{SU}(2)$ is the universal covering of $\mathrm{SO}(4)$. In [35] the explicit example of the figure-eight knot is explained.

Notice that for the figure eight knot $\alpha_{\mathrm{Euc}} = \frac{2\pi}{3}$. From Dunbar's classification or Euclidean orbifolds [18], form any other 2-bridge knot or link $\alpha_{\mathrm{Euc}} > \frac{2\pi}{3}$.

For links we may consider different cone angles on each component, Theorem 4.12 applies. We describe it with one example, the Whitehead link.

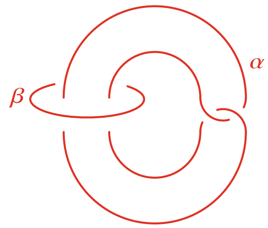
Example 4.4 Consider the cone manifold structures on S^3 with singular locus the Whitehead link, and cone angles α and β (Proposition 4.8 assumes $\alpha = \beta$), Fig. 4.11. Cone manifold structures have been described by several authors, for instance Shmatkov [44]. Here we follow [41].

For $(\alpha, \beta) \in [0, \pi)^2$ there exists a cone manifold structure on S^3 with singular locus the Whitehead link and angles α, and β according to Fig. 4.12.

The curve of Euclidean cone manifolds is described by

$$x^6y^2 - 2x^4y^4 + 2x^4y^2 + x^2y^6 + 2x^2y^4 - 11x^2y^2 + 32 - 48x^2 - 48y^2 + 24y^4 + 24x^4 - 4x^6 - 4y^6 = 0. \qquad (4.3)$$

Fig. 4.11 Cone manifold structure on S^3 with singular locus the Whitehead link and cone angles α and β



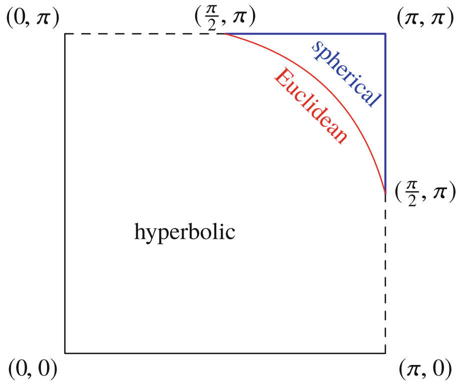


Fig. 4.12 The kind of geometric structures on the Whitehead link according to cone angles α and β

where $x = \pm 2\cos(\alpha/2)$ and $y = \pm 2\cos(\beta/2)$. Here is an explanation of Eq. (4.3). The fundamental group of a two bridge link exterior $S^3 - L$ is generated by two elements μ_1 and μ_2, that are represented by meridians. The variety of $\mathrm{SL}(2,\mathbb{C})$-characters of $\pi_1(S^3 - L)$ is an affine surface in $\mathbb{C}^3$, with coordinates $x([\rho]) = \mathrm{trace}(\rho(\mu_1))$, $y([\rho]) = \mathrm{trace}(\rho(\mu_2))$ and $z([\rho]) = \mathrm{trace}(\rho(\mu_1\mu_2))$, for every conjugacy class (or character) of a representation $\rho\colon \pi_1(S^3 - L) \to \mathrm{SL}(2,\mathbb{C})$. Then the curve (4.3) is the discriminant of the projection of the variety of characters to the plane with coordinates (x, y), intersected with $\mathbb{R}^2$.

For fixed $\beta < \pi$, when $\alpha \to \pi^-$:

- for $\beta < \pi/2$ the cone manifold collapses to a two-dimensional hyperbolic cone manifold with boundary.
- for $\beta = \pi/2$ it collapses to a point (the corresponding orbifold has Nil geometry, see [46] and [38]).
- $\beta \in (\pi/2, \pi]$, the limit is a spherical cone 3-manifold.

This is because the double branched covering along one of the components of the Whitehead link is again S^3, and the other component lifts to a torus link. This assertion can be extrapolated to general hyperbolic links with two bridges, but the limits $\alpha \to \pi^-$ depend on the geometry of the partial double covering.

4.6.2 Montesinos Links

Montesinos links are links $L \subset S^3$ such that the double covering of S^3 branched along L is Seifert fibered, and the fibration is transverse to the branching locus. For instance, 2-bridge links are Montesinos. The Seifert fibration of the double covering induces an orbifold Seifert fibration of the orbifold structure on S^3 with ramification locus L and ramification index 2, see [6] or [11]. The orbifold basis of this fibration is a 2-orbifold, with underlying space a polygon P_L, mirror edges and corner reflectors (corresponding to rational tangles). The polygonal 2-orbifold

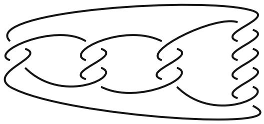


Fig. 4.13 Example of Montesinos knot. When $\alpha \to \pi$, the corresponding hyperbolic cone manifold $C(\alpha)$ collapses to a hyperbolic quadrilateral with angles $\frac{\pi}{2}$, $\frac{\pi}{3}$, $\frac{\pi}{3}$ and $\frac{\pi}{5}$

is geometric: the polygon P_L can be realized in a plane of constant curvature (the angles being π/n for a corner reflector of order $2n$, hence determined by the topology of the link). For a 2-bridge link, P_L is a spherical bigon. For the link L in Fig. 4.13, P_L is a hyperbolic quadrilateral. Notice that when P_L has more than three vertices, then the 2-orbifold has a nontrivial Teichmüller space.

Proposition 4.9 *Let $L \subset S^3$ be a hyperbolic Montesinos link. Consider the cone manifold $C(\alpha)$ with underlying space S^3, branching locus L and cone angle α. Let P_L be the polygonal basis of the orbifold Seifert fibration:*

- *If P_L is spherical, then there exists an angle $\alpha_E \in [\frac{2\pi}{3}, \pi)$ so that $C(\alpha)$ is hyperbolic for $\alpha \in [0, \alpha_E)$, Euclidean for $\alpha = \alpha_E$ and spherical for $\alpha \in (\alpha_E, \pi]$.*
- *Otherwise $C(\alpha)$ is hyperbolic for $\alpha \in [0, \pi)$.*

Furthermore, when P_L is hyperbolic, as $\alpha \to \pi^-$, $C(\alpha)$ Hausdorff converges to the polygon with minimal perimeter among all polygons with given angles.

In the spherical case, the discussion is the same as for two-bridge links. Furthermore, if a collapse occurs before π then P_L must be spherical.

The assertion on the hyperbolic case is proved in [40], including the minimal perimeter of the polygon P_L with given angles.

When P_L is Euclidean, the orbifold has naturally a Nil or Euclidean structure. In the Nil case, for $\alpha > \pi$ the cone manifold $C(\alpha)$ becomes spherical [38]. When P_L is hyperbolic, the natural way to continue the deformations is by means of anti-de Sitter structures [15].

4.6.3 A Cusp Opening

Fix three angles $\alpha, \beta, \gamma \in (0, \pi)$ subject to

$$\alpha + \frac{\gamma}{2} < \pi, \qquad \beta + \frac{\gamma}{2} < \pi.$$

By Andreev's theorem, there exists a truncated hyperbolic tetrahedron with angles α and β at opposite edges, and $\frac{\gamma}{4}$ at the remaining 4 edges. The truncation triangles are totally geodesic and perpendicular to the sides of the tetrahedron, so that we can view the polyhedron as a hyperbolic tetrahedron with vertices outside the hyperbolic space (in the de Sitter sphere). See Fig. 4.14.

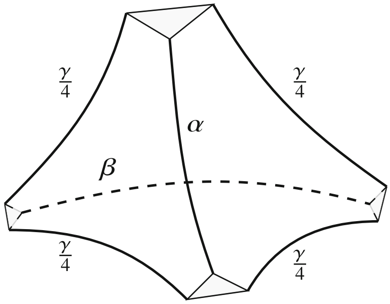


Fig. 4.14 The truncated hyperbolic tetrahedron

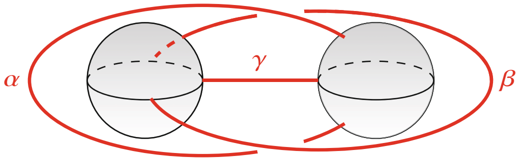


Fig. 4.15 The cone manifold after side pairings of the tetrahedron in Fig. 4.14

To construct a cone manifold identify the faces of the tetrahedron by rotations along the edges of angles α and β. After the identification, the four edges of angles $\frac{\gamma}{4}$ correspond to a single equivalence class. We obtain in this way a cone manifold with totally geodesic boundary consisting of two turnovers $S^2(\alpha, \alpha, \gamma)$ and $S^2(\beta, \beta, \gamma)$, with underlying space $S^2 \times [0, 1]$, and singular locus three arcs of cone angles α, β and γ as in Fig. 4.15.

Notice that when $2\alpha + \gamma = 2\pi$ or when $2\beta + \gamma = 2\pi$, some of the exterior vertices of the truncated tetrahedron in Fig. 4.14 become ideal (i.e. the truncation triangles go to infinity, to an ideal vertex). This means that the corresponding totally geodesic boundary component goes to infinity and the end becomes a cusp, with horospherical cross-section a turnover.

If we furthermore assume $\alpha = \beta$, then we may identify one boundary component with the other by an isometry (turnovers are rigid). In this way we get a family of closed hyperbolic cone manifolds with an embedded totally geodesic turnover when $2\alpha + \gamma \leq 2\pi$, that develops a cusp with horospherical cross-section a turnover when $2\alpha + \gamma \to 2\pi$. This example can be found in [25], see Fig. 4.16.

4.6.4 Borromean Rings

Next we are interested in cone manifold structures on S^3 with singular locus the Borromean rings. Those have been described by many authors, starting by Thurston in his notes [47] for the Euclidean structures, and including for instance [22, 23]. To my knowledge, the different degenerations of hyperbolic structures at angle π are first described in Hodgson's thesis [24], and they are also in [14] (Fig. 4.17).

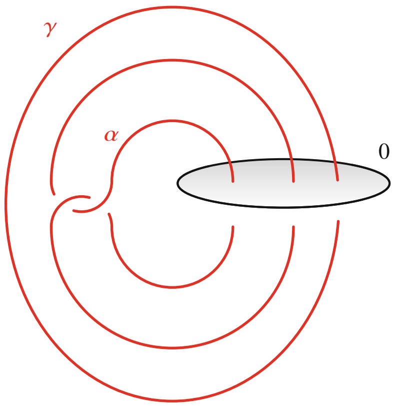


Fig. 4.16 Surgery description of [25], due to Hodgson. When $2\alpha + \gamma < 2\pi$ the turnover is totally geodesic, and when $2\alpha + \gamma \to 2\pi$ it converges to a horospherical turnover

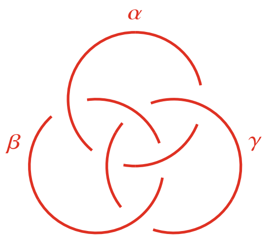


Fig. 4.17 The Borromean rings. They are the singular locus of a hyperbolic cone manifold structure on S^3 with cone angles $\alpha, \beta, \gamma \in [0, \pi)$

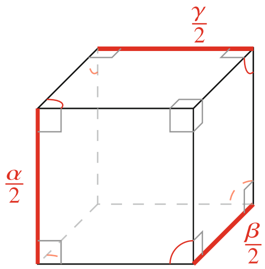


Fig. 4.18 (A Euclidean representation of) the hyperbolic Lambert cube $\mathcal{L}(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$, with three dihedral angles $\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2} \in (0, \frac{\pi}{2})$, the other dihedral angles are $\pi/2$

The building block for the hyperbolic cone manifold structures is the Lambert cube. For $\alpha, \beta, \gamma \in (0, \pi)$, the hyperbolic Lambert cube $\mathcal{L}(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$ is a hyperbolic cube with three dihedral angles $\frac{\alpha}{2}$, $\frac{\beta}{2}$, and $\frac{\gamma}{2}$, as in Fig. 4.18, and all other angles right. By Andreev's theorem, it exists and is unique. Its name comes from its faces, that are Lambert quadrilaterals, Fig. 4.19. The hyperbolic Lambert cube has been considered by several authors, see for instance [12, 17, 29].

We consider eight copies of the Lambert cube $\mathcal{L}(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$, after duplicating it three times, to obtain a polyhedron as in Fig. 4.20. We identify faces of this polyhedron by side pairings along rotations as indicated in Fig. 4.20, so that we get a hyperbolic cone structure on S^3 with singular locus the Borromean rings and cone angles α, β and γ, as explained in Thurston's notes [47].

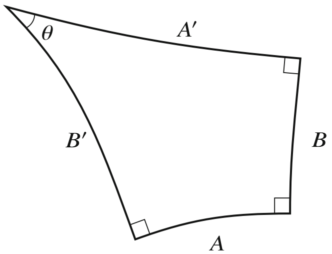


Fig. 4.19 A Lambert quadrilateral with angle $\theta \in (0, \frac{\pi}{2})$. Edges A and B can be arbitrarily short. For a given $\theta \in (0, \frac{\pi}{2})$ the length of A' and B' is bounded below away from zero

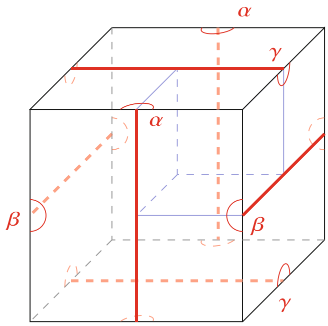


Fig. 4.20 Eight copies of the Lambert cube, after duplicating it three times. We identify the pentagonal faces by rotations along the red axis we obtain the cone manifold $\mathcal{B}(\alpha, \beta, \gamma)$ of Proposition 4.10

Thus we have:

Proposition 4.10 *For every multiangle* $(\alpha, \beta, \gamma) \in [0, \pi)^3$ *there exists a hyperbolic cone structure* $\mathcal{B}(\alpha, \beta, \gamma)$ *on* S^3 *with singular locus the Borromean rings and cone angles* α, β *and* γ.

When some of the angles are zero, we just replace the corresponding edge in the Lambert cube by an ideal point. Notice that Andreev's Theorem applies to the polyhedron of Fig. 4.20, but the computations are easier for the Lambert cube.

Next we ask what happens when some angles converge to π. We do not give the explicit formulas, we just mention that the results below on the limits of Lambert cubes $\mathcal{L}(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$ can be determined from the formulas in [12, 17, 29].

First assume that all angles converge to π.

Lemma 4.1 *When* $\alpha \to \pi^-$, *then the Lambert cube* $\mathcal{L}(\frac{\alpha}{2}, \frac{\alpha}{2}, \frac{\alpha}{2})$ *converges to a point, and after rescaling it converges to a Euclidean cube.*

More precisely, if $\alpha, \beta, \gamma \to \pi^-$ *and the ratios* $\frac{\pi-\alpha}{\pi-\beta}$ *and* $\frac{\pi-\alpha}{\pi-\gamma}$ *converge to positive reals, then* $\mathcal{L}(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$ *converges to a point and, after rescaling it converges to a right rectangular prism.*

Corollary 4.5 *When* $\alpha, \beta, \gamma \to \pi^-$, *and if* $\frac{\pi-\alpha}{\pi-\beta}$ *and* $\frac{\pi-\alpha}{\pi-\gamma}$ *converge to positive real numbers, then* $\mathcal{B}(\alpha, \beta, \gamma)$ *collapses to a point. Furthermore, after rescaling* $\mathcal{B}(\alpha, \beta, \gamma)$ *converges to a Euclidean orbifold.*

This Euclidean orbifold is an almost product and Theorem 4.7 does not apply.

Next assume that one of the angles remains constant.

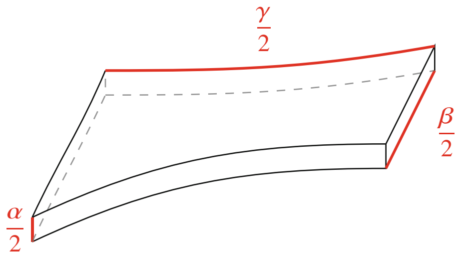


Fig. 4.21 A Lambert cube collapsing to a quadrilateral, when $\beta/2$ and $\gamma/2$ approach $\pi/2$. Four of the Lambert quadrilaterals on the boundary collapse to segments

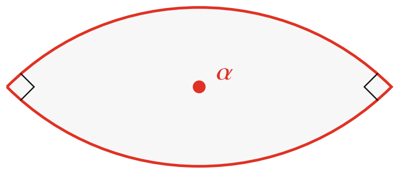


Fig. 4.22 A cone surface that is the limit when $\beta, \gamma \to \pi^-$. The singular components with angles β and γ converge to the segments in the boundary

Lemma 4.2 *Fix $\alpha \in (0, \pi)$. The Hausdorff limit of the Lambert cube $\mathcal{L}(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$ when $\beta, \gamma \to \pi^-$ is a, possibly degenerate, Lambert quadrilateral (a hyperbolic quadrilateral with three right angles and a fourth angle $\alpha/2$, Fig. 4.21), provided that $\frac{\pi-\beta}{\pi-\gamma}$ converges in $[0, +\infty]$.*

Furthermore, any (possibly degenerate) Lambert quadrilateral of angle $\frac{\alpha}{2}$ is realized as a limit, depending on the limit of $\frac{\pi-\beta}{\pi-\gamma}$.

By a possibly degenerate Lambert quadrilateral we mean a triangle with an ideal vertex and two finite vertices, of angles $\frac{\pi}{2}$ and $\frac{\alpha}{2}$.

Again Lemma 4.2 is proved using the formulas for Lambert cubes and quadrilaterals. It is useful to have in mind the following remark, to know what edge lengths can converge to zero:

Remark 4.9 Given $\theta \in (0, \frac{\pi}{2})$, a Lambert quadrilateral is determined by an angle θ and the length of any of the edges, Fig. 4.19. Allowing degenerate Lambert quadrilaterals, the length of an edge takes any value in the interval:

- $[\operatorname{arccosh}(1/\sin(\theta)), +\infty]$, if the edge is adjacent to the vertex of angle θ;
- $[0, +\infty]$, if the edge is disjoint from to the vertex of angle θ.

From Lemma 4.2, by gluing two Lambert quadrilaterals of angle $\frac{\alpha}{2}$ we have:

Corollary 4.6 *For fixed $\alpha \in (0, \pi)$, when $\beta, \gamma \to \pi^-$ and $\frac{\pi-\beta}{\pi-\gamma}$ converges in $[0, +\infty]$, then $\mathcal{B}(\alpha, \beta, \gamma)$ Hausdorff converges to a (possibly degenerate) hyperbolic cone surface with boundary and corners, a bigon with right angles and a cone point α in the interior, Fig. 4.22 (or Fig. 4.25 for the degenerate case).*

Next we fix two angles $\alpha, \beta \in (0, \pi)$ and look at the limit when $\gamma \to \pi^-$. We describe the behavior of its six sides. It can be computed that:

- The sides that are Lambert quadrilaterals of angle $\gamma/2$ collapse to a segment.

- The sides that are Lambert quadrilaterals of angle $\alpha/2$ or $\beta/2$ converge to ideal triangles.

In particular four of the edge lengths converge to zero, four of them converge to infinity, and the remaining four have a non-vanishing finite limit. This can be visualized by a "long" Lambert cube as in Fig. 4.23.

Lemma 4.3 *For fixed $\alpha, \beta \in (0, \pi)$, when $\gamma \to \pi^-$ the diameter of $\mathcal{L}(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$ converges to infinity. There are choices of base points so that the pointed Hausdorff limit is either an ideal triangle of angle $\frac{\alpha}{2}$, an ideal triangle of angle $\frac{\beta}{2}$, or a line. See Fig. 4.23.*

Two phenomena occur simultaneously when $\gamma \to \pi^-$. On the one hand, there is a cusp opening, whose horospherical cross section is a sphere with 4 cone points $S^2(\pi, \pi, \pi, \pi)$ (corresponding to the middle quadrilateral in Fig. 4.23) that separates the cone manifold in two components see Fig. 4.24. On the other hand, each one of these pieces collapses to a hyperbolic cone surface with boundary and finite area, Fig. 4.25. The end of this surface is the quotient of a cusp by an involution, and corresponds to a collapse of the Euclidean cone manifold $S^2(\pi, \pi, \pi, \pi)$ to a segment.

Corollary 4.7 *For fixed $\alpha, \beta \in (0, \pi)$, when $\gamma \to \pi^-$, $\mathcal{B}(\alpha, \beta, \gamma)$ develops a cusp with horospherical cross-section $S^2(\pi, \pi, \pi, \pi)$, that separates $\mathcal{B}(\alpha, \beta, \gamma)$ in two pieces that collapse to cone surfaces as in Fig. 4.25.*

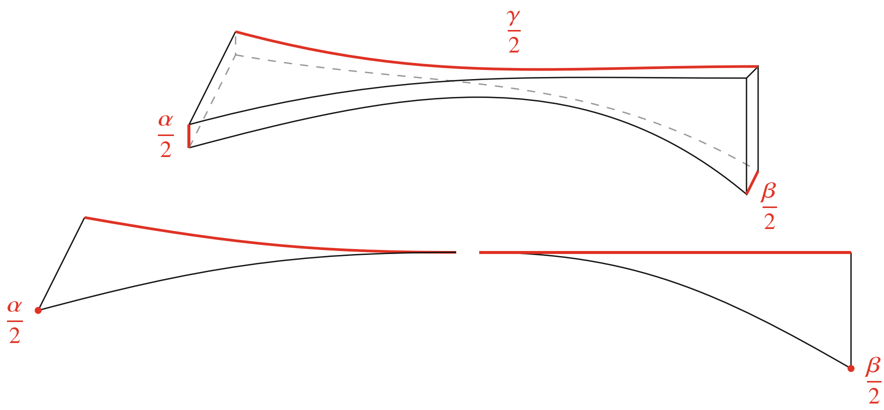


Fig. 4.23 A "long" Lambert cube, when $\gamma/2$ approaches $\pi/2$ (top) and the limiting ideal triangles (bottom)

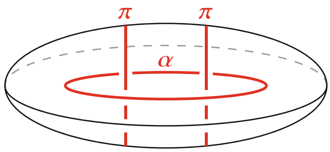


Fig. 4.24 One of the components after splitting $\mathcal{B}(\alpha, \beta, \pi)$ along the Euclidean cone 2-manifold $S^2(\pi, \pi, \pi, \pi)$. It is Seifert fibered over the surface of Fig. 4.25

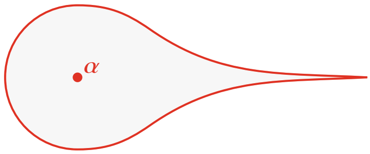


Fig. 4.25 One of the cone surfaces that appear when $\gamma \to \pi^-$ (the other is obtained by replacing α by β)

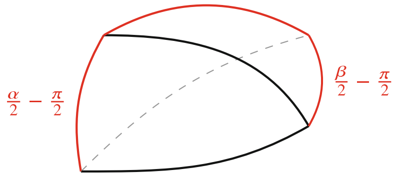


Fig. 4.26 The tetrahedron in Lemma 4.4, with the dihedral angles (when they are not right). The length of an edge is the dihedral angle of the opposite edge, thus $l_\alpha = \frac{\beta}{2} - \frac{\pi}{2}$, $l_\beta = \frac{\alpha}{2} - \frac{\pi}{2}$, and $l_\gamma = \frac{\pi}{2}$

4.6.5 Borromean Rings Revisited: Spherical Structures

Next we consider cone angles $\geq \pi$. For dihedral angles between $\pi/2$ and π, the Lambert cube is spherical, and it has been studied for instance by Díaz [17] and Derevnin and Mednykh [16].

Proposition 4.11 ([17]) *For* $\alpha, \beta, \gamma \in (\pi, 2\pi)$*:*

- *The Lambert cube* $\mathcal{L}(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$ *with dihedral angles* $\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2}$ *is spherical and rigid.*
- S^3 *admits a unique spherical structure with singular locus the Borromean rings and cone angles* (α, β, γ)*,* $\mathcal{B}(\alpha, \beta, \gamma)$*.*

Now we look at the spherical Lambert cube when some dihedral angles approach $\pi/2$ (hence some of the cone angles of $\mathcal{B}(\alpha, \beta, \gamma)$ converges to π).

Lemma 4.4 *When* $\gamma \to \pi^+$ *and* $\alpha, \beta > \pi$ *remain constant,* $\mathcal{L}(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$ *Hausdorff converges to a spherical tetrahedron with right angles, except at two opposite edges, that have angles* $\alpha/2 - \pi/2$ *and* $\beta/2 - \pi/2$*, Fig. 4.26.*

In Lemma 4.4, the edge with dihedral angle $\alpha/2 - \pi/2$ is the result of merging two edges, one with dihedral angle $\alpha/2$ and another one with a right angle, hence its dihedral angle is $\left(\frac{\alpha}{2} + \frac{\pi}{2}\right) - \pi$.

When two of the cone angles converge to π, we have a collapse similar to the hyperbolic case:

Lemma 4.5 *When $\beta, \gamma \to \pi^+$ and $\alpha > \pi$ remains constant, $\mathcal{L}(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$ Hausdorff converges to a spherical Lambert quadrilateral of angle $\frac{\alpha}{2}$, provided that the ratio $\frac{\beta-\pi}{\gamma-\pi}$ converges in $[0, +\infty]$.*

Furthermore, any (possibly degenerate) Lambert quadrilateral of angle $\frac{\alpha}{2}$ is realized as a limit, according to the limit of the ratio $\frac{\beta-\pi}{\gamma-\pi}$.

Finally, the case where all cone angles converge to π^- is similar to the hyperbolic case.

Lemma 4.6 *When $\alpha, \beta, \gamma \to \pi^+$, and the ratios $\frac{\alpha-\pi}{\beta-\pi}$ and $\frac{\alpha-\pi}{\gamma-\pi}$ converge to positive real numbers, then $\mathcal{L}(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$ converges to a point. After rescaling, it converges to a right rectangular prism.*

The translation of the results on Lambert cubes to cone manifolds is the following:

Corollary 4.8

1. *When $\gamma \to \pi^+$, and $\alpha, \beta > \pi$ remain constant, the Hausdorff limit of $\mathcal{B}(\alpha, \beta, \gamma)$ is S^3 with a singular locus as in Fig. 4.27. The singular components of angle α and β intersect the component of angle π and are folded to a segment with cone angle $2\alpha - 2\pi$ and $2\beta - 2\pi$ respectively.*
2. *When $\beta, \gamma \to \pi^+$, $\alpha > \pi$ remains constant and the ratio $\frac{\beta-\pi}{\gamma-\pi}$ converges in $[0, +\infty]$, then $\mathcal{B}(\alpha, \beta, \gamma)$ converges to a cone surface as in Fig. 4.22, possibly degenerate (if the cone point goes to the boundary).*
3. *When $\alpha, \beta, \gamma \to \pi^+$, and the ratios $\frac{\beta-\pi}{\gamma-\pi}$ and $\frac{\alpha-\pi}{\gamma-\pi}$ converge in $(0, +\infty)$, then $\mathcal{B}(\alpha, \beta, \gamma)$ Hausdorff converges to a point, and after rescaling it converges to a Euclidean orbifold.*

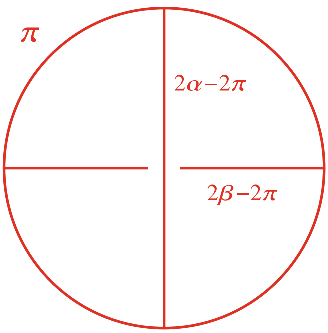


Fig. 4.27 Singular locus of the limit of $\mathcal{B}(\alpha, \beta, \gamma)$ when $\gamma \to \pi^+$

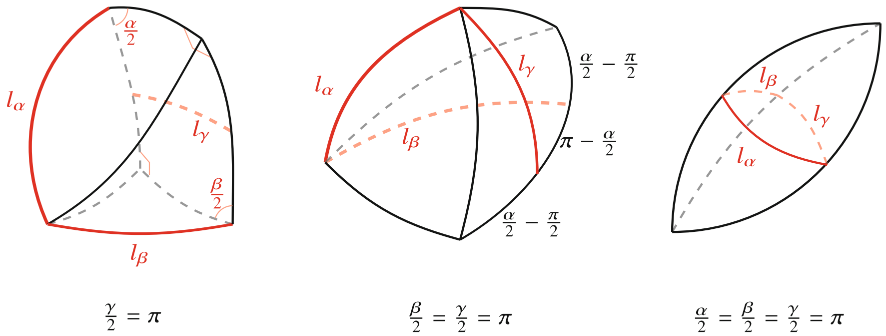


Fig. 4.28 The Hausdorff limit of the spherical Lambert cube when some of the dihedral angles converge to π

We can also consider limits when the cone angles α, β or γ approach 2π; the Hausdorff limits of the spherical Lambert cube $\mathcal{L}(\frac{\alpha}{2}, \frac{\beta}{2}, \frac{\gamma}{2})$ are described in Fig. 4.28.

We describe the limits of the cone manifold in the following remark.

Remark 4.10 When $\alpha \to 2\pi^-$, $\beta \to \beta_0 \in (\pi, 2\pi]$, and $\gamma \to \gamma_0 \in (\pi, 2\pi]$, $\mathcal{B}(\alpha, \beta, \gamma)$ Hausdorff converges to the spherical suspension over a cone surface S. The first singular geodesic converges to a geodesic in S, and, at the limit, the other singular components intersect at the tips of the suspension.

Notice that we allow the limit β_0 or γ_0 to equal 2π. The suspension structure of the remark is obtained from doubling the cones of the Lambert cubes in Fig. 4.28. Hence the Hausdorff limit of $\mathcal{B}(\alpha, \beta, \gamma)$ has a suspension structure for each cone angle that becomes 2π.

Acknowledgements I thank Ken'ichi Ohshika and Athanase Papadopoulos for the oportunity of writing this survey and for their suggestions that improved it. I acknowledge support by the FEDER/AEI/MICIU grant PGC2018-095998-B-I00 "Local and global invariants in geometry".

References

1. C. Adams, Unknotting tunnels in hyperbolic 3-manifolds. Math. Ann. **302**(1), 177–195 (1995)
2. H. Akiyoshi, M. Sakuma, M. Wada, Y. Yamashita, *Punctured Torus Groups and 2-Bridge Knot Groups. I*. Lecture Notes in Mathematics, vol. 1909 (Springer, Berlin, 2007)
3. M. Boileau, J. Porti, Geometrization of 3-orbifolds of cyclic type. Astérisque **272**, 208 (2001)
4. M. Boileau, B. Leeb, J. Porti, Geometrization of 3-dimensional orbifolds. Ann. Math. **162**(1), 195–290 (2005)
5. F. Bonahon, J.-P. Otal, Laminations measurées de plissage des variétés hyperboliques de dimension 3. Ann. Math. **160**(3), 1013–1055 (2004)
6. F. Bonahon, L. Siebenmann, The classification of Seifert fibred 3-orbifolds, in *Low-Dimensional Topology (Chelwood Gate, 1982)*. London Mathematical Society Lecture Note Series, vol. 95 (Cambridge University Press, Cambridge, 1985), pp. 19–85

7. J.F. Brock, K.W. Bromberg, On the density of geometrically finite Kleinian groups. Acta Math. **192**(1), 33–93 (2004)
8. J. Brock, K. Bromberg, R. Evans, J. Souto, Tameness on the boundary and Ahlfors' measure conjecture. Publ. Math. Inst. Hautes Études Sci. **98**, 145–166 (2003)
9. K. Bromberg, Rigidity of geometrically finite hyperbolic cone-manifolds. Geom. Dedicata **105**, 143–170 (2004)
10. K. Bromberg, Projective structures with degenerate holonomy and the Bers density conjecture. Ann. of Math. **166**(1), 77–93 (2007)
11. G. Burde, H. Zieschang, M. Heusener, *Knots*. De Gruyter Studies in Mathematics, vol. 5, extended edition (De Gruyter, Berlin, 2014)
12. P. Buser, A. Mednykh, A. Vesnin, Lambert cubes and the Löbell polyhedron revisited. Adv. Geom. **12**(3), 525–548 (2012)
13. R.D. Canary, D.B.A. Epstein, P. Green, Notes on notes of Thurston, in *Analytical and Geometric Aspects of Hyperbolic Space (Coventry/Durham, 1984)*. London Mathematical Society Lecture Note Series, vol. 111 (Cambridge University Press, Cambridge, 1987), pp. 3–92
14. D. Cooper, C.D. Hodgson, S.P. Kerckhoff, *Three-Dimensional Orbifolds and Cone-Manifolds*.MSJ Memoirs, vol. 5 (Mathematical Society of Japan, Tokyo, 2000). With a postface by Sadayoshi Kojima
15. J. Danciger, A geometric transition from hyperbolic to anti-de Sitter geometry. Geom. Topol. **17**(5), 3077–3134 (2013)
16. D.A. Derevnin, A.D. Mednykh, The volume of the Lambert cube in spherical space. Mat. Zametki **86**(2), 190–201 (2009)
17. R. Díaz, A characterization of Gram matrices of polytopes. Discrete Comput. Geom. **21**(4), 581–601 (1999)
18. W.D. Dunbar, Geometric orbifolds. Rev. Mat. Univ. Complut. Madrid **1**(1–3), 67–99 (1988)
19. W.D. Dunbar, Nonfibering spherical 3-orbifolds. Trans. Am. Math. Soc. **341**(1), 121–142 (1994)
20. N.M. Dunfield, Cyclic surgery, degrees of maps of character curves, and volume rigidity for hyperbolic manifolds. Invent. Math. **136**(3), 623–657 (1999)
21. W.M. Goldman, Characteristic classes and representations of discrete subgroups of Lie groups. Bull. Am. Math. Soc. **6**(1), 91–94 (1982)
22. H.M. Hilden, M.T. Lozano, J.M. Montesinos, W.C. Whitten, On universal groups and three-manifolds. Invent. Math. **87**(3), 441–456 (1987)
23. H.M. Hilden, M.T. Lozano, J.M. Montesinos-Amilibia, On the Borromean orbifolds: geometry and arithmetic, in *Topology'90 (Columbus, OH, 1990)*. Ohio State University Mathemtics Research Institute Publication (de Gruyter, Berlin, 1992), pp. 133–167
24. C.D. Hodgson, Degeneration and regeneration of geometric structures on 3-manifolds. Ph.D. Thesis, Princeton University (1986)
25. C.D. Hodgson, Geometric structures on orbifolds: notes on Thurston's proof. Preliminary version (1988)
26. C.D. Hodgson, S.P. Kerckhoff, Rigidity of hyperbolic cone-manifolds and hyperbolic Dehn surgery. J. Differ. Geom. **48**(1), 1–59 (1998)
27. C.D. Hodgson, S.P. Kerckhoff, Universal bounds for hyperbolic Dehn surgery. Ann. Math. **162**(1), 367–421 (2005)
28. I. Izmestiev, Examples of infinitesimally flexible 3-dimensional hyperbolic cone-manifolds. J. Math. Soc. Japan **63**(2), 581–598 (2011)
29. R. Kellerhals, On the volume of hyperbolic polyhedra. Math. Ann. **285**(4), 541–569 (1989)
30. B. Kleiner, J. Lott, Geometrization of three-dimensional orbifolds *via* Ricci flow. Astérisque 365, 101–177 (2014)
31. S. Kojima, Deformations of hyperbolic 3-cone-manifolds. J. Differ. Geom. **49**(3), 469–516 (1998)
32. A. Kolpakov, Examples of rigid and flexible Seifert fibred cone-manifolds. Glasg. Math. J. **55**(2), 411–429 (2013)

33. R. Mazzeo, G. Montcouquiol, Infinitesimal rigidity of cone-manifolds and the Stoker problem for hyperbolic and Euclidean polyhedra. J. Differ. Geom. **87**(3), 525–576 (2011)
34. R. Mazzeo, H. Weiss, Teichmüller theory for conic surfaces, in *Geometry, Analysis and Probability*. Progress in Mathematics, vol. 310 (Birkhäuser/Springer, Cham, 2017), pp. 127–164
35. A. Mednykh, A. Rasskazov, Volumes and degeneration of cone-structures on the figure-eight knot. Tokyo J. Math. **29**(2), 445–464 (2006)
36. G. Montcouquiol, H. Weiss, Complex twist flows on surface group representations and the local shape of the deformation space of hyperbolic cone-3-manifolds. Geom. Topol. **17**(1), 369–412 (2013)
37. J. Porti, Regenerating hyperbolic and spherical cone structures from Euclidean ones. Topology **37**(2), 365–392 (1998)
38. J. Porti, Regenerating hyperbolic cone structures from Nil. Geom. Topol. **6**, 815–852 (2002)
39. J. Porti, Spherical cone structures on 2-bridge knots and links. Kobe J. Math. **21**(1–2), 61–70 (2004)
40. J. Porti, Regenerating hyperbolic cone 3-manifolds from dimension 2. Ann. Inst. Fourier **63**(5), 1971–2015 (2013)
41. J. Porti, H. Weiss, Deforming Euclidean cone 3-manifolds. Geom. Topol. **11**, 1507–1538 (2007)
42. J.G. Ratcliffe, *Foundations of Hyperbolic Manifolds*. Graduate Texts in Mathematics, vol. 149, 2nd edn. (Springer, New York, 2006)
43. J.-M. Schlenker, Dihedral angles of convex polyhedra. Discrete Comput. Geom. **23**(3), 409–417 (2000)
44. R.N. Shmatkov, Properties of Euclidean Whitehead link cone-manifolds. Siberian Adv. Math. **13**(1), 55–86 (2003)
45. T. Soma, K. Ohshika, S. Kojima, Towards a proof of Thurston's geometrization theorem for orbifolds, no. 568, pp. 1–72 (1985). Hyperbolic geometry and 3-manifolds (Kyoto, 1985).
46. E. Suárez-Peiró, Poliedros de Dirichlet de 3-variedades cónicas y sus deformaciones. Ph.D. Thesis, University of Complutense de Madrid (1998)
47. W.P. Thurston, *The Geometry and Topology of Three-Manifolds* (Princeton University, Princeton, 1976/1980)
48. W.P. Thurston, Three-manifolds with symmetry (preprint, 1982)
49. W.P. Thurston, Three-dimensional manifolds, Kleinian groups and hyperbolic geometry. Bull. Amer. Math. Soc. **6**(3), 357–381 (1982)
50. M. Troyanov, Prescribing curvature on compact surfaces with conical singularities. Trans. Am. Math. Soc. **324**(2), 793–821 (1991)
51. H. Weiss, Local rigidity of 3-dimensional cone-manifolds. J. Differ. Geom. **71**(3), 437–506 (2005)
52. H. Weiss, Global rigidity of 3-dimensional cone-manifolds. J. Differ. Geom. **76**(3), 495–523 (2007)
53. H. Weiss, The deformation theory of hyperbolic cone-3-manifolds with cone-angles less than 2π. Geom. Topol. **17**(1), 329–367 (2013)
54. Q. Zhou, The moduli space of hyperbolic cone structures. J. Differ. Geom. **51**(3), 517–550 (1999)

Chapter 5
A Survey of the Thurston Norm

Takahiro Kitayama

Abstract We present an overview of the study of the Thurston norm, introduced by W. P. Thurston in the seminal paper "A norm for the homology of 3-manifolds" (written in 1976 and published in 1986). We first review fundamental properties of the Thurston norm of a 3-manifold, including a construction of codimension-1 taut foliations from norm-minimizing embedded surfaces, established by D. Gabai. In the main part we describe relationships between the Thurston norm and other topological invariants of a 3-manifold: the Alexander polynomial and its various generalizations, Reidemeister torsion, the Seiberg–Witten invariant, Heegaard Floer homology, the complexity of triangulations and the profinite completion of the fundamental group. Some conjectures and questions on related topics are also collected.

Keywords Thurston norm · Knot genus · 3-manifold · Knot · Fibration · Foliation · Alexander polynomial · Reidemeister torsion · Teichmüller polynomial · Seiberg–Witten invariant · Adjunction inequality · Heegaard Floer homology · Knot Floer homology · Twisted Alexander polynomial · Higher-order Alexander polynomial · L^2-torsion · Normal surface · Complexity of a 3-manifold · Profinite rigidity

2020 Mathematics Subject Classification Primary: 57K31; Secondary: 57K30, 57K41

5.1 Introduction

In the seminal paper [202] written in 1976, Thurston introduced a seminorm on the first real cohomology group of a 3-manifold, called the *Thurston norm*. It

T. Kitayama (✉)
Graduate School of Mathematical Sciences, The University of Tokyo, Tokyo, Japan
e-mail: kitayama@ms.u-tokyo.ac.jp

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_5

measures the topological complexity of embedded surfaces dual to a given integral cohomology class. The unit ball of the seminorm constitutes a convex polyhedron and Thurston described the distribution of the cohomology classes represented by fibrations over a circle in terms of the top-dimensional faces of the polyhedron. Thurston also showed that every compact leaf of a codimension-1 taut foliation minimizes the Thurston norm. The Thurston norm has become a fundamental tool in the study of incompressible surfaces, fibrations over a circle and codimension-1 foliations.

The roots of the Thurston norm go back to the study of the genus of a knot in the 3-sphere, and it has been extensively studied in various contexts up to the present. The goal of this chapter is to present an overview of the study of the Thurston norm, with an emphasis on its relationships with other topological invariants, and without aiming for completeness. We summarize some of the results discussed in this chapter in the following.

Inspired by the foundational work [202] of Thurston, Gabai [74] developed sutured manifold theory, extended by Scharlemann [193], and established a construction of codimension-1 taut foliations from embedded surfaces minimizing the Thurston norm. As consequences Gabai resolved the Property R conjecture and the Poénaru conjecture, and showed the equivalence of the Thurston norm and the Gromov norm [83] on the second homology group. Sutured manifold theory provides an efficient algorithm to compute the Thurston norm. Tollefson and Wang [203, 204], and later Cooper and Tillmann [32] described another algorithm via normal surface theory.

One of the most fundamental algebraic invariants related to the Thurston norm is the *Alexander polynomial*, equivalent to *Milnor torsion* [149, 205, 211]. It is well known that the classical Alexander polynomial of a knot in the 3-sphere gives a lower bound on its genus. As a generalization, McMullen [147] introduced the *Alexander norm* for a general 3-manifold, and showed that it gives a lower bound on the Thurston norm. Also, gauge theory has been a successful tool to study the complexity of embedded surfaces in 3- and 4-manifolds. In the *adjunction inequality* for a 3-manifold the *Seiberg–Witten invariant* [216] gives a lower bound on the Thurston norm [6, 114]. By the equivalence of Milnor torsion and the Seiberg–Witten invariant of a 3-manifold [148, 208], the above two lower bounds coincide [114, 214]. Furthermore, Kronheimer and Mrowka [119] described the Thurston norm in terms of solutions of the Seiberg–Witten monopole equations.

These relationships were generalized in *Heegaard Floer homology* [177, 180] and *monopole Floer homology* [120], which provide categorifications of the Milnor torsion and the Seiberg–Witten invariant of a 3-manifold. These homology theories were also shown to be equivalent [28–30, 125–128]. Ozsváth and Szabó [177] showed that Heegaard Floer homology determines the Thurston norm of a closed 3-manifold, and Ni [165] showed that Heegaard Floer homology detects fiberedness of a closed 3-manifold. Knot Floer homology [178, 189] provides a categorification of the classcal Alexander polynomial of a knot. Ozsváth and Szabó [177] showed that knot Floer homology determines the genus of a knot, and Ghiggini [79], Ni [163], and Juhász [102, 103] showed that knot Floer homology detects fiberedness of a knot.

Twisted Alexander polynomials [132, 215] associated with linear representations, and higher-order Alexander polynomials [27, 85] with coefficients in skew-fields are more direct generalizations of the Alexander polynomial. These polynomials give generalized lower bounds on the Thurston norm [48, 54, 87]. Furthermore, Friedl and Vidussi [64, 65, 67, 69] showed that twisted Alexander polynomials detect fiberedness of a 3-manifold, and Friedl, Nagel and Vidussi [60, 71] showed that twisted Alexander polynomials determine the Thurston norm. The L^2-Alexander invariant or torsion [38, 131] are "polynomial-like" L^2-invariants, generalizing the L^2-torsion [140]. Friedl and Lück [58], and Liu [137] showed that the L^2-Alexander torsion determines the Thurston norm.

Boileau and Friedl [13], Bridson, Reid and Wilton [16, 17], and Liu [138] showed certain rigidity results of the Thurston norm and fiberedness of a 3-manifold on the profinite completion of the fundamental group. Also, Jaco, Rubinstein, Spreer and Tillmann [97, 98, 100] introduced a $\mathbb{Z}/2\mathbb{Z}$-analogue of the Thurston norm and showed that it gives lower bounds on minimal numbers of tetrahedra in triangulations and ideal triangulations of a 3-manifold.

The influence of the Thurston norm is not limited to low-dimensional topology, and the following significant topics are, for example, unfortunately beyond the scope of this article. The universal L^2-torsion defines an equivalence class of a pair of convex polytopes for a torsion-free group satisfying certain conditions. Such an equivalence class can be regarded as the unit ball of the (dual) Thurston norm, and as already shown in [56, 57, 59, 61, 63, 73, 78, 89, 90, 107], there should be a fruitful theory for the "Thurston norm of groups". Calegari [19–23] studied a group-theoretical interpretation of the Thurston norm in terms of the stable commutative length. Also, Flores, Kahrobaei and Koberda [45] proposed a public-key and a symmetric-key cryptographic schemes based on the Thurston norm of hyperbolic 3-manifolds.

For foundational results on the Thurston norm there are already excellent expositions in [77, 168, 193], and also in [24, Chapter 10] and [105, Chapter 2]. See also the survey [192, Section 12] of the impact of Thurston's work on knot theory in the first volume of this series of books. For terminology and developments of the study of 3-manifolds we refer the reader to the book [5].

Throughout we do not attempt to state results in their greatest generality, and we do not make any claims to originality.

Organization

Section 5.2 provides a brief review of the definition and fundamental properties of the Thurston norm, including the correspondence between embedded surfaces minimizing the Thurston norm and codimension-1 taut foliations. Section 5.3 describes the relationships between the norms on the first cohomology group associated with the Alexander and Teichmüller polynomials and the Thurston norm. Section 5.4 discusses adjunction inequalities from Seiberg–Witten theory for 3- and 4-manifolds. Section 5.5 summarizes the facts that Heegaard Floer homology and knot Floer homology detect the Thurston norm, knot genus and fiberedness of a 3-manifold and a knot. Section 5.6 deals with twisted Alexander polynomials, higher-order Alexander polynomials and L^2-Alexander torsion. Here results on the Thurston norm are described in terms of Reidemeister torsion. Section 5.7 contains constructions of the Thurston norm ball via normal surface theory and applications of a $\mathbb{Z}/2\mathbb{Z}$-analogue of the Thurston norm to the study of complexity of a 3-manifold. Section 5.8 is devoted to explain certain rigidity results of the Thurston norm on the profinite completion of the fundamental group. In Sect. 5.9 we conclude by collecting some conjectures and questions on the Thurston norm and related topics.

Conventions and Notation

All surfaces and manifolds are understood to be compact, connected and oriented unless we say specifically otherwise. For a link L in S^3 we denote by X_L the complement of an open tubular neighborhood of L. For an integral domain R we denote by $Q(R)$ its quotient field.

5.2 Foundations of the Thurston Norm

First we briefly review the definition of the Thurston norm of a 3-manifold and its fundamental properties. We summarize original results by Thurston [202] and Gabai [74, 76] on the polyhedron structure of the unit ball of the Thurston norm, the distribution of cohomology classes represented by fibrations over a circle, the correspondence between embedded surfaces minimizing the Thurston norm and codimension-1 taut foliations, and the equivalence of the Thurston and Gromov norms.

5.2.1 Thurston Norm

We begin with the definition of the Thurston norm of a 3-manifold M [202].

For a surface S with connected components $S_1, S_2, \ldots, S_k$ its *complexity* $\chi_-(S)$ is defined by

$$\chi_-(S) = \sum_{i=1}^{k} \max\{-\chi(S_i), 0\},$$

where χ is the Euler characteristic. Every cohomology class in $H^1(M;\mathbb{Z})$ is represented by a smooth map $M \to S^1$ and the properly embedded surface obtained as the inverse image of any regular value represents the Poincaré dual of the cohomology class. The *Thurston norm* x_M on $H^1(M;\mathbb{Z})$ of M is defined by

$$x_M(\phi) = \min\{\chi_-(S)\ ;\ S \text{ is a properly embedded surface in } M \text{ dual to } \phi\}$$

for $\phi \in H^1(M;\mathbb{Z})$.

In [202] Thurston first showed that x_M is a seminorm on $H^1(M;\mathbb{Z})$. Key observations are that the k-multiple of a homology class is represented by $|k|$ disjoint properly embedded surfaces representing the homology class, and that the "double curved sum" of two properly embedded surfaces with transverse intersection represents the sum of their homology classes. Since x_M is linear on each ray through the origin, it extends to $H^1(M;\mathbb{Q})$. Since x_M is a convex function, it extends to all of $H^1(M;\mathbb{R})$ in a unique continuous way. We denote also by x_M the extended seminorm on $H^1(M;\mathbb{R})$. Moreover, for $\phi \in H^1(M;\mathbb{R})$ with $x_M(\phi) = 0$ the ray through ϕ comes arbitrarily near lattice points, and if nonzero multiple $a\phi$ is near enough to a lattice point l, then the integer $x_M(l) = x_M(l - a\phi)$ must be 0. Thus ϕ can be approximated by multiples of lattice points l with $x_M(l) = 0$. Summarizing, we state the following theorem [202, Theorem 1]:

Theorem 5.2.1 ([202]) *The Thurston norm x_M uniquely extends to $H^1(M;\mathbb{R})$ as a seminorm:*

1. $x_M(a\phi) = |a| x_M(\phi)$,
2. $x_M(\phi + \psi) \le x_M(\phi) + x_M(\psi)$,

for ϕ, $\psi \in H^1(M;\mathbb{R})$ and $a \in \mathbb{R}$. Moreover, $x_M^{-1}(\{0\})$ is spanned by integral cohomology classes dual to properly embedded surfaces in M with non-negative Euler characteristic. □

Remark 5.2.2 More generally, the seminorm can be defined on $H_2(M, A;\mathbb{R})$ for any submanifold A in ∂M, as Scharlemann described in [193]. □

In general, x_M is only a seminorm, but Theorem 5.2.1, in particular, shows that x_M is nondegenerate for a *hyperbolic* 3-*manifold*, i.e., a 3-manifold whose interior admits a complete Riemannian metric of constant sectional curvature -1 and finite volume.

A properly embedded surface S is called *norm-minimizing* if $\chi_-(S) = x_M(\phi)$ for its dual $\phi \in H^1(M;\mathbb{Z})$. Every connected norm-minimizing surface S with negative Euler characteristic is incompressible since any compression of such a surface S along a simple closed curve not bounding any disc in S would reduce $\chi_-(S)$.

Example 5.2.3 The Thurston norm is a generalization of the knot genus: The *genus* $g(K)$ of a knot K in S^3 is the minimum genus of Seifert surfaces of K. Every norm-minimizing surface in the complement X_K of K dual to a generator $\psi \in H^1(X_K; \mathbb{Z})$ corresponds to a minimal genus Seifert surface of K, and we have

$$x_{X_K}(\phi) = 2g(K) - 1$$

for a nontrivial knot K. □

Example 5.2.4 Let M be a 3-manifold fibering over a circle with a fiber surface S. Then every incompressible surface in M representing the same homology class in $H_2(M, \partial M; \mathbb{Z})$ as S is isotopic to S.

We give a sketch of the proof as in [42, Lemma 5.1]. First such an incompressible surface S' in M lifts homeomorphically to an incompressible surface $\overline{S}'$ in the infinite cyclic covering $S \times \mathbb{R}$ of M corresponding with the fibration. Then the inclusion-induced homomorphism $\pi_1 \overline{S}' \to \pi_1(S \times \mathbb{R})$ is an isomorphism. If it would be not surjective, then an argument with van Kampen's theorem would imply that $\pi_1(S \times \mathbb{R})$ is not finitely generated. We thus see that S' is isotopic to S.

An immediate consequence is that S is norm-minimizing. We will see a more general result in Theorem 5.2.14 for codimension-1 foliations on M. □

5.2.2 Norm Balls and Fibrations Over a Circle

We next discuss the structure of the unit ball of the Thurston norm and the distribution of the cohomology classes represented by fibrations over a circle.

The *Thurston norm ball* of a 3-manifold M, denoted by B_M, is the unit ball of x_M:

$$B_M = \{\phi \in H^1(M; \mathbb{R}) \ ; \ x_M(\phi) \leq 1\}.$$

A seminorm determines its unit ball and vice versa. We set

$$\mathrm{Ann}(x_M^{-1}(\{0\})) = \{\alpha \in H_1(M; \mathbb{R}) \ ; \ \langle \phi, \alpha \rangle = 0 \text{ for all } \phi \in x_M^{-1}(\{0\})\},$$

where $\langle \cdot, \cdot \rangle$ is the Kronecker pairing. Note that if x_M is nondegenerate, then $\mathrm{Ann}(x_M^{-1}(\{0\})) = H_1(M; \mathbb{R})$. The *dual Thurston norm* x_M^* on $\mathrm{Ann}(x_M^{-1}(\{0\}))$ is defined by

$$x_M^*(\alpha) = \sup\{\langle \phi, \alpha \rangle \ ; \ \phi \in B_M\}$$

for $\alpha \in \mathrm{Ann}(x_M^{-1}(\{0\}))$. Theorem 5.2.1 implies that x_M^* is a norm on $\mathrm{Ann}(x_M^{-1}(\{0\}))$. The *dual Thurston norm ball* of M, denoted by B_M^*, is the unit ball of x_M^*:

$$\begin{aligned} B_M^* &= \{\alpha \in \mathrm{Ann}(x_M^{-1}(\{0\}) \; ; \; x_M^*(\alpha) \leq 1\} \\ &= \{\alpha \in \mathrm{Ann}(x_M^{-1}(\{0\}) \; ; \; \langle \phi, \alpha \rangle \leq 1 \text{ for all } \phi \in B_M\}. \end{aligned}$$

The unit ball of a seminorm, a priori, may be an arbitrary convex body symmetric in origin, but Thurston [202, Theorem 2] showed that the structure of B_M is more restrictive.

Theorem 5.2.5 ([202]) *The dual Thurston norm ball B_M^* of a 3-manifold M is a convex polytope in $H_1(M;\mathbb{R})$ with finitely many vertices $\pm\alpha_1, \ldots, \pm\alpha_k \in \mathrm{Ann}(x_M^{-1}(\{0\})) \cap H_1(M;\mathbb{Z})$, and we have*

$$B_M = \{\phi \in H^1(M;\mathbb{R}) \; ; \; |\langle \phi, \alpha_i \rangle| \leq 1 \, \textit{for } 1 \leq i \leq k\}.$$

Theorem 5.2.5 is a formal consequence of the fact that x_M is $\mathbb{Z}$-valued on the integral lattice $H^1(M;\mathbb{Z})$.

Corollary 5.2.6 ([202]) *The Thurston norm ball B_M of a 3-manifold M is a (possibly noncompact) convex polyhedron in $H^1(M;\mathbb{R})$ with finitely many vertices in $H^1(M;\mathbb{Q})$.* □

Remark 5.2.7 In this chapter a *convex polyhedron* in a real affine linear space refers to a closed convex subset such that every point on the boundary lies in only finitely many maximal convex subsets of the boundary. A *convex polytope* refers to a compact convex polyhedron. □

A cohomology class $\phi \in H^1(M;\mathbb{Z})$ is called *fibered* if M fibers over a circle such that the fibers are dual to ϕ. Since integration of a nonsingular closed 1-form on M with integer periods defines a fibration over a circle, $\phi \in H^1(M;\mathbb{Z})$ is fibered if and only if ϕ is represented by a nonsingular closed 1-form on M.

An observation is that since every nonsingular closed 1-form on M remains nonsingular after sufficiently small perturbation, the subset of cohomology classes of nonsingular closed 1-forms is open in $H^1(M;\mathbb{R})$. Also, a nonsingular closed 1-form on M defines a codimension-1 foliation on M, which we will discuss in Sect. 5.2.3. Based on the study of general position of incompressible surfaces with respect to codimension-1 foliations [202, Theorem 4], Thurston [202, Theorem 5] described the distribution of fibered classes in terms of B_M as follows:

Theorem 5.2.8 ([202]) *Let M be a 3-manifold fibering over a circle with fiber of negative Euler characteristic. There are some top-dimensional faces of B_M such that $\phi \in H^1(M;\mathbb{Z})$ is fibered if and only if ϕ lies in the interior of the cone on one of the faces.* □

Remark 5.2.9 For a 3-manifold M fibering over a circle with fiber of nonnegative Euler characteristic, x_M vanishes on $H^1(M; \mathbb{R})$. □

Such top-dimensional faces of B_M as in Theorem 5.2.8 are called *fibered faces* of B_M.

A 3-manifold M is *atoroidal* if M contains no incompressible torus.

Corollary 5.2.10 ([202]) *Let M be an atoroidal 3-manifold with $b_1(M) > 1$. Then there exists an incompressible surface which is not the fiber of a fibration over a circle.* □

Remark 5.2.11 In the proof of Thurston's hyperbolization theorem [105, 170, 171] Corollary 5.2.10 played a significant role to reduce the exceptional (semi)fibered case to the generic case. □

For a 3-manifold M with $b_1(M) = 1$ its norm balls B_M and B_M^* are closed intervals centered at origins, possibly consisting only of origins. The following examples together with one for the complement of the 3-link chain in S^3 are given in [202, Examples 1, 2, 3]. See also [202, Section 4] for a large variety of shapes for (dual) Thurston norm balls.

Example 5.2.12 Let L be the Whitehead link. Let $\mu_1, \mu_2 \in H_1(X_L; \mathbb{R})$ be a basis represented by meridians of the two components of L, and $\mu_1^*, \mu_2^* \in H^1(X_L; \mathbb{R})$ its dual basis. Then B_{X_L} is the diamond with vertices $\pm\mu_1^*, \pm\mu_2^*$ and $B_{X_L}^*$ is the square with vertices $\pm\mu_1 \pm \mu_2$. All the 2-dimensional faces of B_{X_L} are fibered faces. □

Example 5.2.13 Let L be the Borromean rings. Let $\mu_1, \mu_2, \mu_3 \in H_1(X_L; \mathbb{R})$ be a basis represented by meridians of the three components of L, and $\mu_1^*, \mu_2^*, \mu_3^* \in H^1(X_L; \mathbb{R})$ its dual basis. Then B_{X_L} is the octahedron with vertices $\pm\mu_1^*, \pm\mu_2^*, \pm\mu_3^*$ and $B_{X_L}^*$ is the cube with vertices $\pm\mu_1 \pm \mu_2 \pm \mu_3$. All the 3-dimensional faces of B_{X_L} are fibered faces. □

5.2.3 Norm-Minimizing Surfaces and Codimension-1 Foliations

Here we describe the correspondence between norm-minimizing surfaces and certain codimension-1 foliations.

A *codimension-1 foliation* $\mathcal{F}$ on a 3-manifold M is a decomposition of M into possibly noncompact immersed 2-dimensional submanifolds called leaves such that M is covered by a collection of charts of the form $\mathbb{R}^2 \times \mathbb{R}$ where the leaves pass through a given chart in slices of the form $\mathbb{R}^2 \times \{z\}$ for $z \in \mathbb{R}$. A codimension-1 foliation $\mathcal{F}$ on M is *transversely oriented* if some vector field on M transverse to the leaves of $\mathcal{F}$ is fixed.

A Reeb component is the foliation on the solid torus $D^2 \times S^1$ described as follows: Consider the decomposition $\mathbb{R}^2 \times [0, \infty)$ into planes $\mathbb{R}^2 \times \{z\}$ for $z \in [0, \infty)$ and the action $(x, y, z) \to 2(x, y, z)$ on $\mathbb{R}^2 \times [0, \infty)$ by $\mathbb{Z}$. The quotient

$(\mathbb{R}^2 \times [0, \infty) \setminus \{(0, 0, 0)\})/\mathbb{Z}$ is a solid torus, and the induced foliation on it is a Reeb foliation.

As a generalization of the fact that every fiber of a fibration over a circle is norm-minimizing, Thurston [202, Corollary 2] proved the following theorem. A 3-manifold is *irreducible* if every embedded 2-sphere in M bounds an embedded 3-ball in M.

Theorem 5.2.14 ([202]) *Let M be an irreducible 3-manifold with empty or toroidal boundary and $\mathcal{F}$ a codimension-1 transversely oriented foliation on M such that $\mathcal{F}$ has no Reeb components, and each component of ∂M is either transverse to $\mathcal{F}$ or is a leaf of $\mathcal{F}$. Then every compact leaf of $\mathcal{F}$ is norm-minimizing.* □

For such a codimension-1 foliation $\mathcal{F}$ as in Theorem 5.2.14 we can consider the Euler class of the bundle of planes tangent to the leaves. The key ingredient of the proof is that the dual Thurston norm of the Poincaré dual of the Euler class is less than or equal to 1.

A codimension-1 transversely oriented foliation $\mathcal{F}$ on M is *taut* if there exists a closed curve in M transversally intersecting each leaf of $\mathcal{F}$. Every taut foliation has no Reeb components [81]. Let $\mathcal{F}$ be a codimension-1 foliation on M. We say that a leaf L of $\mathcal{F}$ is of depth 0 if L is compact. Having defined depth $j \leq k$ leaves we say that L is of depth $k + 1$ if $\overline{L} \setminus L$ is a union of depth $j \leq k$ leaves and contains a leaf of depth k. We say that $\mathcal{F}$ is *of finite depth* if there exists an integer k such that the depth of every leaf of $\mathcal{F}$ is defined to be less than k.

Developing the theory of sutured manifolds, Gabai [74, Theorem 5.5] proved the following theorem, which can be seen as the converse of Theorem 5.2.14:

Theorem 5.2.15 ([74]) *Let M be an irreducible 3-manifold with empty or toroidal boundary and S a norm-minimizing surface in M representing a nontrivial class in $H_2(M, \partial M; \mathbb{Z})$. Then there exists a codimension-1 transversely oriented taut foliation $\mathcal{F}$ on M of finite depth such that $\mathcal{F}$ is transverse to ∂M, S is a leaf of $\mathcal{F}$ and $\mathcal{F}|_{\partial M}$ is a suspension of homeomorphisms of S^1.* □

Gabai's construction of such a foliation as in Theorem 5.2.15 used a so-called sutured manifold hierarchy [74, 193]: A *sutured manifold* (M, γ) is a 3-manifold M equipped with a decomposition of ∂M into two subsurfaces $R_\pm$ meeting along a possibly empty system γ of simple closed curves such that R_- and R_- are transversely oriented inwards and outwards respectively. Under the assumptions in Theorem 5.2.15 there exists a sequence of sutured manifolds $(M_0, \gamma_0), \ldots, (M_n, \gamma_n)$, where (M_0, γ_0) is obtained by decomposing M along S, (M_i, γ_i) is obtained by decomposing M_{i-1} along certain type of properly embedded surface, and (M_n, γ_n) is a collection of 3-balls with single simple closed curves. A codimension-1 transversely oriented taut foliation on M is constructed inductively on the hierarchy.

Together with Theorem 5.2.14, the construction provides an effective algorithm to compute the Thurston norm. Applying the techniques to knots and links, Gabai [75] gave tables of the genera of knots with 10 or fewer crossings and links with 9 or fewer crossings.

Remark 5.2.16 Based on the idea, Lackenby [130] showed that it is in NP to determine the Thurston norm of a given first cohomology class. Computations of the Thurston norm were described by Oertel [168] in terms of branched surfaces, and by Mosher [154–156] in terms of pseudo-Anosov flows. We will see another algorithm via normal surface theory in Sect. 5.7.1. □

As a corollary of Theorems 5.2.14, 5.2.15, Gabai [74, Corollary 6.13] proved the following, which has been conjectured by Thurston [202].

Corollary 5.2.17 *Let $p\colon \widetilde{M} \to M$ be an n-fold covering. Then*

$$x_{\widetilde{M}}(p^*(\phi)) = n x_M(\phi)$$

for $\phi \in H^1(M;\mathbb{R})$. □

It is worth pointing out here that for a finite covering $p\colon \widetilde{M} \to M$, $\phi \in H^1(M;\mathbb{Z})$ is fibered if and only if $p^*(\phi)$ is fibered, which is an immediate consequence of Stallings' fibration theorem.

In [76, Theorem 2] Gabai showed the following stronger result than Theorem 5.2.15 in the case of knots in S^3:

Theorem 5.2.18 ([76]) *Let K be a knot in S^3 and S a minimal genus Seifert surface of K. Then there exits a codimension-1 taut foliation $\mathcal{F}$ on $K(0)$ of finite depth such that the capped off surface S is a leaf of $\mathcal{F}$.* □

As a corollary Gabai [76, Corollary 5] proved the following:

Corollary 5.2.19 ([76]) *For a knot K in S^3, $K(0)$ is prime and the genus $g(K)$ is equal to the minimal genus of an embedded nonseparating surface in $K(0)$.* □

The *Property R conjecture* asserts that if $K(0)$ is homeomorphic to $S^2 \times S^1$, then K is the unknot. The *Poénaru conjecture* is stronger and asserts that if $K(0)$ is reducible, then K is the unknot. Corollary 5.2.19 gave the positive proofs of these conjectures.

Remark 5.2.20 Let M be an irreducible 3-manifold with toroidal boundary, not being a cable space and not homeomorphic to $T^2 \times [0,1]$. Generalizing a result of Sela [194], Baker and Taylor [7] showed that for all but finitely many slopes of ∂M, the Thurston norm of M equals that of the result of the Dehn filling along a slope plus the so-called winding norm. □

Another corollary [76, Corollary 6] of Theorem 5.2.18 is the following:

Corollary 5.2.21 ([76]) *A knot K in S^3 is fibered if and only if $K(0)$ is fibered.* □

Scharlemann [193] realized that much of Gabai's theory could work only in terms of sutured manifolds without any reference to foliations. See also [25] for results on a generalization of the Thurston norm for sutured manifolds.

5.2.4 Singular and Gromov Norms

Using Theorem 5.2.15, Gabai [74, Corollary 6.18] showed the equivalence of the Thurston norm, its singular one and the Gromov norm of a 3-manifold M.

The *singular Thurston norm* $x_{M,s}$ on $H^1(M;\mathbb{Z})$ is defined by

$$x_{M,s}(\phi) = \min\Big\{\frac{1}{n}\chi_-(S)\ ;\ f\colon (S,\partial S)\to (M,\partial M) \text{ is a proper map from}$$
$$\text{a surface } S \text{ such that } f_*([S,\partial S]) \text{ is dual to } n\phi\Big\},$$

for $\phi \in H^1(M;\mathbb{Z})$. It is straightforward to see that $x_{M,s}$ uniquely extends to $H^1(M;\mathbb{R})$ as a seminorm.

For a singular k-chain $\sum_i a_i\sigma_i \in C_k(M,\partial M;\mathbb{R})$ its norm is defined to be the sum $\sum_i |a_i|$ of its absolute values of the coefficients. The *Gromov norm* or l^1-*seminorm* $||c||_1$ of $c \in H_k(M,\partial M;\mathbb{R})$ is the induced seminorm [83]:

$$||c||_1 = \inf\left\{\sum_i |a_i|\ ;\ \sum_i a_i\sigma_i \text{ is a singular 2-cycle representing } c\right\}.$$

We also denote by $||\cdot||_1$ the seminorm on $H^*(M;\mathbb{R})$ induced by Poincaré duality.

The first equality in the following theorem has been conjectured by Thurston [202]. See also [186] for a combinatorial proof.

Theorem 5.2.22 ([74]) *The following equality holds on* $H^1(M;\mathbb{R})$*:*

$$x_M = x_{M,s} = \frac{1}{2}||\cdot||_1.$$

As a special case, Gabai [74, Corollary 6.23] proved the following generalization of Dehn's lemma for higher genus surfaces.

Corollary 5.2.23 ([74]) *Let* $f\colon S\to M$ *be a map from a surface with connected boundary such that* $f|_{\partial S}$ *is an embedding and* $f^{-1}(f(\partial S)) = \partial S$*. Then there exists an embedded surface* S' *in* M *such that* $\partial S' = \partial S$ *and the genus of* S' *is less than or equal to that of* S*.* □

In particular, Corollary 5.2.23 shows equality of the embedded and immersed genera of knots [74, Corollary 6.22]:

Corollary 5.2.24 ([74]) *The genus* $g(K)$ *of a knot* K *in* S^3 *is equal to the minimal genus of immersed surfaces* S *in* S^3 *bounding* K *which are nonsingular along* K*.* □

5.3 Alexander and Teichmüller Polynomials

We describe the lower bound on the Thurston norm by the Alexander polynomial, following McMullen [147]. This lower bound is then restated in terms of abelian Reidemeister torsion. We also discuss the Teichmüller polynomial associated with a fibered face of the Thurston norm ball, introduced by McMullen [146].

5.3.1 Alexander Polynomial

It is well known that the classical Alexander polynomial $\Delta_K(t) \in \mathbb{Z}[t, t^{-1}]$ of a knot K in S^3 gives a lower bound on the genus $g(K)$:

$$2g(K) \geq \deg \Delta_K(t),$$

where equality holds if K is a fibered knot. Following McMullen [147], we describe a generalization of this inequality on the Thurston norm and the Alexander polynomial of a general 3-manifold.

Let M be a 3-manifold with empty or toroidal boundary. We denote by $H_1(M)_f$ the free abelian group obtained by dividing $H_1(M; \mathbb{Z})$ by the torsion submodule. We denote by $\overline{M}$ the maximal free abelian cover of M, which is the cover of M associated with the canonical projection $\pi_1 M \to H_1(M)_f$. Since $H_1(M)_f$ acts on $\overline{M}$ by deck transformations, $H_1(\overline{M}; \mathbb{Z})$ is a finitely generated module over the group ring $\mathbb{Z}[H_1(M)_f]$. The *Alexander polynomial* $\Delta_M \in \mathbb{Z}[H_1(M)_f]$ of M is the order of $H_1(\overline{M}; \mathbb{Z})$ over $\mathbb{Z}[H_1(M)_f]$, which is well-defined up to multiplication by elements of $\pm H_1(M)_f$: In general, for a finitely generated module L over a noetherian UFD R and an exact sequence

$$R^l \xrightarrow{r} R^m \to L \to 0$$

with $l \geq m$, the *order* of L is the greatest common divisor of the m-minors of a representation matrix r, and is well-defined up to multiplication by units in R.

Example 5.3.1 For a knot K in S^3, Δ_{X_K} coincides with the classical Alexander polynomial $\Delta_K(t) \in \mathbb{Z}[t, t^{-1}]$ under the identification of $H_1(X_K; \mathbb{Z})$ with the infinite cyclic group generated by t. □

McMullen [147, Theorem 1.1] introduced the *Alexander norm* $||\cdot||_A$ on $H^1(M; \mathbb{R})$ and showed an inequality between the Thurston and Alexander norms as follows: We write $\Delta_M = \sum_{h \in H_1(M)_f} a_h h$ for $a_h \in \mathbb{Z}$. If $\Delta_M = 0$, then we define $||\cdot||_A = 0$. Otherwise, we define

$$||\phi||_A = \max\{\langle \phi, h - h' \rangle \ ; \ h, h' \in H_1(M)_f \text{ such that } a_h a_{h'} \neq 0\}$$

for $\phi \in H^1(M; \mathbb{R})$. It is clear that $||\cdot||_A$ is a seminorm on $H^1(M; \mathbb{R})$.

Theorem 5.3.2 ([147]) *Let M be a 3-manifold with empty or toroidal boundary. Then*

$$x_M(\phi) \geq ||\phi||_A - \begin{cases} 1 + b_3(M) & \text{if } H^1(M;\mathbb{Z}) \text{ is generated by } \phi, \\ 0 & \text{if } b_1(M) > 1, \end{cases}$$

for $\phi \in H^1(M;\mathbb{Z})$. Furthermore, equality holds if ϕ is fibered with $M \neq S^1 \times S^2$ and $M \neq S^1 \times D^2$. □

Remark 5.3.3 As Dunfield [39] showed, there are examples of 3-manifolds M fibering over a circle with $b_1(M) > 1$ such that x_M and $||\cdot||_A$ do not agree. □

It is known that $2g(K) = \deg \Delta_K(t)$ for all knots up to 10 crossings or less (see for example [75]). McMullen [147, Theorem 7.1] showed that the Thurston and Alexander norms agree for all the tabulated links with 9 or fewer crossings in [191] except 9^3_{21}, and possibly 9^2_{41}, 9^2_{50} and 9^3_{15}.

5.3.2 Abelian Torsion

We discuss a corresponding result to Theorem 5.3.2 in terms of abelian Reidemeister torsion. We will see the precise definition of Reidemeister torsion in Sect. 5.6.1.

Let M be a 3-manifold with empty or toroidal boundary with a CW-complex structure. We denote by $\mathbb{Q}(H_1(M)_f)$ the quotient field of $\mathbb{Z}[H_1(M)_f]$. The *abelian torsion* or *Milnor torsion* $\tau(M) \in \mathbb{Q}(H_1(M)_f)$ of M is the Reidemeister torsion associated with the canonical projection $\pi_1 M \to H_1(M)_f$, which is the algebraic torsion of the twisted chain complex $C_*(\overline{M}) \otimes_{\mathbb{Z}[H_1(M)_f]} \mathbb{Q}(H_1(M)_f)$ of the CW-complex $\overline{M}$. The topological invariant $\tau(M)$ is well-defined up to multiplication by elements of $\pm H_1(M)_f$, and is known to be symmetric, i.e., $\tau(M)$ is invariant up to multiplication by elements of $\pm H_1(M)_f$ under the involution on $\mathbb{Q}(H_1(M)_f)$ reversing the elements of $H_1(M)_f$.

Example 5.3.4 For a knot K in S^3,

$$\tau(X_K) = \frac{\Delta_K(t)}{t-1}$$

under the identification of $H_1(X_K;\mathbb{Z})$ with the infinite cyclic group generated by t. □

Turaev [205, 211] showed that $\tau(M)$ determines the Alexander polynomial Δ_M, and vise versa:

Theorem 5.3.5 ([205, 211]) *Let M be a 3-manifold with empty or toroidal boundary. If $H_1(M)_f$ is an infinite cyclic group generated by t, then*

$$\tau(M) = \begin{cases} \frac{\Delta_M}{(t-1)^2} & \text{if } \partial M = \emptyset, \\ \frac{\Delta_M}{t-1} & \text{if } \partial M \neq \emptyset. \end{cases}$$

If $b_1(M) > 1$, then

$$\tau(M) = \Delta_M.$$

A cohomology class $\phi \in H^1(M; \mathbb{Z})$ induces a ring homomorphism $\mathbb{Z}[H_1(M)_f] \to \mathbb{Z}[t, t^{-1}]$ by sending $h \in H_1(M)_f$ to $t^{\langle \phi, h \rangle}$. We define $\tau_\phi(M) \in \mathbb{Q}(t)$ to be the reduction of $\tau(M)$ by the induced homomorphism, which is the algebraic torsion of $C_*(\overline{M}) \otimes_{\mathbb{Z}[H_1(M)_f]} \mathbb{Q}(t)$.

Theorem 5.3.2 is restated in terms of τ_ϕ as follows. We define

$$\deg \left(a_l t^l + a_{l+1} t^{l+1} + \cdots + a_m t^m \right) = m - l$$

for $l, m \in \mathbb{Z}$ with $l < m$ and $a_i \in \mathbb{Z}$ with $a_l a_m \neq 0$, and further define

$$\deg \frac{p(t)}{q(t)} = \deg p(t) - \deg q(t)$$

for $p(t), q(t) \in \mathbb{Z}[t, t^{-1}] \setminus \{0\}$.

Theorem 5.3.6 *Let M be a 3-manifold with empty or toroidal boundary. Then*

$$x_M(\phi) \geq \deg \tau_\phi(M)$$

for $\phi \in H^1(M; \mathbb{Z})$. Furthermore, equality holds if ϕ is fibered with $M \neq S^1 \times S^2$ and $M \neq S^1 \times D^2$. □

It is well known that $\Delta_K(t)$ is monic for a fibered knot K in S^3. More generally, $\tau_\phi(M)$ is represented by a monic polynomial divided by $(t-1)^{1+b_3(M)}$ for a fibered class $\phi \in H^1(M; \mathbb{Z})$. We will discuss more the property in Remark 5.6.5.

In analogy with the Thurston norm, Turaev [210, 212] introduced a seminorm on $H^1(X; \mathbb{R})$ for a finite 2-dimensional complex X, and numerical functions on $H_2(M; \mathbb{Q}/\mathbb{Z})$ and on the torsion subgroup of $H_1(M; \mathbb{Z})$ for a 3-manifold M. Turaev showed that the Alexander polynomial and abelian Reidemeister torsion give lower bounds also on these functions. See [62, 166] for further studies of such analogues of the Thurston norm.

5.3.3 Teichmüller Polynomial

Here we give a brief exposition of the Teichmüller polynomial introduced by McMullen [146].

A (codimension-1) lamination $\mathcal{L}$ on a 3-manifold M is a codimension-1 foliation on a closed subset of M. A lamination $\mathcal{L}$ on M is *transversely orientable* if there is a nonsingular vector field on a neighborhood of the underlying closed subset of $\mathcal{L}$ in M transverse to the leaves. A *geodesic lamination* λ on a hyperbolic surface S is a decomposition of a closed subset of S into simple geodesics.

Let M be a hyperbolic 3-manifold having a fibered face F of B_M. Let $\phi \in H^1(M;\mathbb{Z})$ be a fibered class in the cone on F, and $\psi\colon S \to S$ a pseudo-Anosov monodromy of a fibration representing ϕ. Then ψ has an expanding invariant geodesic lamination λ on S. Let $\mathcal{L}$ be the lamination on M obtained as the mapping torus of $\psi|_\lambda$. Based on results by Fried [46], McMullen [146, Corollary 3.2] showed that the isotopy class of $\mathcal{L}$ depends only on F.

We denote by $\overline{\mathcal{L}}$ the preimage of $\mathcal{L}$ by the maximal free abelian covering $\overline{M} \to M$. A *transeversal* for $\overline{\mathcal{L}}$ is a compact totally disconnected subset of $\overline{\mathcal{L}}$ such that there is an open neighborhood U of T with a homeomorphism $(U, T) \to (T \times \mathbb{R}^2, T \times \{0\})$. Note that the free abelian group $H_1(M)_f$ acts on the set of transversals for $\mathcal{L}$. We define $T(\overline{\mathcal{L}})$ to be the abelian group generated by all transversals $[T]$ for $\overline{\mathcal{L}}$ modulo the following relations:

1. $[T] = [T'] + [T'']$, if T is a disjoint union of T' and T'',
2. $[T] = [T']$, if there is an open neighborhood U of $T \cup T'$ with homeomorphisms $(U, T) \to (T \times \mathbb{R}^2, T \times \{0\})$, $(U, T') \to (T' \times \mathbb{R}^2, T' \times \{0\})$.

A consequence of the compactness of $\mathcal{L}$ is that $T(\overline{\mathcal{L}})$ is a finitely generated $\mathbb{Z}[H_1(M)_f]$-module.

Now the *Teichmüller polynomial* $\Theta_F \in \mathbb{Z}[H_1(M)_f]$ of F is defined to be the order of $T(\overline{\mathcal{L}})$ over $\mathbb{Z}[H_1(M)_f]$, which is well-defined up to multiplication by elements of $\pm H_1(M)_f$. McMullen showed that Θ_F is monic and symmetric.

McMullen [146, Theorem 6.1] introduced the *Teichmüller norm* $||\cdot||_{\Theta_F}$ on $H^1(M;\mathbb{R})$ and showed its relation with the Thurston norm as follows: We write $\Theta_F = \sum_{h \in H_1(M)_f} a_h h$ for $a_h \in \mathbb{Z}$ and define

$$||\phi||_{\Theta_F} = \max\{\langle \phi, h - h' \rangle \ ; \ h, h' \in H_1(M)_f \text{ such that } a_h a_{h'} \neq 0\}$$

for $\phi \in H^1(M;\mathbb{R})$. It is clear that $||\cdot||_{\Theta_F}$ is a seminorm on $H^1(M;\mathbb{R})$.

Theorem 5.3.7 ([146]) *Let F be a fibered face of B_M of a hyperbolic 3-manifold M. Then there exists a face D of the unit ball of $||\cdot||_{\Theta_F}$ such that the cones on F and D coincides.* □

Together with a computational formula of Θ_F in terms of train tracks on fibers [146, Theorem 3.6], Theorem 5.3.7 provides an effective algorithm to determine a fibered face of B_M for a hyperbolic 3-manifold M from a single fiber and the monodromy on it.

Using Theorem 5.3.2, McMullen [146, Theorem 7.1] proved the following theorem:

Theorem 5.3.8 ([146]) *Let F be a fibered face of B_M of a hyperbolic 3-manifold M with $b_1(M) > 1$. Then there exists a unique face A of the unit ball of the Alexander norm containing F. Furthermore, if the lamination $\mathcal{L}$ associated with F is transversely orientable, then $F = A$ and Δ_M divides Θ_F.* □

McMullen [146] also showed that for a fibered class $\phi \in H^1(M; \mathbb{Z})$ lying in the cone on a fibered face F, the dilatation $\lambda(\phi)$ of its monodromy is the largest root of the polynomial equation $\sum_{h \in H} a_h t^{\phi(h)} = 0$ obtained by evaluating Θ_F by ϕ, and that the function $\frac{1}{\log \lambda(\phi)}$ extends to the cone on F as a real-analytic function which is strictly concave, extending results in [47, 143]. See also [199].

Dowdall, Kapovich and Leininger [35, 36] introduced analogues of the Teichmüller polynomial and proved analogous results for free-by-cyclic groups. In their work a hyperbolic 3-manifold fibering over a circle and its fibered face of the Thurston norm ball are replaced by a free-by-cyclic group and a component of its Bieri–Neumann–Strebel invariant [11].

5.4 Seiberg–Witten Invariant

Here we are concerned with adjunction inequalities, which give relationships between the Seiberg–Witten invariant of a 4-manifold and the complexity of embedded surfaces in the manifold, and between the Seiberg–Witten invariant of a 3-manifold and its Thurston norm. As a related topic we also discuss the harmonic norm on the cohomology group associated with a Riemannian metric.

5.4.1 Seiberg–Witten Theory

We briefly review Seiberg–Witten theory [216] in the case of a closed smooth 4-manifold with $b_2^+(N) > 1$. (Here $b_2^+(N)$ is the dimension of a maximal positive-definite subspace $H_+^2(N; \mathbb{R})$ of the intersection pairing on $H^2(N; \mathbb{R})$.) For the details we refer the reader to the expositions [94, 120, 133, 157].

Recall that the Lie group $\mathrm{Spin}^c(n) = \mathrm{Spin}(n) \times_{\pm 1} U(1)$ is a central extension of $SO(n)$ by $U(1)$. A Spin^c-structure on a Riemannian n-manifold X is a lifting of the principal $SO(n)$-frame bundle on X to a principal $\mathrm{Spin}^c(n)$-bundle. We denote by $\mathrm{Spin}^c(X)$ the set of equivalence classes of Spin^c-structures on X. The set $\mathrm{Spin}^c(X)$ has a free and transitive action by $H^2(X; \mathbb{Z})$, and we write $\mathfrak{s} + c$ for the image of

$\mathfrak{s}$ by $c \in H^2(X;\mathbb{Z})$. The first Chern class of the principal $U(1)$-bundle associated with a Spin^c-structure on X defines the map $c_1\colon \mathrm{Spin}^c(X) \to H^2(X;\mathbb{Z})$.

Let N be a closed Riemannian 4-manifold with a metric g, and $\widetilde{P}$ a spin^c structure on N. We denote by L the determinant line bundle of $\widetilde{P}$ and by $S^{\pm}$ the two complex spin bundles associated with $\widetilde{P}$. For a connection on L we have Dirac operators $D_A\colon \Gamma(S^{\pm}) \to \Gamma(S^{\mp})$ on the set of sections of $S^{\pm}$, defined using Levi–Civitá connection on the frame bundle on X. The *Seiberg–Witten monopole equations* associated with $\widetilde{P}$ are the following pair of nonlinear elliptic equations for unitary connections A on L and sections ψ of S^+:

$$F_A^+ = \psi \otimes \psi^* - \frac{|\psi|^2}{2} Id,$$
$$D_A(\psi) = 0,$$

where we identify S^+ with its dual via an anti-complex isomorphism and ψ^* is its image of ψ. We denote by $\mathcal{M}_N(\widetilde{P})$ the quotient of the space of gauge-equivalence classes of solutions to the equations. The moduli space $\mathcal{M}_N(\widetilde{P})$ is known to be compact [216]. A class $c \in H^2(N;\mathbb{Z})$ is called a *Seiberg–Witten monopole class* if $c = c_1(\mathfrak{s})$ for some $\mathfrak{s} \in \mathrm{Spin}^c(M)$ representing $\widetilde{P}$ with nonempty $\mathcal{M}(\widetilde{P}, g)$.

If $b_2^+(N) > 1$, then the *Seiberg–Witten invariant* $SW_N\colon \mathrm{Spin}^c(N) \to \mathbb{Z}$ is defined as follows: Let $\widetilde{P}$ be a Spin^c-structure on N representing $\mathfrak{s} \in \mathrm{Spin}^c(N)$. We denote by $C_N(\widetilde{P})$ the space of gauge-equivalence classes of pairs (A, ψ) with $\psi \neq 0$, which is a classifying space of the group $(S^1)^N$. There is a universal S^1-bundle over $C_N(\widetilde{P})$ whose Chern class μ generates $H^2(C_N(\widetilde{P});\mathbb{Z})$. After a perturbation of $\mathcal{M}_N(\widetilde{P})$ by an addition of $i\eta$ to the right hand side of the first Seiberg–Witten monopole equation for a generic (real) self-dual 2-form η on N, the resulting moduli space $\mathcal{M}$ is known to become a smooth submanifold in $C_N(\widetilde{P})$ of dimension

$$d(\mathfrak{s}) = \frac{\langle c_1(\mathfrak{s})^2, [N]\rangle - (2\chi(N) + 3\sigma(N))}{4},$$

where $\sigma(N)$ is the signature of N. Moreover, choosing an orientation of the real vector space $H^2_+(N;\mathbb{R}) \oplus H^1(N;\mathbb{R})$ gives an orientation of $\mathcal{M}$. If $d(\mathfrak{s})$ is odd, then we define $SW_N(\mathfrak{s}) = 0$, and otherwise we define

$$SW_N(\mathfrak{s}) = \langle \mu^{d(\mathfrak{s})}, [\mathcal{M}]\rangle.$$

This is an invariant of $\mathfrak{s}$, which is independent of the choice of the Riemannian metric of N and the perturbation term η [216]. The invariant SW_N takes nonzero value only on finitely many Spin^c-structures, and changing the orientation of $H^2_+(N;\mathbb{R}) \oplus H^1(N;\mathbb{R})$ reverses its sign. A class $c \in H^2(N;\mathbb{Z})$ is called a *Seiberg–Witten basic class* if $c = c_1(\mathfrak{s})$ for some $\mathfrak{s} \in \mathrm{Spin}^c(N)$ with $SW_N(\mathfrak{s}) \neq 0$. Note that every Seiberg–Witten basic class is a Seiberg–Witten monopole class.

5.4.2 Seiberg–Witten Invariant of a 3-Manifold

The *Seiberg–Witten invariant* $SW_M\colon \mathrm{Spin}^c(M) \to \mathbb{Z}$ of a closed 3-manifold M with $b_1(M) > 1$ can be defined by

$$SW_M(\mathfrak{s}) = SW_{M\times S^1}(\pi^*\mathfrak{s})$$

for $\mathfrak{s} \in \mathrm{Spin}^c(M)$, where $\pi^*\mathfrak{s} \in \mathrm{Spin}^c(M \times S^1)$ is the pullback of $\mathfrak{s}$ by the projection $\pi\colon M \times S^1 \to M$. Note that all solutions to the Seiberg–Witten monopole equations on $M \times S^1$ are known to be S^1-invariant [169]. One can also define SW_M directly in terms of Seiberg–Witten monopole equations on M as in Sect. 5.4.1. As in the case of a 4-manifold, a class $c \in H^2(M;\mathbb{Z})$ is called a *Seiberg–Witten monopole class* if $c = c_1(\mathfrak{s})$ for some $\mathfrak{s} \in \mathrm{Spin}^c(M)$ representing a Spin^c-structure $\widetilde{P}$ on M with nonempty $\mathcal{M}_{M\times S^1}(\pi^*\widetilde{P})$ for the pullback $\pi^*\widetilde{P}$ on $M \times S^1$. Also, a class $c \in H^2(M;\mathbb{Z})$ is called a *Seiberg–Witten basic class* if $c = c_1(\mathfrak{s})$ for some $\mathfrak{s} \in \mathrm{Spin}^c(M)$ with $SW_M(\mathfrak{s}) \neq 0$.

For a closed 3-manifold M, Turaev [206, 207] introduced a refinement of the abelian torsion $\tau(M)$ as an integer-valued function T_M on $\mathrm{Spin}^c(M)$ (or on the set of so-called Euler structures on M), called *Turaev's torsion function* of M. When $b_1(M) > 1$, $\tau(M)$ is represented by an element of $\mathbb{Z}[H_1(M)_f]$ (Theorem 5.3.5), and T_M satisfies

$$\tau(M) = \sum_{h\in H_1(M;\mathbb{Z})} T_M(\mathfrak{s} - PD(h))[h]$$

for $\mathfrak{s} \in \mathrm{Spin}^c(M)$, where $PD(h) \in H^2(M;\mathbb{Z})$ is the Poincaré dual of h. Similarly, when $H_1(M)_f$ is an infinite cyclic group generated by t, $\tau(M)$ is represented by an element of the Novikov ring $\mathbb{Z}((t)) = \mathbb{Z}[[t]][t^{-1}]$ (Theorem 5.3.5), and T_M satisfies

$$\tau(M) = \sum_{h\in H_1(M;\mathbb{Z})} T_M(\mathfrak{s} - PD(h))[h] \in \mathbb{Z}((t))$$

for $\mathfrak{s} \in \mathrm{Spin}^c(M)$.

In terms of T_M, Turaev [208, Theorem 1] refined the equivalence of the Seiberg–Witten invariant and the abelian torsion shown by Meng and Taubes [148]:

Theorem 5.4.1 ([148, 208]) *For a closed* 3*-manifold M with $b_1(M) > 1$, the Seiberg–Witten invariant and Turaev's torsion function of M coincides up to sign:*

$$SW_M = \pm T_M.$$

Theorem 5.4.1 similarly extends to the case $b_1(M) = 1$ [148, 208]. See [162] for the case $b_1(M) = 0$.

The following is the *adjunction inequality* for 3-manifolds. See [6, 114] for the details. We will discuss more on adjunction inequalities in Sect. 5.4.3.

Theorem 5.4.2 ([6, 114]) *Let M be a closed irreducible 3-manifold with $b_1(M) > 1$ and $c \in H^2(M;\mathbb{Z})$ a Seiberg–Witten basic class. Then*

$$x_M(\phi) \geq |\langle c \cup \phi, [M]\rangle|$$

for $\phi \in H^1(M;\mathbb{R})$. □

Kronheimer and Mrowka [119, Theorem 1] showed that the Thurston norm is determined by the Seiberg–Witten monopole classes:

Theorem 5.4.3 ([119]) *Let M be a closed irreducible 3-manifold with $b_1(M) > 1$. Then*

$$x_M(\phi) = \max\{|\langle c \cup \phi, [M]\rangle| \,;\ c \in H^2(M;\mathbb{Z}) \text{ is a Seiberg–Witten monopole class}\}$$

for $\phi \in H^1(M;\mathbb{R})$. □

Corollary 5.4.4 ([119]) *Let M be a closed irreducible 3-manifold with $b_1(M) > 1$. Then the convex hull of the Seiberg–Witten monopole classes in $H^2(M;\mathbb{R})$ is equal to B_M^*.* □

As described by Kronheimer [114] and Vidussi [214], Theorems 5.4.1, 5.4.2, 5.4.3 deduce Theorem 5.3.2 for closed irreducible 3-manifolds M with $b_1(M) > 1$.

5.4.3 Complexity of Surfaces in a 4-Manifold

Adjunction inequalities give relationships between the Seiberg–Witten invariants of a 4-manifold and the genus of embedded surfaces in the manifold. The terminology arises from the adjunction formula for a smooth algebraic curve C in an algebraic surface X:

$$\chi_-(C) = C \cdot C - \langle c_1(X), [C]\rangle.$$

The genus of the algebraic curve of degree d in $\mathbb{C}P^2$ is given by $\frac{(d-1)(d-2)}{2}$. A conjecture attributed to Thom states that the genus of the algebraic curve is minimal among smoothly embedded surfaces in $\mathbb{C}P^2$ representing $d[\mathbb{C}P^1] \in H_2(\mathbb{C}P^2;\mathbb{Z})$. With the advance of the Seiberg–Witten monopole equations, Kronheimer and Mrowka [116, 117] and Morgan, Szabó and Taubes [158] proved the Thom conjecture for holomorphic curves in a general Kähler surface with nonnegative intersection. Later, Ozsváth and Szabó [172, Theorem 1.1 and Corollary 1.2] proved the symplectic Thom conjecture in its complete generality:

Theorem 5.4.5 ([172]) *The genus of an embedded symplectic surface in a closed symplectic 4-manifold is minimal among smoothly embedded surfaces representing the same homology class.* □

The theorem for Kähler surfaces follows as a special case of Theorem 5.4.5:

Corollary 5.4.6 *The genus of an embedded holomorphic curve in a Kähler surface is minimal among smoothly embedded surfaces representing the same homology class.* □

The following are adjunction inequalities shown by Morgan, Szabó and Taubes [158, Proposition 4.2] and Ozsváth and Szabó [172, Corollary 1.7]:

Theorem 5.4.7 ([158]) *Let N be a smooth closed 4-manifold with $b_2^+(M) > 1$ and $c \in H^2(N; \mathbb{Z})$ a Seiberg–Witten basic class. Then for a smoothly embedded surface Σ in N with nonpositive Euler characteristic and $[\Sigma] \cdot [\Sigma] \geq 0$, we have*

$$\chi_-(\Sigma) \geq [\Sigma] \cdot [\Sigma] + \langle c, [\Sigma] \rangle.$$

Theorem 5.4.8 ([172]) *Let N be a smooth closed 4-manifold with $b_2^+(M) > 1$ and $c \in H^2(N; \mathbb{Z})$ a Seiberg–Witten basic class. Suppose that $d(\mathfrak{s}) = 0$ for any $\mathfrak{s} \in \mathrm{Spin}^c(N)$ associated with a basic class in $H^2(N; \mathbb{Z})$. Then for a smoothly embedded surface in N with nonpositive Euler characteristic and $[\Sigma] \cdot [\Sigma] < 0$, we have*

$$\chi_-(\Sigma) \geq [\Sigma] \cdot [\Sigma] + |\langle c, [\Sigma] \rangle|.$$

See [173] for further refinements of Theorems 5.4.7, 5.4.8.

There is also an adjunction inequality by Fintushel and Stern [43] for embedded spheres:

Theorem 5.4.9 ([43]) *Let N be a smooth closed 4-manifold with $b_2^+(M) > 1$. Suppose that there exists a Seiberg–Witten basic class. Then there exist no smoothly embedded spheres Σ such that $\Sigma \cdot \Sigma \geq 0$ and $[\Sigma] \neq 0$.* □

See [8, 9] for a refinement of Theorem 5.4.9 in terms of the so-called Bauer-Furuta invariants. See also [113, 198] for adjunction-type inequalities for families of embedded surfaces.

Now we describe results on the relationship between complexity of embedded surfaces in circle bundles over a 3-manifold and its Thurston norm.

Let N be a smooth closed 4-manifold. We define a function $x_N : H_2(N; \mathbb{Z}) \to \mathbb{Z}$ by

$$x_N(\alpha) = \min\{\chi_-(\Sigma) \; ; \; \Sigma \text{ is an embedded surface representing } \alpha\}$$

for $\alpha \in H_2(N; \mathbb{Z})$.

Using Agol's virtual fibering theorem [1, 2] (see Theorem 5.9.8) and considering the Seiberg–Witten invariants of finite covers, Friedl and Vidussi [70, Theorem 1.1], and Nagel [159, Theorem 5.6] showed the following theorem:

Theorem 5.4.10 ([70, 159]) *Let M be a closed irreducible 3-manifold which is not a Seifert fibered space and not covered by a torus bundle, and let $p : N \to M$ be an* *oriented circle bundle. Then*

$$x_N(\alpha) \geq |\alpha \cdot \alpha| + x_M(p_*\alpha)$$

for $\alpha \in H_2(N; \mathbb{Z})$. □

Remark 5.4.11 Kronheimer [115] proved the same inequality as in Theorem 5.4.10 for the case $N = M \times S^1$ such that M is a closed irreducible 3-manifold whose Thurston norm does not identically vanish. □

Let M be a closed 3-manifold. We set Ξ_M to be the inverse image by the canonical map $H^2(M; \mathbb{Z}) \to H^2(M; \mathbb{R})$ of the set of nonzero classes $w \in H^2(M; \mathbb{R})$ such that $v + 2w$ lies on an edge of B_M^* for some vertex v of B_M^*. Note that Ξ_M is a finite set.

Friedl and Vidussi [70, Corollary 1.3] also showed that equality in Theorem 5.4.10 holds for all but finitely many circle bundles over a 3-manifold which is not exceptional:

Theorem 5.4.12 ([70]) *Let M be a closed irreducible 3-manifold which is not a closed graph manifold such that $\Delta_M^\phi \neq 0$ for all nontrivial $\phi \in H^1(M; \mathbb{Z})$, and let $p\colon N \to M$ be an oriented circle bundle with Euler class not in Ξ_N. Then*

$$x_N(\alpha) = |\alpha \cdot \alpha| + x_M(p_*\alpha)$$

for $\alpha \in H_2(N; \mathbb{Z})$. □

Remark 5.4.13 Friedl and Vidussi showed Theorem 5.4.12 also for nonpositively curved graph manifolds. By the work by Agol [2], Liu [136], and Przytycki and Wise [187, 217] the so-called virtually special theorem holds for irreducible nonpositively curved 3-manifolds. As a consequence, Agol's virtual fibering theorem also holds for such 3-manifolds. □

5.4.4 Harmonic Norm

We discuss relationships between the harmonic norm associated with a Riemannian metric and the Thurston norm.

Let M be a closed Riemannian 3-manifold with a metric h. The L^2-norm $||\cdot||_h$ on the vector space $\Omega^k(M)$ of k-forms on M is associated with the inner product

$$\langle \alpha, \beta \rangle = \int_M \alpha \wedge *\beta$$

for $\alpha, \beta \in \Omega^k(M)$, where $*$ is the Hodge star operator. The k-th homology group $H^k(M; \mathbb{R})$ is identified with the subspace of harmonic k-forms, and $||\cdot||_h$ induces a norm on $H^k(M; \mathbb{R})$. The induced norm is called the *harmonic norm* and is also denoted by $||\cdot||_h$. As it comes from a positive-definite inner product, the unit ball of $||\cdot||_h$ is a smooth ellipsoid.

In the study of the Seiberg–Witten monopole equations Kronheimer and Mrowka [118, 119] showed that the Thurston norm x_M is characterized in terms of the harmonic norm:

Theorem 5.4.14 ([119]) *Let M be a closed irreducible 3-manifold not containing non-separating tori. Then*

$$x_M(\phi) = \frac{1}{4\pi}\inf_h ||s_h||_h ||\phi||_h$$

for $\phi \in H^1(M;\mathbb{R})$, where s_h is the scalar curvature of h, and the infimum is taken over all Riemannian metrics h on M.

Remark 5.4.15 The original statement of [119, Theorem 2] is in terms of the dual Thurston norm x_M^* on $H_2(M;\mathbb{R})$. As explained in [18, Theorem 5.1], it is equivalent to that of Theorem 5.4.14. □

See [15, 197] for extensions of Theorem 5.4.14 in another approach studying harmonic 1-forms. See also [106].

For a closed hyperbolic 3-manifold M, by Mostow rigidity, the harmonic norm is uniquely determined by the underlying topology of M. We denote it by $||\cdot||_{L^2}$. Refining results of Bergeron, Şengün and Venkatesh [10] as well as Theorem 5.4.14, Brock and Dunfield [18] showed the following inequalities between the two norms $||\cdot||_{L^2}$ and x_M:

Theorem 5.4.16 ([18]) *Let M be a closed hyperbolic 3-manifold. Then*

$$\frac{\pi}{\sqrt{\mathrm{Vol}(M)}} x_M(\phi) \leq ||\phi||_{L^2} \leq \frac{10\pi}{\sqrt{\mathrm{inj}(M)}} x_M(\phi)$$

for $\phi \in H^1(M;\mathbb{R})$, where $\mathrm{inj}(M)$ is the injectivity radius of M, which is half the length of the shortest closed geodesic in M. □

Brock and Dunfield used the theory of minimal surfaces to prove Theorem 5.4.16. They also showed that the inequality is qualitatively sharp [18, Theorem 1.3].

Since the scalar curvature of a hyperbolic metric h is -6, specializing Theorem 5.4.14 to such h gives

$$\frac{2\pi}{3\sqrt{\mathrm{Vol}(M)}} x_M(\phi) \leq ||\phi||_{L^2},$$

which is weaker than the first inequality in Theorem 5.4.16. Lin [134] gave a gauge-theoretic proof of the stronger inequality.

5.5 Floer Homology

We look at the fact that Floer homology detects the Thurston norm and fiberedness of a 3-manifold.

With a motivation to better understand the Seiberg–Witten invariant, Ozsváth and Szabó [179, 180] introduced *Heegaard Floer homology*:

$$\widehat{HF}(M),\ HF^{\infty}(M),\ HF^{+}(M),\ HF^{-}(M).$$

Analogously, based directly on the Seiberg–Witten monopole equations, Kronheimer and Mrowka [120] introduced *monopole Floer homology*:

$$\widetilde{HM}_*(M),\ \widehat{HM}_*(M),\ \overline{HM}_*(M).$$

Passing through *embedded contact homology* ECH introduced by Hutchings and Taubes [93, 95, 96], Heegaard Floer homology and Monopole Floer homology were shown to be equivalent by Colin, Ghiggini and Honda [28–30], and Kutluhan, Li and Taubes [125–129].

In the following we focus on results in terms only of Heegaard Floer homology. But by the equivalence of the theories corresponding results hold also in terms of monopole Floer homology. For details including the definitions of the Floer homology groups, we refer the reader to the expositions [82, 92, 104, 176, 178, 181, 183] for Heegaard Floer homology, and to [120, 133] for monopole Floer homology. For combinatorial computations of Heegaard Floer homology see the survey article [142] and the references given there.

5.5.1 Heegaard Floer Homology

A *Heegaard diagram* for a closed 3-manifold M is a Heegaard surface Σ of genus g together with two systems α and β of simple closed curves $\alpha_1, \dots, \alpha_g$ and $\beta_1, \dots, \beta_g$ on Σ representing generators of $H_1(\Sigma; \mathbb{Z})$ and bounding disks in the two handlebodies in M respectively. Heegaard Floer homology is constructed by taking a Heegaard diagram (Σ, α, β) for M and applying Lagrangian intersection Floer theory [44, 72] to the tori $\alpha_1 \times \cdots \times \alpha_g$ and $\beta_1 \times \cdots \times \beta_g$ in the symmetric product of g copies of Σ [179, 180].

Heegaard Floer homology assigns to M a finitely generated abelian group $\widehat{HF}(M)$ and finitely generated $\mathbb{Z}[U]$-modules $HF^{\infty}(M)$, $HF^{+}(M)$, $HF^{-}(M)$, where U is a formal variable in the polynomial ring $\mathbb{Z}[U]$. Each group $HF^{\circ}(M)$ has the following decomposition over $\mathrm{Spin}^c(M)$:

$$HF^{\circ}(M) = \bigoplus_{\mathfrak{s}\in \mathrm{Spin}^c(M)} HF^{\circ}(M, \mathfrak{s}).$$

Furthermore, each group $HF^\circ(M,\mathfrak{s})$ carries an absolute $\mathbb{Z}/2\mathbb{Z}$-grading, and we can take the Euler characteristic $\chi(HF^\circ(M,\mathfrak{s}))$ with respect to the grading. These four flavors of Heegaard Floer homology are related by the following exact triangles:

$$\cdots \to HF^-(M,\mathfrak{s}) \to HF^\infty(M,\mathfrak{s}) \to HF^+(M,\mathfrak{s}) \to HF^-(M,\mathfrak{s}) \to \cdots,$$
$$\cdots \to \widehat{HF}(M,\mathfrak{s}) \to HF^+(M,\mathfrak{s}) \to HF^+(M,\mathfrak{s}) \to \widehat{HF}(M,\mathfrak{s}) \to \cdots,$$
$$\cdots \to HF^-(M,\mathfrak{s}) \to HF^-(M,\mathfrak{s}) \to \widehat{HF}(M,\mathfrak{s}) \to HF^-(M,\mathfrak{s}) \to \cdots.$$

Ozsváth and Szabó [180, Theorem 1.2] showed that $HF^+(M)$ is a categolification of Turaev's torsion function T_M:

Theorem 5.5.1 ([180]) *Let M be a closed 3-manifold and $\mathfrak{s} \in \mathrm{Spin}^c(M)$ such that $c_1(\mathfrak{s})$ is not torsion. Then*

$$\chi(HF^+(Y,\mathfrak{s})) = \pm T_M(\mathfrak{s}).$$

When M is a rational homology 3-sphere, $HF^+(M,\mathfrak{s})$ carries an absolute $\mathbb{Q}$-grading. The *correction term* or *d-invariant* $d(M,\mathfrak{s})$ is defined to be the minimal grading of nontorsion elements in the image of the map $\pi\colon HF^\infty(M,\mathfrak{s}) \to HF^+(M,\mathfrak{s})$. Ozsváth and Szabó [174, Theorem 1.3] showed that $d(M,\mathfrak{s})$ and $\chi(\mathrm{Coker}\,\pi)$ determine the Casson invariant of an integral homology sphere M.

Ozsváth and Szabó [177, Theorem 1.1] also showed that $\widehat{HF}(M)$ detects the Thurston norm.

Theorem 5.5.2 ([177]) *Let M be a closed 3-manifold M. Then*

$$x_M(\phi) = \min\{|\langle c_1(\mathfrak{s}) \cup \phi, [M]\rangle| \;;\; \mathfrak{s} \in \mathrm{Spin}^c(M) \text{ such that } \widehat{HF}(M,\mathfrak{s}) \neq 0\},$$

for $\phi \in H^1(M;\mathbb{R})$. □

Ni [164, Theorem 1.1] showed that $HF^+(M)$ detects fiberedness of M:

Theorem 5.5.3 ([164]) *Let M be a closed irreducible 3-manifold and S a properly embedded surface in M of negative Euler characteristic. If the group*

$$\bigoplus_{\mathfrak{s}\in \mathrm{Spin}^c(M) \text{ with } \langle c_1(\mathfrak{s}),[S]\rangle = \chi_-(S)} HF^+(M,\mathfrak{s})$$

is isomorphic to $\mathbb{Z}$, then M fibers over a circle with fiber S. □

Theorems 5.5.1, 5.5.2, 5.5.3 recover Theorem 5.3.6 on the abelian torsion.

5.5.2 Knot Floer Homology

Ozsváth and Szabó [178], and Rasmussen [189] independently defined the *knot Floer homology* $\widehat{HFK}(L)$ for a null-homologous link L in a closed 3-manifold M. This finitely generated abelian group $\widehat{HFK}(L)$ refines $\widehat{HF}(M)$ in the sense that there exists a spectral sequence from $\widehat{HFK}(L)$ converging to $\widehat{HF(M)}$.

In the case of a knot K in S^3, $\widehat{HFK}(K)$ is bigraded:

$$\widehat{HFK}(K) = \bigoplus_{(i,j)\in\mathbb{Z}^2} \widehat{HFK}_j(K,i),$$

where i and j are called the *Alexander grading* and the *homological grading* respectively. The group $\widehat{HFK}(K)$ is a categorification of the classical Alexander polynomial $\Delta_K(t)$ of K:

Theorem 5.5.4 ([178, 189]) *For a knot K in S^3 we have*

$$\Delta_K(t) = \sum_{(i,j)\in\mathbb{Z}^2} (-1)^j \left(\operatorname{rank} \widehat{HFK}_j(K,i)\right) t^i.$$

As shown in [175], $\widehat{HFK}(K)$ of an alternating knot K is completely determined by $\Delta_K(t)$ and the signature of K. See also [188].

As a consequence of the proof of Theorem 5.5.2, Ozsváth and Szabó [177, Theorem 1.2] also showed that $\widehat{HFK}(K)$ determines the knot genus $g(K)$:

Theorem 5.5.5 ([177]) *For a knot K in S^3 we have*

$$g(K) = \max\{i \in \mathbb{Z}\,;\, \bigoplus_{j\in\mathbb{Z}} \widehat{HFK}_j(K,i) \neq 0\}.$$

Theorem 5.5.5 implies that $\widehat{HFK}(K)$ detects the unknot. See [165, 182] for the case of links. In particular, Ozsváth and Szabó [182] showed that the Thurston and Alexander norms agree for the complements of alternating links in S^3.

Results by Ghiggini [79], Ni [163], and Juhász [102, 103] showed that $\widehat{HFK}(K)$ detects fiberedness of K:

Theorem 5.5.6 ([79, 102, 103, 163]) *A knot K in S^3 is fibered if and only if*

$$\bigoplus_{j\in\mathbb{Z}} \widehat{HFK}_j(K, g(K))$$

is isomorphic to $\mathbb{Z}$. □

More generally, Juhász [101] introduced the sutured Floer homology $SFH(M,\gamma)$ for sutured manifolds (M,γ) satisfying certain conditions, generalizing $\widehat{HF}(M)$ and $\widehat{HFK}(K)$. See [51] for the decategorification of $SFH(M,\gamma)$ and the relationship between $SFH(M,\gamma)$ and the Thurston norm for sutured manifolds. See also [4].

Kronheimer and Mrowka [121–123] extended instanton and monopole Floer homology to sutured manifolds. They also introduced a knot invariant $KHI(K)$ being a categorification of the classical Alexander polynomial and detecting the knot genus and fiberedness of knots.

5.6 Torsion Invariants

The Alexander polynomial has been generalized in different ways to three flavors of nonabelian Alexander polynomials: twisted Alexander polynomials introduced by Lin [132] and Wada [215], higher-order Alexander polynomials by Cochran [27] and Harvey [85], and L^2-Alexander invariant by Li and Zhang [131]. As seen in the equivalence of the Alexander polynomial and abelian torsion (Theorem 5.3.5), these generalized Alexander polynomials are also systematically studied in terms of nonabelian Reidemeister torsion: Reidemeister torsion associated with linear representations, higher-order Reidemeister torsion introduced by Friedl [48], and L^2-Alexander torsion by Dubois, Friedl and Lück [38]. We describe relationships between these invariants and the Thurston norm.

5.6.1 Reidemeister Torsion

We briefly review Reidemeister torsion associated with linear representations. See [149, 161, 209, 211] for details on Reidemeister torsion.

Let $C_* = (C_n \xrightarrow{\partial_n} C_{n-1} \to \cdots \to C_0)$ be a finite-dimensional acyclic chain complex over a commutative field $\mathbb{F}$, and let $c = \{c_i\}$ be a basis of C_*. We choose a basis b_i of $\operatorname{Im}\partial_{i+1}$ for each i. Taking a lift $\tilde{b}_{i-1}$ of $\tilde{b}_{i-1}$ in C_i and combining it with b_i, we have a basis $b_i\tilde{b}_{i-1}$ of C_i for each i. The *algebraic torsion* $\tau(C_*, c) \in \mathbb{F} \setminus \{0\}$ is defined as:

$$\tau(C_*, c) = \prod_{i=0}^{n} [b_i\tilde{b}_{i-1}/c_i]^{(-1)^{i+1}},$$

where $[b_i\tilde{b}_{i-1}/c_i]$ is the determinant of the base change matrix from c_i to $b_i\tilde{b}_{i-1}$. It can be checked that $\tau(C_*, c)$ does not depend on the choice of b_i and $b_i\tilde{b}_{i-1}$.

Let X be a connected CW-complex and R a noetherian UFD (e.g., R equals $\mathbb{Z}$ or $\mathbb{F}$). The cellular chain complex $C_*(\widetilde{X})$ of its universal cover $\widetilde{X}$ is a left $\mathbb{Z}[\pi_1 X]$-module. We think of $C_*(\widetilde{X})$ also as a right $\mathbb{Z}[\pi_1 X]$-module, using the involution of $\mathbb{Z}[\pi_1 X]$ reversing elements of $\pi_1 X$. Let $\rho\colon \pi_1 X \to \mathrm{GL}(n, R)$ be a representation.

For each nonnegative integer i the *i-th twisted homology group* $H_1^{\rho}(X; \mathbb{R}^n)$ is defined as:

$$H_i^{\rho}(X; R^n) = H_i(C_*(\widetilde{X}) \otimes_{\mathbb{Z}[\pi_1 X]} R^n).$$

The Reidemeister torsion $\tau_\rho(X) \in \mathbb{F}$ associated with a representation $\rho\colon \pi_1 X \to GL(n, \mathbb{F})$ is defined as follows. If $H_*^{\rho}(X; \mathbb{F}^n)$ does not vanish, then we define $\tau_\rho(X) = 0$. Otherwise, we choose a lift $\tilde{e}$ in $\widetilde{X}$ of each cell e of X, and define

$$\tau_\rho(X) = \tau\left(C_*(\widetilde{X}) \otimes_{\mathbb{Z}[\pi_1 X]} \mathbb{F}^n, \{\tilde{e} \otimes f_j\}_{e, 1 \le j \le n}\right),$$

where $f_1, \dots, f_n$ is the standard basis of $\mathbb{F}^n$. It is known that $\tau_\rho(X)$ is well-defined as a simple homotopy invariant up to multiplication by elements of $(\pm 1)^n \det \rho(\pi_1 X)$. Reidemeister torsion $\tau_\rho(X)$ is invariant under conjugation of representations ρ.

Remark 5.6.1 Turaev introduced a refinement of Reidemeister torsion $\tau_\rho(X)$ as an element of $\mathbb{F}$ without any indeterminacy, by fixing an orientation of $H_*(X; \mathbb{R})$ and an Euler structure of X, which is an equivalence class of the choice of lifts $\tilde{e}$ [209, 211]. □

5.6.2 Twisted Alexander Polynomials

We describe the results by Friedl and Vidussi [64, 65, 67, 69, 71], and Friedl and Nagel [60] that twisted Alexander polynomials detect the Thurston norm and fiberedness of a 3-manifold. For more details on twisted Alexander polynomials we refer the reader to the survey papers [37, 66, 151].

Let M be a 3-manifold with empty or toroidal boundary, $\psi\colon \pi_1 M \to F$ a homomorphism to a free abelian group F and $\rho\colon \pi_1 X \to \mathrm{GL}(n, R)$ a representation over a noetherian UFD R. We write $\psi \otimes \rho\colon \pi_1 M \to \mathrm{GL}(n, R[F])$ for the tensor representation defined by $\psi \otimes \rho(\gamma) = \psi(\gamma)\rho(\gamma)$ for $\gamma \in \pi_1 M$. Then $H_i^{\psi\otimes\rho}(M; R[F]^n)$ is a finitely generated $R[F]$-module for each i. The *i-th twisted Alexander polynomial* $\Delta_{M,i}^{\psi,\rho} \in R[F]$ of M associated with ψ and ρ is defined to be its order over $R[F]$, which is well-defined up to multiplication by units in $R[F]$. We set $\Delta_M^{\psi,\rho} = \Delta_{M,1}^{\psi,\rho}$. Twisted Alexander polynomials $\Delta_{M,i}^{\psi,\rho}$ are invariant under conjugation of representations ρ.

When $\psi\colon \pi_1 M \to H_1(M)_f$ is the canonical projection and ρ is the trivial representation, $\Delta_M^{\psi,\rho}$ coincides with the usual Alexander polynomial Δ_M of M defined in Sect. 5.3.1. We identify $H^1(M; \mathbb{Z})$ with $\mathrm{Hom}(\pi_1 M, \mathbb{Z})$ and $R[\mathbb{Z}]$ with the polynomial ring $R[t, t^{-1}]$ so that $1 \in \mathbb{Z}$ corresponds to t. Then for $\phi \in H^1(M; \mathbb{Z})$ twisted Alexander polynomials $\Delta_{M,i}^{\phi,\rho}$ are in $R[t, t^{-1}]$.

The Reidemeister torsion $\tau_{\psi\otimes\rho}(M) \in Q(R)(F)$ associated with $\psi\otimes\rho\colon \pi_1 M \to GL(n, Q(R)(F))$ is defined as in Sect. 5.6.1, where $Q(R)(F)$ is the quotient field of $R[F]$. For $\phi \in H^1(M;\mathbb{Z})$, $\tau_{\phi\otimes\rho}(M) \in Q(R)(t)$ is also defined. We define

$$\deg\left(a_l t^l + a_{l+1}t^{l+1} + \cdots + a_m t^m\right) = m - l$$

for $l, m \in \mathbb{Z}$ with $l < m$ and $a_i \in R$ with $a_l a_m \neq 0$, and further define

$$\deg \frac{p(t)}{q(t)} = \deg p(t) - \deg q(t)$$

for $p(t), q(t) \in R[t, t^{-1}] \setminus \{0\}$.

The following is a relationship between Reidemeister torsion and twisted Alexander polynomials. See also [109, 110].

Proposition 5.6.2 ([52, 54, 209]) *Let M be a 3-manifold with empty or toroidal boundary. For a homomorphism $\psi\colon \pi_1 M \to F$ and a representation $\rho\colon \pi_1 M \to \mathrm{GL}(n, R)$, if $\Delta_M^{\psi,\rho} \neq 0$, then*

$$\tau_{\psi\otimes\rho}(M) = \frac{\Delta_M^{\psi,\rho}}{\Delta_{M,0}^{\psi,\rho}\Delta_{M,2}^{\psi,\rho}}.$$

Remark 5.6.3 It can be checked for any ψ and ρ that $\Delta_{M,0}^{\psi,\rho} \neq 0$ and $\Delta_{M,3}^{\psi,\rho} = 1$, and that $\Delta_{M,2}^{\psi,\rho} = 0$ if and only if $\Delta_M^{\psi,\rho} = 0$. The second one follows from the first one and an Euler characteristic argument. See [66, Proposition 3.2]. □

Proposition 5.6.2, in particular, shows that $\Delta_M^{\psi,\rho} = 0$ if and only if $H_*^{\psi\otimes\rho}(M; Q(R)(F)^n) = 0$, and that for $\phi \in H^1(M;\mathbb{Z})$, if $\Delta_M^{\phi,\rho} \neq 0$, then

$$\deg \tau_{\phi\otimes\rho}(M) = \deg \Delta_M^{\phi,\rho} - \deg \Delta_{M,0}^{\phi,\rho} - \deg \Delta_{M,2}^{\phi,\rho}.$$

An advantage of twisted Alexander polynomials and the corresponding Reidemeister torsion is that if a representation ρ is given explicitly, then these invariants can be combinatorially computed, for example, by Fox derivatives for a presentation of the fundamental group.

Friedl and Kim [52, Theorems 1.1, 1.2] generalized Theorem 5.3.6 to twisted Alexander polynomials:

Theorem 5.6.4 ([52]) *Let M be a 3-manifold with empty or toroidal boundary and $\rho\colon \pi_1 M \to \mathrm{GL}(n, R)$ a representation. For $\phi \in H^1(M;\mathbb{Z})$, if $\Delta_M^{\phi,\rho} \neq 0$, then*

$$x_M(M) \geq \frac{1}{n} \deg \tau_{\phi\otimes\rho}(M).$$

Furthermore, $\Delta_M^{\phi,\rho} \neq 0$ *and equality holds if* ϕ *is fibered with* $M \neq S^1 \times D^2$ *and* $M \neq S^1 \times D^2$. □

Remark 5.6.5 Friedl and Kim [52, Theorems 1.3] also showed that under the assumptions of Theorem 5.6.4, if $\phi \in H^1(M;\mathbb{Z})$ is fibered, then $\Delta_M^{\phi,\rho}$ is monic, i.e., its highest and lowest coefficients are units in R. As explained in [52, Proposition 6.1], this theorem can be deduced from Theorem 5.6.4. □

Generalizing the Alexander norm, Friedl and Kim [54, Theorems 3.1, 3.2] also defined the *twisted Alexander norm* $||\cdot||_A^\rho$ on $H^1(M;\mathbb{R})$ associated with a representation $\rho\colon \pi_1 M \to \mathrm{GL}(n,R)$ and generalized Theorem 5.3.2. Let $\psi\colon \pi_1 M \to H_1(M)_f$ be the canonical projection, and write $\Delta_M^{\psi,\rho} = \sum_{h\in H_1(M)_f} a_h h$ for $a_h \in R$. If $\Delta_M^{\psi,\rho} = 0$, we define $||\cdot||_A^\rho = 0$. Otherwise, we define

$$||\phi||_A^\rho = \max\{\langle\phi, h-h'\rangle \mid h,h' \in H_1(M)_f \text{ such that } a_h a_{h'} \neq 0\}$$

for $\phi \in H^1(M;\mathbb{R})$. It is clear that $||\cdot||_A^\rho$ is a seminorm on $H^1(M;\mathbb{R})$ for any ρ. When ρ is the trivial representation, $||\cdot||_A^\rho = ||\cdot||_A$.

Theorem 5.6.6 ([54]) *Let* M *be a* 3*-manifold with empty or toroidal boundary and* $\rho\colon \pi_1 M \to \mathrm{GL}(n,R)$ *a representation. Suppose that* $b_1(M) > 0$*. Then*

$$x_M(\phi) \geq \frac{1}{n}||\phi||_A^\rho$$

for $\phi \in H^1(M;\mathbb{R})$*. Furthermore, equality holds for* ϕ *in the cone on a fibered face of* B_M *with* $M \neq S^1 \times S^2$ *and* $M \neq S^1 \times D^2$. □

Theorems 5.6.4, 5.6.6 also provide a fibering obstruction. Fibering obstructions on twisted Alexander polynomials in various level of generality were proved in [26, 49, 54, 80, 111, 112, 185].

As a corollary of the duality of (refined) Reidemeister torsion, Friedl, Kim and the author [55, Theorem 1.4] proved the following theorem.

Theorem 5.6.7 ([55]) *Let* M *be an irreducible* 3*-manifold with empty or toroidal boundary such that* $M \neq S^1 \times D^2$*, and let* $\rho\colon \pi_1 M \to U(n)$ *be a representation. For* $\phi \in H^1(M;\mathbb{Z})$ *whose restriction to any component of* ∂M *is nontrivial, if* $\Delta_M^{\phi,\rho} \neq 0$*, then*

$$\deg \tau_{\phi\otimes\rho}(M) \equiv n x_M(\phi) \mod 2.$$

Using the virtually special theorem by Agol [2], Liu [136] and Przytycki and Wise [187, 217], Friedl and Vidussi [71, Theorem 1.2, Corollary 5.10] with an extension by Friedl and Nagel [60, Theorem 1.3] showed that twisted Alexander polynomials determine the Thurston norm:

Theorem 5.6.8 ([60, 71]) *Let M be an irreducible 3-manifold with empty or toroidal boundary. Then there exists a representation $\rho\colon \pi_1 M \to \mathrm{GL}(n,\mathbb{C})$ with finite image such that $\Delta_M^{\phi,\rho} \neq 0$ and*

$$x_M(\phi) = \frac{1}{n} \deg \tau_{\phi\otimes\rho}(M)$$

for all $\phi \in H^1(M;\mathbb{Z})$. □

Corollary 5.6.9 ([71]) *Let M be an irreducible 3-manifold with empty or toroidal boundary. Suppose that $b_1(M) > 1$. Then there exists a representation $\rho\colon \pi_1 M \to \mathrm{GL}(n,\mathbb{C})$ with finite image such that*

$$x_M(\phi) = \frac{1}{n} ||\phi||_A^{\rho}$$

for all $\phi \in H^1(M;\mathbb{R})$. □

As explained in [71, Section 6] Theorem 5.6.8 gives an effective algorithm to compute the Thurston norm. We will see another algorithm in terms of normal surface theory in Sect. 5.7.1. Also, Theorem 5.6.8 and Corollary 5.6.9, in particular, show that the Thurston norm is an invariant of fundamental groups of 3-manifolds. We will discuss more on this point of view in Sect. 5.8.

Friedl and Vidussi [64, 65, 67, 69] also showed that twisted Alexander polynomials detect fiberedness of 3-manifolds. Based on different ideas using Novikov–Sikorav homology, Sikorav [196] showed the fibering detection theorem for general $\phi \in H^1(M;\mathbb{R})$.

Theorem 5.6.10 ([69]) *Let M be a 3-manifold with empty or toroidal boundary. If $\phi \in H^1(M;\mathbb{Z})$ is not fibered, then there exists a representation $\rho\colon \pi_1 M \to \mathrm{GL}(n,\mathbb{Z})$ with finite image such that $\Delta_M^{\phi,\rho} = 0$.* □

As a corollary of the fibering detection, together with the study of the Seiberg–Witten invariants of symplectic 4-manifolds, Friedl and Vidussi [64, 66, 68, 69] further showed that a closed 4-manifold which carries a free circle action admits a symplectic structure if and only if the orbit 3-manifold is fibered. The 'if' direction generalizes earlier work of Thurston [200]. See also [14, 84].

5.6.3 Higher-Order Alexander Polynomials

Cochran [27] and Harvey [85] introduced higher-order Alexander polynomials, analogues of the Alexander polynomial with coefficients in skew fields and showed that their degrees give lower bounds on the Thurston norm. Following Friedl [38, 48], we describe results in terms of corresponding higher-order Reidemeister torsion.

Let Γ be a torsion-free elementary-amenable group. By [34, 124] $\mathbb{Z}[\Gamma]$ is a right (and left) Ore domain, i.e., $\mathbb{Z}[\Gamma]$ embeds in its classical right ring of quotient $\mathbb{Q}(\Gamma) = \mathbb{Z}[\Gamma]\,(\mathbb{Z}[\Gamma] \setminus \{0\})^{-1}$. The Dieudonné determinant defines a canonical isomorphism $K_1(\mathbb{Q}(\Gamma)) \to \mathbb{Q}(\Gamma)^{\times}_{ab}$, where $\mathbb{Q}(\Gamma)^{\times}_{ab}$ is the abelianization of the multiplicative group $\mathbb{Q}(\Gamma) \setminus \{0\}$.

Let M be a 3-manifold with empty or toroidal boundary. We define the *higher-order Reidemeister torsion* $\tau_\rho(M) \in \mathbb{Q}(\Gamma)^{\times}_{ab} \cup \{0\}$ associated with an epimorphism $\rho\colon \pi_1 M \to \Gamma$ onto a torsion-free elementary-amenable group as follows: If $H_*^\rho(M; \mathbb{Q}(\Gamma)) \neq 0$, then we define $\tau_\rho(M) = 0$. Otherwise, we take a CW-complex structure of M, choose a lift $\tilde{e}$ of each cell e, and define

$$\tau_\rho(M) = \tau\left(C_*(\widetilde{M}) \otimes_{\mathbb{Z}[\pi_1 M]} \mathbb{Q}(\Gamma), \{\tilde{e} \otimes 1\}\right),$$

where $\tau(\cdot) \in \mathbb{Q}(\Gamma)^{\times}_{ab}$ is the algebraic torsion defined by replacing the usual determinant by the Dieudonné determinant. The invariant $\tau_\rho(M)$ is well-defined up to multiplication by elements in $\pm\Gamma$.

A pair (ρ, ϕ) of an epimorphism $\pi_1 M \to \Gamma$ onto a torsion-free elementary-amenable group and $\phi \in H^1(M; \mathbb{Z})$ is *admissible* if there exists a homomorphism $\phi_\Gamma\colon \Gamma \to \mathbb{Z}$ such that $\phi_\Gamma \circ \rho\colon \pi_1 M \to \mathbb{Z}$ coincides with ϕ under the identification $\operatorname{Hom}(\pi_1 M, \mathbb{Z})$ with $H^1(M; \mathbb{Z})$. For an admissible pair (ρ, ϕ) we define $\deg_\phi\colon \mathbb{Q}(\Gamma)^{\times}_{ab} \cup \{0\} \to \mathbb{Z} \cup \{-\infty\}$ as follows: Given $p = \sum_{g \in \Gamma} a_g g \in \mathbb{Z}[\Gamma] \setminus \{0\}$, we set

$$\deg_\phi p = \max\{\phi_\Gamma(g) - \phi_\Gamma(g')\ ;\ a_g a_{g'} \neq 0\},$$

and then define

$$\deg_\phi pq^{-1} = \deg_\phi p - \deg_\phi q$$

for $p, q \in \mathbb{Z}[\Gamma] \setminus \{0\}$, which induces a homomorphism $\mathbb{Q}(\Gamma)^{\times}_{ab} \to \mathbb{Z}$. We extend this to $\deg_\phi 0 = -\infty$. Now we have an integer-valued invariant $\deg_\phi \tau_\rho(M)$.

Example 5.6.11 Let M be a 3-manifold with empty or toroidal boundary. Examples of admissible pairs for M are given by *rational derived series* introduced by Cochran [27] and Harvey [85]: We set $\pi_r^{(0)} = \pi_1 M$ and inductively define

$$\pi_r^{(i)} = \{\gamma \in \pi_r^{(i-1)}\ ;\ \gamma^k \in [\pi_r^{(i-1)}, \pi_r^{(i-1)}] \text{ for some nonzero } k \in \mathbb{Z}\}.$$

Then for any n, $\pi_1 M/\pi_r^{(n)}$ is a poly-torsion-free-abelian group, and is, in particular, a torsion-free elementary-amenable group. We write $\rho_n\colon \pi_1 M \to \pi_1 M/\pi_r^{(n)}$ for the quotient map. Then (ρ_n, ϕ) is an admissible pair for any n and $\phi \in H^1(M; \mathbb{Z})$, and we can define $\tau_{\rho_n}(M)$ and $\deg_\phi \tau_{\rho_n}(M)$. The invariant $\tau_{\rho_n}(M)$ is called the *higher-order Reidemeister torsion of order n*. The invariant $\tau_{\rho_0}(M)$ of order 0 coincides with the abelian torsion of M. □

Extending the results by Cochran [27, Theorem 7.1] and Harvey [85, Theorem 10.1], Friedl and Harvey [48, Theorem 1.2], [87, Theorem 3.1] proved the following theorem, generalizing Theorem 5.3.6:

Theorem 5.6.12 ([48, 87]) *Let M be a 3-manifold with empty or toroidal boundary, and (ρ, ϕ) an admissible pair for M. Then*

$$x_M(\phi) \geq \deg_\phi \tau_\rho(M).$$

Furthermore, equality holds if ϕ is fibered with $M \neq S^1 \times S^2$ and $M \neq S^1 \times D^2$.

Theorem 5.6.12 also provides a fibering obstruction. Friedl [50] gave more fibering obstructions on higher-order Reidemeister torsion in terms of Novikov-Sikorav homology.

By the duality of higher-order Reidemeister torsion Friedl and Kim [53, Theorem 4.4] proved the following theorem:

Theorem 5.6.13 ([53]) *Let M be a closed 3-manifold or the complement of a link in S^3 and (ρ, ϕ) an admissible pair for M. Then*

$$\max\{\deg_\phi \tau_\rho(M), 0\} \equiv x_M(\phi) \mod 2.$$

An advantage of higher-order Alexander polynomials or higher-order Reidemeister torsion is that these invariants have monotonicity concerning epimorphisms ρ. Extending the result by Cochran [27, Theorem 5.4], Friedl [48, Theorem 1.3] and Harvey [86, Theorem 2.2, Corollary 2.10] proved the following:

Theorem 5.6.14 ([48, 86]) *Let M be a 3-manifold with empty or toroidal boundary, and $(\rho\colon \pi_1 M \to \Gamma, \phi)$ an admissible pair for M. Let $\varphi\colon \Gamma \to \Gamma'$ be an epimorphism such that $(\varphi \circ \rho, \phi)$ is admissible. Then*

$$\deg_\phi \tau_{\varphi\circ\rho}(M) \leq \deg_\phi \tau_\rho(M).$$

Corollary 5.6.15 ([27, 86]) *Let M be a 3-manifold with empty or toroidal boundary and $\phi \in H^1(M; \mathbb{Z})$. Then*

$$\deg_\phi \tau_{\rho_{n-1}}(M) \leq \deg_\phi \tau_{\rho_n}(M)$$

for each positive integer n.

5.6.4 L^2-Alexander Torsion

Li and Zhang [131] introduced the L^2-Alexander invariant, an L^2-analogue of higher-order Alexander polynomials. Dubois, Friedl and Lück [38] introduced the

L^2-Alexander torsion, the corresponding Reidemeister tosion generalizing the L^2-torsion [140]. Later, Friedl and Lück [56] introduced the universal L^2-torsion, which is a further generalization of the invariants. For more details on these L^2-invariants we refer the reader to the survey papers [37, 59, 141]. For basic terminology in L^2-theory see [140].

Let Γ be a torsion-free group. We denote by $L^2(\Gamma)$ the Hilbert space of formal sums $\sum_{\gamma \in \Gamma} a_\gamma \gamma$ for $a_\gamma \in \mathbb{C}$ such that $\sum_{\gamma \in \Gamma} |a_\gamma|^2 < \infty$. The group von Neumann algebra $\mathcal{N}(\Gamma)$ of Γ is defined to be the algebra of bounded Γ-equivalent operators on $L^2(\Gamma)$. We denote by $\mathcal{U}(\Gamma)$ the Ore localization of $\mathcal{N}(\Gamma)$ with respect to the multiplicative subset of nonzero divisors. (It is the algebra of affiliated operators.) Now we consider the division closure $\mathcal{D}(\Gamma)$ of $\mathbb{Z}[\Gamma]$ in $\mathcal{U}(\Gamma)$, which is the smallest subring S of $\mathcal{U}(\Gamma)$ containing $\mathbb{Z}[\Gamma]$ so that any element of S invertible in $\mathcal{U}(\Gamma)$ is already invertible in S. We say that Γ *satisfies the Atiyah conjecture* if given any $m \times n$ matrix A in $\mathbb{Z}[\Gamma]$ the L^2-dimesnion of the kernel of the map $L^2(\Gamma)^m \to L^2(\Gamma)^n$ sending $v \to vA$ for $v \in L^2(\Gamma)^m$ is a natural number. It is an open question whether all torsion-free groups satisfy the Atiyah conjecture. By [135] Γ satisfies the Atiyah conjecture if and only if $\mathcal{D}(\Gamma)$ is a skew-field.

Let M be an aspherical 3-manifold with empty or toroidal boundary. It is one of the consequences of the virtually special theorem [2, 136, 187, 217] that if M is not a closed graph manifold, then $\pi_1 M$ satisfies the Atiyah conjecture. Suppose that $\pi_1 M$ satisfies the Atiyah conjecture. We define $\rho_u^{(2)}(\widetilde{M}) \in \mathcal{D}(\pi_1 M)^\times_{ab}$ by

$$\rho_u^{(2)}(\widetilde{M}) = \tau\left(C_*(\widetilde{M}) \otimes_{\mathbb{Z}[\pi_1 M]} \mathcal{D}(\Gamma), \{\tilde{e} \otimes 1\}\right) \in \mathcal{D}(\pi_1 M)^\times_{ab}$$

as in the definition of higher-order Reidemeister torsion. The invariant $\rho_u^{(2)}(\widetilde{M})$ is well-defined up to multiplication by elements in $\pm\pi_1 M$ and coincides with the one introduced by Friedl and Lück in [56], called *the universal L^2-torsion* on M.

As described in [56, Section 2.4] the Fuglede–Kadison determinant defines a homomorphism $\det_{\mathcal{N}(\pi_1 M)} : \mathcal{D}(\pi_1 M)^\times_{ab} \to \mathbb{R}$. The image $\det_{\mathcal{N}(\pi_1 M)}(\rho_u^{(2)}(\widetilde{M}))$ coincides with the L^2-torsion $\rho^{(2)}(\widetilde{M})$ of M, which is equal to

$$-\frac{1}{6\pi} \sum \operatorname{Vol} M_i,$$

where M_i are the hyperbolic pieces in the JSJ decomposition of M. More generally, for $\phi \in H^1(M; \mathbb{Z})$ and $t \in \mathbb{R}_{>0}$, we can consider the Fuglede–Kadison determinant $\det_{\mathcal{N}(\pi_1 M),t}\, \mathcal{D}(\pi_1 M)^\times_{ab} \to \mathbb{R}$ twisted by the character on $H_1(M)_f$ sending $h \in H_1(M)_f$ to $t^{\langle \phi, h\rangle}$. We define a function $\bar{\rho}^{(2)}(\widetilde{M}; \phi) : \mathbb{R}_{>0} \to \mathbb{R}$ by

$$\bar{\rho}^{(2)}(\widetilde{M}; \phi)(t) = \det{}_{\mathcal{N}(\pi_1 M),t}(\rho_u^{(2)}(\widetilde{M})).$$

The invariant $\bar{\rho}^{(2)}(\widetilde{M}; \phi)$ is well-defined up to multiplication by functions of the form t^r for some $r \in \mathbb{R}$, and coincides with the *(full) L^2-Alexander torsion* of M introduced by Dubois, Friedl and Lück [38]. In fact, the L^2-Alexander torsion itself is defined for any irreducible 3-manifold M with empty or toroidal boundary. The following example complements the case of graph manifolds.

Example 5.6.16 Let M be a graph manifold which is not homeomorphic to $S^1 \times S^2$ nor to $S^1 \times D^2$. Dubois, Friedl and Lück [38, Theorem 1.1] showed that $\bar{\rho}^{(2)}(\widetilde{M};\phi)$ is represented by the function

$$\bar{\rho}(t) = \begin{cases} 1 & \text{if } t \le 1, \\ t^{x_M(\phi)} & \text{if } t \ge 1 \end{cases}$$

for $\phi \in H^1(M;\mathbb{Z})$. See also [91]. □

Based on the virtually special theorem [2, 136, 187, 217], Friedl and Lück [58, Theorem 0.1], and Liu [137, Theorem 1.2] independently proved the following theorem:

Theorem 5.6.17 ([58, 137]) *Let M be an irreducible 3-manifold with empty or toroidal boundary which is not homeomorphic to $S^1 \times D^2$. Then the limits $\lim_{t\to\infty} \frac{\bar{\rho}^{(2)}(\widetilde{M};\phi)}{\ln(t)}$ and $\lim_{t\to 0} \frac{\bar{\rho}^{(2)}(\widetilde{M};\phi)}{\ln(t)}$ exist, and*

$$\lim_{t\to\infty} \frac{\bar{\rho}^{(2)}(\widetilde{M};\phi)}{\ln(t)} - \lim_{t\to 0} \frac{\bar{\rho}^{(2)}(\widetilde{M};\phi)}{\ln(t)} = -x_M(\phi)$$

for $\phi \in H^1(M;\mathbb{Z})$. □

Remark 5.6.18 In Theorem 5.6.17 the difference of the limits on the left hand side can be regarded as the 'degree' of the function $\bar{\rho}^{(2)}(\widetilde{M};\phi)$. □

See [57, Theorem 0.2] for a related theorem on the ϕ-twisted L^2-Euler characteristic and the Thurston norm.

We now discuss an equivalence class of a pair of convex polytopes in $H_1(M;\mathbb{R})$ associated with the universal L^2-torsion.

The *Minkowski sum* of convex polytopes P and Q in $H_1(M;\mathbb{R})$ is defined by

$$P + Q = \{p + q \; ; \; p \in P \text{ and } q \in Q\}.$$

Two convex polytopes P and Q in $H_1(M;\mathbb{R})$ are *translation equivalent* if $Q = P + \{v\}$ for some $v \in H_1(M;\mathbb{R})$. We denote by $\mathfrak{P}(M)$ the set of translation equivalence classes of convex polytopes in $H_1(M;\mathbb{R})$. The Minkowski sum induces the structure of a commutative monoid on $\mathfrak{P}(M)$. We denote by $\mathfrak{G}(M)$ the Grothendieck group of $\mathfrak{P}(M)$. Let $\varphi\colon \pi_1 M \to H_1(M)_f$ be the canonical projection. Taking a section $H_1(M)_f \to \pi_1 M$, we can identify $\mathbb{Z}[\pi_1 M]$ with $\mathbb{Z}[\operatorname{Ker}\varphi][H_1(M)_f]$. We define a map $\mathcal{P}\colon \mathbb{Z}[\pi_1 M] \to \mathfrak{P}(M)$ as follows: For $f = \sum_{h\in H_1(M)_f} a_h h \in \mathbb{Z}[\operatorname{Ker}\varphi][H_1(M)_f] \setminus \{0\}$ we define $\mathcal{P}(f)$ to be the convex hull of all h with $a_h \neq 0$ in $H_1(M;\mathbb{R})$. The map extends as a homomorphism $\mathcal{P}\colon \mathcal{D}(\pi_1 M) \setminus \{0\} \to \mathfrak{G}(M)$, which further induces a homomorphism $\mathcal{P}\colon \mathcal{D}(\pi_1 M)^\times_{ab} \to \mathfrak{G}(M)$.

Friedl and Lück [56, Theorem 3.35] showed that the universal L^2-torsion determines the (dual) Thurston norm ball:

Theorem 5.6.19 ([56]) *Let M be an aspherical 3-manifold with empty or toroidal boundary such that $\pi_1 M$ satisfies the Atiyah conjecture. Then*

$$[B_M^*] = 2 \cdot \mathcal{P}(\rho_u^{(2)}(\widetilde{M})) \in \mathfrak{G}(M).$$

More generally, the above construction associates an equivalence class of a pair of convex polytopes $\mathcal{P}(\Gamma)$ also to a torsion-free group Γ satisfying the Atiyah conjecture and having a finite classifying space $B\Gamma$. For a group Γ admitting a presentation with two generators and one relator, reinterpreting results by Friedl, Schreve and Tillmann [61, 63], Friedl, Lück and Tillmann [59, Theorems 3.1, 5.4] described a combinatorial construction of $\mathcal{P}(\Gamma)$ from such a presentation, and showed that $\mathcal{P}(\Gamma)$ determines the Bieri–Neumann–Strebel invariant of Γ. See also [184] for related results. Funke and Kielak [73] studied $\mathcal{P}(\Gamma)$ and its relationships with the Bieri–Neumann–Strebel invariant and higher-order Alexander polynomials for free-by-cyclic groups Γ.

5.7 Triangulations

We discuss relationships between triangulations of a 3-manifold and its Thurston norm. There are algorithms to compute the Thurston norm ball and its fibered faces from triangulations in normal surface theory. Also, a $\mathbb{Z}/2\mathbb{Z}$-analogue of the Thurston norm is known to give a lower bound on the minimal number of tetrahedra in triangulations.

5.7.1 Thurston Norm Via Normal Surfaces

Algorithms to compute the Thurston norm ball in terms of normal surface theory are given by Tollefson and Wang [203, 204], and Cooper and Tillmann [32]. Here we overview a construction of the Thurston norm ball, following [32].

Let M be a closed irreducible 3-manifold and $\mathcal{T}$ a triangulation of M with t tetrahedra. Here we mean triangulations to be more general than simplicial triangulations. We allow triangulations to have simplices with self-identifications on their boundary. A triangulation is called 0-*efficient* if every normal 2-sphere bounds a 3-ball contained in a small neighborhood of a vertex. Recall that there are 7 types of normal discs in a tetrahedron Δ: 4 triangles around the vertices of Δ and 3 quadrilaterals separating the vertices of Δ into 2 pairs. A normal surface is an embedded surface in M whose intersection with each tetrahedron of $\mathcal{T}$ is a collection of disjoint normal discs. A fundamental fact in normal surface theory is that every incompressible surface in M is isotopic to a normal surface.

If we also take into account of transverse orientations of normal surfaces, there are 2 equivalence classes for each type of normal discs. We first consider the linear subspace of the real vector space of dimension $14t$ with a basis consisting of the equivalence classes of transversely oriented normal discs in the tetrahedra of $\mathcal{T}$. We denote by $NS^{\nu}(\mathcal{T})$ the linear subspace defined by the so-called matching equations: for each equivalence class of transversely oriented arcs γ in each triangle shared by 2 tetrahedra $\Delta_{\pm}$,

$$t_- + q_- = t_+ + q_+,$$

where $t_{\pm}$ and $q_{\pm}$ are the coefficients of the equivalence classes of transversely oriented triangle and quadrilateral in $\Delta_{\pm}$ respectively containing γ in their boundary. We denote by $NS^{\nu}_{+}(\mathcal{T})$ the subset of $NS^{\nu}(\mathcal{T})$ consisting of elements whose coefficients are all nonnegative. An element of $NS^{\nu}_{+}(\mathcal{T})$ is *admissible* if at most one type of quadrilateral in each tetrahedron is allowed to have nonzero coefficients.

By the construction every admissible integral point of $NS^{\nu}_{+}(\mathcal{T})$ is represented by a transversely oriented normal surface in M, and there are a linear map $\chi^*\colon NS^{\nu}(\mathcal{T}) \to \mathbb{R}$ and a surjective homomorphism $h\colon NS^{\nu}(\mathcal{T}) \to H_2(M;\mathbb{R})$ corresponding to the Euler characteristic and homology class of a normal surface respectively [32, Lemma 3, Proposition 4]. The set $P(\mathcal{T})$ of all elements of $NS^{\nu}_{+}(\mathcal{T})$ such that the sum of the coefficients is equal to 1 is a compact convex polytope in $NS^{\nu}(\mathcal{T})$. We define $B(\mathcal{T})$ to be the convex hull of the points $\frac{v}{|\chi^*(v)|}$, where v is an admissible vertex of $P(\mathcal{T})$ satisfying $\chi^*(v) < 0$. Now we can state the following theorem [32, Theorem 5]:

Theorem 5.7.1 ([32]) *Let M be a closed irreducible atoroidal 3-manifold with $b_1(M) > 0$, and $\mathcal{T}$ a simplicial or 0-efficient triangulation. Then $h(B(\mathcal{T}))$ coincides with B_M.*

Together with Haken's algorithm to check whether the complement of an open tubular neighborhood of an embedded surface S is homeomorphic to the product $S \times [0, 1]$ [145], Theorem 5.7.1 also gives an algorithm to determine the fibered faces of B_M [32, Algorithm 6]. An alternative algorithm to construct B_M is given in [204, Algorithm 5.9].

5.7.2 $\mathbb{Z}/2\mathbb{Z}$-Thurston Norm and Complexity of 3-Manifolds

Jaco, Rubinstein and Tillmann [100] introduced a $\mathbb{Z}/2\mathbb{Z}$-analogue of the Thurston norm. Let M be a closed irreducible 3-manifold. Every cohomology class in $H^1(M;\mathbb{Z}/2\mathbb{Z})$ is the Poincaré dual of the homology class represented by a possibly nonorientable embedded surface with some components in M. The $\mathbb{Z}/2\mathbb{Z}$-*Thurston norm* $x_{M,2}$ is the function on $H^1(M;\mathbb{Z}/2\mathbb{Z})$ defined by

$$x_{M,2}(\phi) = \min\{\chi_-(S)\ ;\ S \text{ is a possibly nonorientable embedded surface dual to } \phi\}$$

for $\phi \in H^1(M;\mathbb{Z}/2\mathbb{Z})$.

The *complexity* $c(M)$ of M is the minimal number of tetrahedra in triangulations of M. The number agrees with the one defined by Matveev [144] unless M is homeomorphic to S^3, $\mathbb{R}P^3$ or $L(3,1)$.

Generalizing earlier work [99], Jaco, Rubinstein, Spreer and Tillmann [100, Theorems 1, 2], [97, Theorems 1, 3], and Nakamura [160, Theorems 1.1, 1.2] showed that $x_{M,2}$ gives lower bounds on $c(M)$:

Theorem 5.7.2 ([97, 160]) *Let M be a closed irreducible 3-manifold not homeomorphic to $\mathbb{R}P^3$. Then*

$$c(M) \geq 1 + 2x_{M,2}(\phi).$$

Furthermore, if equality holds, then M is a lens space. □

Theorem 5.7.3 ([97, 100, 160]) *Let M be a closed irreducible 3-manifold and suppose $H^1(M;\mathbb{Z}/2\mathbb{Z})$ contains a subgroup H of rank 2. Then*

$$c(M) \geq 2 + \sum_{\phi \in H\setminus\{0\}} x_{M,2}(\phi).$$

Furthermore, if equality holds, then M is a generalized quaternionic space. □

Jaco, Rubinstein, Spreer and Tillmann [98, Theorem 1] also showed that another $\mathbb{Z}/2\mathbb{Z}$-analogue of the Thurston norm gives a lower bound on the minimal number of tetrahedra in ideal triangulations of cusped hyperbolic 3-manifolds.

For rational homology 3-spheres M, Ni and Wu [167, Corollary 1.2] gave a lower bound on $x_{M,2}$ by the d-invariant $d(M,\mathfrak{s})$, defined via gradings in Heegaard Floer homology [174].

5.8 Profinite Rigidity

What properties of 3-manifolds are determined by the set of finite quotients of their fundamental groups, or the profinite completions of their fundamental groups? We describe results on the Thuston norm and fiberedness by Boileau and Friedl [12], Bridson, Reid and Wilton [16, 17], and Liu [138]. We refer the reader to the survey [190] for recent work on profinite rigidity of residually finite groups.

The *profinite completion* $\widehat{\pi}$ of a group π is defined to be the limit $\lim_{\leftarrow} \pi/\Gamma$ of the inverse system $\{\pi/\Gamma\}_\Gamma$, where Γ runs over all finite index normal subgroups of

π. It easily follows from consideration of finite abelian quotients that if the profinite completions $\widehat{\pi}$ and $\widehat{\pi}'$ of finitely generated groups π and π' are isomorphic, then so are $H_1(\pi; \mathbb{Z})$ and $H_1(\pi'; \mathbb{Z})$.

Bridson, Reid and Wilton [16, Theorem A, Corollary 1.1], [17, Theorem C] showed the following rigidity theorems on fiberedness of 3-manifolds M with $b_1(M) = 1$:

Theorem 5.8.1 ([16]) *Let M_1 and M_2 be 3-manifolds with $b_1(M_1) = b_2(M_2) = 1$. Suppose that $\widehat{\pi_1 M_1}$ and $\widehat{\pi_1 M_2}$ are isomorphic. If M_1 has nonempty incompressible boundary and fibers over a circle such that $\pi_1 M_1$ is isomorphic to a semidirect product of the free group of rank r and $\mathbb{Z}$, then so does M_2.* □

Theorem 5.8.2 ([17]) *Let M_1 and M_2 be 3-manifolds with $b_1(M_1) = b_2(M_2) = 1$. Suppose that $\widehat{\pi_1 M_1}$ and $\widehat{\pi_1 M_2}$ are isomorphic. If M_1 is a closed hyperbolic 3-manifold fibering over a circle with fiber of genus g, then so is M_2.* □

Using different methods, Boileau and Friedl [13, Theorems 1.1, 4.6] showed the following theorems:

Theorem 5.8.3 ([13]) *Let M_1 and M_2 be aspherical 3-manifolds with empty or toroidal boundary. Suppose that there exists an isomorphism $\widehat{\pi_1 M_1} \to \widehat{\pi_1 M_2}$ such that the induced isomorphism $H_1\widehat{(M_1; \mathbb{Z})} \to H_1\widehat{(M_2; \mathbb{Z})}$ is induced by an isomorphism $f\colon H_1(M_1; \mathbb{Z}) \to H_1(M_2; \mathbb{Z})$. Then*

$$x_{M_1}(f^*\phi) = x_{M_2}(\phi)$$

for $\phi \in H_1(M_2; \mathbb{Z})$. Furthermore, $f^*\phi$ is fibered if and only if so is ϕ. □

Theorem 5.8.4 ([13]) *Let M_1 and M_2 be aspherical 3-manifolds with empty or toroidal boundary such that $H_1(M_1; \mathbb{Z})$ and $H_2(M_2; \mathbb{Z})$ are infinite cyclic groups. Let $\phi_1 \in H_1(M_1; \mathbb{Z})$ and $\phi_2 \in H_1(M_2; \mathbb{Z})$ be generators. Suppose that there exists an isomorphism $\widehat{\pi_1 M_1} \to \widehat{\pi_1 M_2}$. Then*

$$x_{M_1}(\phi_1) = x_{M_2}(\phi_2).$$

Furthermore, ϕ_1 is fibered if and only if so is ϕ_2. □

The proofs of Theorems 5.8.3, 5.8.4 rest on the facts that 3-manifold groups are good in the sense of Serre [195], and that the profinite completion of a 3-manifold group contains enough information on certain twisted Alexander polynomials determining the Thurston norm and fiberedness as in Theorems 5.6.8, 5.6.10.

For knots in S^3 Theorems 5.8.1, 5.8.4 can be restated as follows:

Corollary 5.8.5 ([13, 16]) *Let J and K be two knots in S^3 such that $\widehat{\pi_1 X_J}$ and $\widehat{\pi_1 X_K}$ are isomorphic. Then $g(J) = g(K)$. Furthermore, J is fibered if and only if K is fibered.* □

Bridson and Reid [16, Theorem A, Proposition 3.10] showed that the profinite completions of fundamental groups distinguish each of the complements of the trefoil knot and the figure-eight knot, and the Gieseking manifold among 3-manifolds. Also, Boileau and Friedl [13, Theorem 1.5] showed that every torus knot is also distinguished among knots in S^3.

For examples, Hempel's pairs [88] give examples of Seifert fibered spaces the profinite completions of whose fundamental groups are isomorphic but not satisfying the assumption as in Theorem 5.8.3. For hyperbolic 3-manifolds, Liu [138, Theorems 1.2, 1.3] strengthen Theorem 5.8.3:

Theorem 5.8.6 ([138]) *Let M_1 and M_2 be hyperbolic 3-manifolds. Suppose that there exists an isomorphism $\Phi\colon \widehat{\pi_1 M_1} \to \widehat{\pi_1 M_2}$. Then the following hold:*

1. *There exists a unit $\mu \in \widehat{\mathbb{Z}}$ such that $\Phi_*\colon \widehat{H_1(M)_f} \to \widehat{H_1(M_2)_f}$ is induced by an isomorphism $f\colon H_1(M_1)_f \to H_1(M_2)_f$ composed with the multiplication by μ.*
2. *We have*

$$x_{M_1}(f^*\phi) = x_{M_2}(\phi)$$

for $\phi \in H^1(M_2; \mathbb{Z})$. Furthermore, $f^*\phi$ is fibered if and only if so is ϕ.

□

Using Theorem 5.8.6 as one of the key ingredients, Liu[138, Theorem 1.1] proved the following theorem:

Theorem 5.8.7 ([138]) *For the fundamental group π of a hyperbolic 3-manifold there exists only finitely many 3-manifold groups whose profinite completions are isomorphic to $\widehat{\pi}$.* □

In [12] Boileau and Friedl showed that the Thurston norms of all finite covers of an aspherical 3-manifold determine whether it is a hyperbolic manifold, a graph manifold, or a mixed manifold, i.e., the JSJ decomposition is nontrivial and contains at least one hyperbolic component. Ueki [213] showed that the Alexander polynomial of a knot in S^3 is determined by the profinite completion of its knot group.

5.9 Conjectures and Questions

We conclude by collecting some conjectures and questions on the Thurston norm and related topics.

5.9.1 Realization Problem

In [202, Section 4] Thurston already gave a large variety of shapes for (dual) Thurston norm balls. However, the following naive question has been open since the Thurston norm was introduced:

Question 5.9.1 Which polyhedrons in $\mathbb{R}^n$ are realized as the (dual) Thurston norm balls of 3-manifolds? □

See [59, Question 6.11] for a restatement of this question and another one in terms of the universal L^2-torsion, and see also [184] for a related result.

5.9.2 Complexity Functions for Circle Bundles

In Theorem 5.4.12 Friedl and Vidussi showed that for all but finitely many circle bundles N over a non-exceptional 3-manifold M the complexity function $x_N\colon H_2(N;\mathbb{Z}) \to \mathbb{Z}$ is attained by the Thurston norm x_M. As remarked in [70, Section 1.3] we can ask whether the theorem holds for all circle bundles:

Question 5.9.2 Let M be a closed irreducible 3-manifold which is not a closed graph manifold such that $\Delta_M^{\phi} \neq 0$ for all nontrivial $\phi \in H^1(M;\mathbb{Z})$, and let $p\colon N \to M$ be an oriented circle bundle. Then does the equality

$$x_N(\alpha) = |\alpha \cdot \alpha| + x_M(p_*\alpha)$$

hold for any oriented circle bundle $p\colon N \to M$ and any $\alpha \in H_2(N;\mathbb{Z})$? □

5.9.3 Twisted Alexander Polynomials for Hyperbolic Knots

Let K be a hyperbolic knot in S^3 and $\phi \in H^1(X_K;\mathbb{Z})$ a generator. A holonomy representation $\rho\colon \pi_1 X_K \to \mathrm{PSL}(2,\mathbb{C})$ of the hyperbolic structure has a lift $\tilde{\rho}\colon \pi_1 X_K \to \mathrm{SL}(2,\mathbb{C})$ [33, 201]. Thus Reidemeister torsion $\tau_{\phi\otimes\tilde{\rho}}(X_K) \in \mathbb{C}(t)$ is defined, and can be checked to be in $\mathbb{C}[t,t^{-1}]$. Considering Turaev's refinement of $\tau_{\phi\otimes\tilde{\rho}}(X_K)$, Dunfield, Friedl and Jackson [40] introduced *hyperbolic torsion polynomial* $\mathcal{T}_K \in \mathbb{C}[t,t^{-1}]$ without any indeterminacy.

Based on experimental results for knots with at most 15 crossings, Dunfield, Friedl and Jackson [40, Conjecture 1.7] proposed the following conjecture:

Conjecture 5.9.3 ([40]) Let K be a hyperbolic knot in S^3. Then

$$\deg \mathcal{T}_K = 4g(K) - 2.$$

Furthermore, K is fibered if and only if the leading coefficient of $\mathcal{T}_K$ is equal to 1.□

Morifuji and Tran [150, 152, 153] showed that Conjecture 5.9.3 holds for a certain class of 2-bridge knots. Later, Agol and Dunfield [3] showed that equality in Conjecture 5.9.3 holds for all libroid hyperbolic knots in S^3, including all 2-bridge knots. The class of libroid knots is closed under Murasugi sum and contains all special arborescent knots obtained from plumbing oriented bands. See [153] for a generalization of Conjecture 5.9.3 for links.

5.9.4 Higher-Order Alexander Polynomials and the Knot Genus

Theorems 5.6.12, 5.6.14 naturally raise the question whether the higher-order Reidemeister torsion determines the Thurston norm. Dubois, Friedl and Lück [38, Conjecture 4.4] proposed the following conjecture:

Conjecture 5.9.4 ([38]) Let K be a knot in S^3 and $\phi \in H^1(X_K;\mathbb{Z})$ a generator. Then there exists an epimorphism $\rho\colon \pi_1 X_K \to \Gamma$ onto a torsion-free elementary-amenable group such that the pair (ρ, ϕ) is admissible and

$$\deg_\phi \tau_\rho(X_K) = 2g(K) - 1.$$

The following theorem proved by Friedl, Schreve and Tillmann [61, Theorem 3], in particular, shows that there are 'enough' epimorphisms from knot groups onto torsion-free elementary-amenable groups:

Theorem 5.9.5 ([61]) *Let M be an irreducible 3-manifold with empty or toroidal boundary, which is not a closed graph manifold. Then $\pi_1 M$ is a residually torsion-free elementary amenable group, i.e., for any nontrivial $\gamma \in \pi_1 M$ there exists an epimorphism $\rho\colon \pi_1 M \to \Gamma$ onto a torsion-free elementary-amenable group such that $\rho(\gamma)$ is nontrivial.* □

5.9.5 Lower Bounds on Complexity of 3-Manifolds

We have seen in Theorems 5.7.2, 5.7.3 that the $\mathbb{Z}/2\mathbb{Z}$-Thurston norm gives lower bounds on the complexity of 3-manifolds.

The following question was asked by Jaco, Rubinstein and Tillmann in [99, Section 1].

Question 5.9.6 ([99]) Determine an effective bound for the complexity of a closed irreducible 3-manifold M using a rank k subgroup of $H^1(M;\mathbb{Z}/2\mathbb{Z})$ for $k \geq 3$. □

5.9.6 Thurston Norm Balls of Finite Covers

As described in [5, Proposition 5.4.9], the following is a consequence of the virtually special theorem [2, 136, 187, 217] and the work of Cooper, Long and Reid [31]:

Theorem 5.9.7 *Let M be an irreducible 3-manifold with empty or toroidal boundary which is not a closed graph manifold. Then for each positive integer n there exists a finite cover $\widetilde{M}$ of M such that $B_{\widetilde{M}}$ has at least n top-dimensional faces.* □

The following is a version of Agol's virtual fibering theorem [1, 2] with a generalization by Kielak [108].

Theorem 5.9.8 ([1, 2, 108]) *Let M be an irreducible 3-manifold with empty or toroidal boundary which is not a closed graph manifold. Then there exists a finite covering $p\colon \widetilde{M} \to M$ such that for every nonfibered $\phi \in H^1(M;\mathbb{R})$, the pullback $p^*(\phi)$ lies in the cone on the boundary of a fibered face of $B_{\widetilde{M}}$.* □

As a corollary of Theorems 5.9.7, 5.9.8 we have the following:

Corollary 5.9.9 *Let M be an irreducible 3-manifold with empty or toroidal boundary which is not a closed graph manifold. Then for each positive integer n there exists a finite cover $\widetilde{M}$ of M such that $B_{\widetilde{M}}$ has at least n fibered faces.* □

See also [41, 139] for related results.

The following questions (also for nonpositively curved graph manifolds in Question 5.9.10) were asked by Aschenbrenner, Friedl and Wilton in [5, Questions 7.5.5, 7.5.6].

Question 5.9.10 ([5]) Let M be an irreducible 3-manifold with empty or toroidal boundary which is not a closed graph manifold. Does there exist a finite cover $\widetilde{M}$ of M such that all top-dimensional faces of $B_{\widetilde{M}}$ are fibered? □

Question 5.9.11 ([5]) Let M be an irreducible 3-manifold with empty or toroidal boundary which is not a graph manifold. Does there exist a finite cover $\widetilde{M}$ of M such that at least one top-dimensional face of $B_{\widetilde{M}}$ is not fibered? □

Acknowledgements The author wishes to thank Ken'ichi Ohshika and Athanase Papadopoulos for inviting him to write this survey article and for a careful check of it greatly improving the presentation. The author also wishes to thank Stefan Friedl for many helpful suggestions, and Martin Bridson, Alan Reid and Makoto Sakuma for valuable comments. The author was supported by JSPS KAKENHI Grant Numbers JP18K13404, JP18KK0071, JP18KK0380 and JP21H00986.

References

1. I. Agol, Criteria for virtual fibering. J. Topol. **1** (2), 269–284 (2008)
2. I. Agol, The virtual Haken conjecture. With an appendix by Agol, Daniel Groves, and Jason Manning. Doc. Math. 18, 1045–1087 (2013)

3. I. Agol, N.M. Dunfield, *Certifying the Thurston Norm via* $\mathrm{SL}(2, \mathbb{C})$*-twisted Homology*. What's next? Annals of Mathematics Studies, vol. 205 (Princeton University Press, Princeton, NJ, 2020), pp. 1–20
4. I. Altman, The sutured Floer polytope and taut depth-one foliations. Algebr. Geom. Topol. **14**(4), 1881–1923 (2014)
5. M. Aschenbrenner, S. Friedl, H. Wilton, 3*-Manifold Groups*. EMS Series of Lectures in Mathematics (European Mathematical Society (EMS), Zürich, 2015). xiv+215 pp.
6. D. Auckly, The Thurston norm and three-dimensional Seiberg-Witten theory. Osaka J. Math. **33**(3), 737–750 (1996)
7. K.L. Baker, S.A. Taylor, Dehn filling and the Thurston norm. J. Differ. Geom. **112**(3), 391–409 (2019)
8. S. Bauer, A stable cohomotopy refinement of Seiberg-Witten invariants. II. Invent. Math. **155**(1), 21–40 (2004)
9. S. Bauer, M. Furuta, A stable cohomotopy refinement of Seiberg-Witten invariants. I. Invent. Math. **155**(1), 1–19 (2004)
10. N. Bergeron, M.H. Şengün, A. Venkatesh, Torsion homology growth and cycle complexity of arithmetic manifolds. Duke Math. J. **165**(9), 1629–1693 (2016)
11. R. Bieri, W.D. Neumann, R. Strebel, A geometric invariant of discrete groups. Invent. Math. **90**(3), 451–477 (1987)
12. M. Boileau, S. Friedl, The virtual Thurston seminorm of 3-manifolds. Osaka J. Math. **56**(1), 51–63 (2019)
13. M. Boileau, S. Friedl, The profinite completion of 3-manifold groups, fiberedness and the Thurston norm. What's next? Annals of Mathematics Studies, vol. 205 (Princeton University Press, Princeton, NJ, 2020), pp. 21–44
14. J. Bowden, The topology of symplectic circle bundles. Trans. Am. Math. Soc. **361**(10), 5457–5468 (2009)
15. H.L. Bray, D.L. Stern, Scalar curvature and harmonic one-forms on three-manifolds with boundary (2019). arXiv:1911.06803
16. M.R. Bridson, A.W. Reid, *Profinite Rigidity, Fibering, and the Figure-Eight Knot*. What's next? Annals of Mathematics Studies, vol. 205 (Princeton Univ. Press, Princeton, NJ, 2020), pp. 45–64
17. M.R. Bridson, A.W. Reid, H. Wilton, Profinite rigidity and surface bundles over the circle. Bull. Lond. Math. Soc. **49**(5), 831–841 (2017)
18. J.F. Brock, N.M. Dunfield, Norms on the cohomology of hyperbolic 3-manifolds. Invent. Math. **210**(2), 531–558 (2017)
19. D. Calegari, Surface subgroups from homology. Geom. Topol. *12*(4), 1995–2007 (2008)
20. D. Calegari, Faces of the scl norm ball. Geom. Topol. **13**(3), 1313–1336 (2009)
21. D. Calegari, *scl*. MSJ Memoirs, vol. 20 (Mathematical Society of Japan, Tokyo, 2009). xii+209 pp.
22. D. Calegari, Certifying incompressibility of noninjective surfaces with scl. Pacific J. Math. **262**(2), 257–262 (2013)
23. D. Calegari, C. Gordon, Knots with small rational genus. Comment. Math. Helv. **88**(1), 85–130 (2013)
24. A. Candel, L. Conlon, *Foliations II*. Graduate Studies in Mathematics, vol. 60 (American Mathematical Society, Providence, RI, 2003). xiv+545 pp.
25. J. Cantwell, L. Conlon, The sutured Thurston norm, in *Foliations 2012* (World Sci. Publ., Hackensack, NJ, 2013), pp. 41–66
26. J.C. Cha, Fibred knots and twisted Alexander invariants. Trans. Am. Math. Soc. **355**(10), 4187–4200 (2003)
27. T.D. Cochran, Noncommutative knot theory. Algebr. Geom. Topol. **4**, 347–398 (2004)
28. V. Colin, P. Ghiggini, K. Honda, Equivalence of Heegaard Floer homology and embedded contact homology via open book decompositions. Proc. Natl. Acad. Sci. U. S. A. **108**(20), 8100–8105 (2011)

29. V. Colin, P. Ghiggini, K. Honda, The equivalence of Heegaard Floer homology and embedded contact homology via open book decompositions II (2012). arXiv:1208.1077
30. V. Colin, P. Ghiggini, K. Honda, The equivalence of Heegaard Floer homology and embedded contact homology III: from hat to plus (2012). arXiv:1208.1526
31. D. Cooper, D.D. Long, A.W. Reid, Essential closed surfaces in bounded 3-manifolds. J. Am. Math. Soc. **10**(3), 553–563 (1997)
32. D. Cooper, S. Tillmann, The Thurston norm via normal surfaces. Pacific J. Math. **239**(1), 1–15 (2009)
33. M. Culler, Lifting representations to covering groups. Adv. Math. **59**(1), 64–70 (1986)
34. J. Dodziuk, P. Linnell, V. Mathai, T. Schick, S. Yates, Approximating L^2-invariants and the Atiyah conjecture. Dedicated to the memory of Jürgen K, Moser. Comm. Pure Appl. Math., vol. 56(7) (2003), pp. 839–873
35. S. Dowdall, I. Kapovich, C.J. Leininger, Dynamics on free-by-cyclic groups. Geom. Topol. **19**(5), 2801–2899 (2015)
36. S. Dowdall, I. Kapovich, C.J. Leininger, McMullen polynomials and Lipschitz flows for free-by-cyclic groups. J. Eur. Math. Soc. **19**(11), 3253–3353 (2017)
37. J. Dubois, S. Friedl, W. Lück, Three flavors of twisted invariants of knots. *Introduction to Modern Mathematics*. Advances in Mathematics (ALM), vol. 33 (International Press, Somerville, MA, 2015), pp. 143–169
38. J. Dubois, S. Friedl, W. Lück, The L^2-Alexander torsion of 3-manifolds. J. Topol. **9**(3), 889–926 (2016)
39. N.M. Dunfield, Alexander and Thurston norms of fibered 3-manifolds. Pacific J. Math. **200**(1), 43–58 (2001)
40. N.M. Dunfield, S. Friedl, N. Jackson, Twisted Alexander polynomials of hyperbolic knots. Exp. Math. **21**(4), 329–352 (2012)
41. N.M. Dunfield, D. Ramakrishnan, Increasing the number of fibered faces of arithmetic hyperbolic 3-manifolds. Am. J. Math. **132**(1), 53–97 (2010)
42. A.L. Edmonds, C. Livingston, Group actions on fibered three-manifolds. Comment. Math. Helv. **58**(4), 529–542 (1983)
43. R. Fintushel, R.J. Stern, Immersed spheres in 4-manifolds and the immersed Thom conjecture. Turkish J. Math. **19**(2), 145–157 (1995)
44. A. Floer, Morse theory for Lagrangian intersections. J. Differ. Geom. **28**(3), 513–547 (1988)
45. R. Flores, D. Kahrobaei, T. Koberda, A cryptographic application of the Thurston norm. Int. J. Comput. Math. Comput. Syst. Theory **5**(1), 15–24 (2020)
46. D. Fried, The geometry of cross sections to flows. Topology **21**(4), 353–371 (1982)
47. D. Fried, Flow equivalence, hyperbolic systems and a new zeta function for flows. Comment. Math. Helv. **57**(2), 237–259 (1982)
48. S. Friedl, Reidemeister torsion, the Thurston norm and Harvey's invariants. Pacific J. Math. **230**(2), 271–296 (2007)
49. S. Friedl, Twisted Reidemeister torsion, the Thurston norm and fibered manifolds. Geom. Dedicata **172**, 135–145 (2014)
50. S. Friedl, Novikov homology and non-commutative Alexander polynomials. J. Knot Theory Ramif. **26**(2), 1740013, 30 pp. (2017)
51. S. Friedl, A. Juhász, J. Rasmussen, The decategorification of sutured Floer homology. J. Topol. **4**(2), 431–478 (2011)
52. S. Friedl, T. Kim, The Thurston norm, fibered manifolds and twisted Alexander polynomials. Topology **45**(6), 929–953 (2006)
53. S. Friedl, T. Kim, The parity of the Cochran-Harvey invariants of 3-manifolds. Trans. Am. Math. Soc. **360**(6), 2909–2922 (2008)
54. S. Friedl, T. Kim, Twisted Alexander norms give lower bounds on the Thurston norm. Trans. Am. Math. Soc. **360**(9), 4597–4618 (2008)
55. S. Friedl, T. Kim and T. Kitayama, *Poincaré duality and degrees of twisted Alexander polynomials*, Indiana Univ. Math. J. **61** (2012), no. 1, 147–192.

56. S. Friedl and W. Lück, *Universal L^2-torsion, polytopes and applications to 3-manifolds*, Proc. Lond. Math. Soc. (3) **114**(6), 1114–1151 (2017)
57. S. Friedl, W. Lück, L^2-Euler characteristics and the Thurston norm. Proc. Lond. Math. Soc. (3) **118**(4), 857–900 (2019)
58. S. Friedl, W. Lück, The L^2-torsion function and the Thurston norm of 3-manifolds. Comment. Math. Helv. **94**(1), 21–52 (2019)
59. S. Friedl, W. Lück, S. Tillmann, Groups and polytopes, in *Breadth in Contemporary Topology*. Proc. Sympos. Pure Math., vol. 102 (Amer. Math. Soc., Providence, RI, 2019), pp. 57–77
60. S. Friedl, M. Nagel, Twisted Reidemeister torsion and the Thurston norm: graph manifolds and finite representations. Illinois J. Math. **59**(3), 691–705 (2015)
61. S. Friedl, K. Schreve, S. Tillmann, Thurston norm via Fox calculus. Geom. Topol. **21**(6), 3759–3784 (2017)
62. S. Friedl, D. Silver, S. Williams, The Turaev and Thurston norms. Pacific J. Math. **284**(2), 365–382 (2016)
63. S. Friedl, S. Tillmann, Two-generator one-relator groups and marked polytopes. Ann. Inst. Fourier (Grenoble) **70**(2), 831–879 (2020)
64. S. Friedl, S. Vidussi, Symplectic $S^1 \times N^3$, subgroup separability, and vanishing Thurston norm. J. Am. Math. Soc. **21**(2), 597–610 (2008)
65. S. Friedl, S. Vidussi, Twisted Alexander Polynomials and fibered 3-manifolds, in *Low-Dimensional and Symplectic Topology*, vol. 82. Proc. Sympos. Pure Math. (Amer. Math. Soc., Providence, RI, 2011), pp. 111–130
66. S. Friedl, S. Vidussi, A survey of twisted Alexander polynomials, in *The Mathematics of Knots*, vol. 1. Contrib. Math. Comput. Sci. (Springer, Heidelberg, 2011), pp. 45–94
67. S. Friedl, S. Vidussi, Twisted Alexander polynomials detect fibered 3-manifolds. Ann. Math. (2) **173**(3), 1587–1643 (2011)
68. S. Friedl, S. Vidussi, Construction of symplectic structures on 4-manifolds with a free circle action. Proc. Roy. Soc. Edinburgh Sect. A **142**(2), 359–370 (2012)
69. S. Friedl, S. Vidussi, A vanishing theorem for twisted Alexander polynomials with applications to symplectic 4-manifolds. J. Eur. Math. Soc. **15**(6), 2027–2041 (2013)
70. S. Friedl, S. Vidussi, Minimal genus in 4-manifolds with a free circle action. Adv. Math. **250**, 570–587 (2014)
71. S. Friedl, S. Vidussi, The Thurston norm and twisted Alexander polynomials. J. Reine Angew. Math. **707**, 87–102 (2015)
72. K. Fukaya, Y.-G. Oh, H. Ohta, K. Ono, *Lagrangian Intersection Floer Theory: Anomaly and Obstruction. Part I*. AMS/IP Studies in Advanced Mathematics, vol. 46(1) (American Mathematical Society, Providence, RI; International Press, Somerville, MA, 2009), xii+396 pp.
73. F. Funke, D. Kielak, Alexander and Thurston norms, and the Bieri-Neumann-Strebel invariants for free-by-cyclic groups. Geom. Topol. **22**(5), 2647–2696 (2018)
74. D. Gabai, Foliations and the topology of 3-manifolds. J. Differ. Geom. **18**(3), 445–503 (1983)
75. D. Gabai, Foliations and genera of links. Topology **23**(4), 381–394 (1984)
76. D. Gabai, Foliations and the topology of 3-manifolds, III. J. Differ. Geom. **26**(3), 479–536 (1987)
77. D. Gabai, 3 lectures on foliations and laminations on 3-manifolds, in *Laminations and Foliations in Dynamics, Geometry and Topology* (Stony Brook, NY, 1998). Contemp. Math., vol. 269 (Amer. Math. Soc., Providence, RI, 2001), pp. 87–109
78. G. Gardam, D. Kielak, A.D. Logan, JSJ decompositions and polytopes for two-generator one-relator groups (2021). arXiv:2101.02193
79. P. Ghiggini, Knot Floer homology detects genus-one fibred knots. Am. J. Math. **130**(5), 1151–1169 (2008)
80. H. Goda, T. Kitano, T. Morifuji, Reidemeister torsion, twisted Alexander polynomial and fibered knots. Comment. Math. Helv. **80**(1), 51–61 (2005)
81. S.E. Goodman, Closed leaves in foliations of codimension one. Comment. Math. Helv. **50**(3), 383–388 (1975)

82. J.E. Greene, Heegaard Floer homology. Notices Amer. Math. Soc. **68**(1), 19–33 (2021)
83. M. Gromov, Volume and bounded cohomology. Inst. Hautes Études Sci. Publ. Math. 56 (1982), 5–99 (1983)
84. B. Hajduk, R. Walczak, Symplectic forms invariant under free circle actions on 4-manifolds. Trans. Am. Math. Soc. **358**(5), 1953–1970 (2006)
85. S.L. Harvey, Higher-order polynomial invariants of 3-manifolds giving lower bounds for the Thurston norm. Topology **44**(5), 895–945 (2005)
86. S.L. Harvey, Monotonicity of degrees of generalized Alexander polynomials of groups and 3-manifolds. Math. Proc. Cambridge Philos. Soc. **140**(3), 431–450 (2006)
87. S. Harvey, S. Friedl, Non-commutative multivariable Reidemester torsion and the Thurston norm. Algebr. Geom. Topol. **7**, 755–777 (2007)
88. J. Hempel, Some 3-manifold groups with the same finite quotients (2014). arXiv:1409.3509
89. F. Henneke, D. Kielak, Agrarian and L^2-invariants (2018). arXiv:1809.08470
90. F. Henneke, D. Kielak, The agrarian polytope of two-generator one-relator groups. J. Lond. Math. Soc. (2) **102**(2), 722–748 (2020)
91. G. Herrmann, The L^2-Alexander torsion for Seifert fiber spaces. Arch. Math. (Basel) **109**(3), 273–283 (2017)
92. J. Hom, A survey on Heegaard Floer homology and concordance. J. Knot Theory Ramif. **26**(2), 1740015, 24 pp. (2017)
93. M. Hutchings, An index inequality for embedded pseudoholomorphic curves in symplectizations. J. Eur. Math. Soc. **4**(4), 313–361 (2002)
94. M. Hutchings, C.H. Taubes, An introduction to the Seiberg-Witten equations on symplectic manifolds, in *Symplectic Geometry and Topology*, Park City, UT, 1997, IAS/Park City. Math. Ser., vol. 7 (Amer. Math. Soc., Providence, RI, 1999), pp. 103–142
95. M. Hutchings, C.H. Taubes, Gluing pseudoholomorphic curves along branched covered cylinders I. J. Symplectic Geom. **5**(1), 43–137 (2007)
96. M. Hutchings, C.H. Taubes, Gluing pseudoholomorphic curves along branched covered cylinders II. J. Symplectic Geom. **7**(1), 29–133 (2009)
97. W. Jaco, J.H. Rubinstein, J. Spreer, S. Tillmann, $\mathbb{Z}_2$-Thurston norm and complexity of 3-manifolds, II. Algebr. Geom. Topol. **20**(1), 503–529 (2020)
98. W. Jaco, J.H. Rubinstein, J. Spreer, S. Tillmann, On minimal ideal triangulations of cusped hyperbolic 3-manifolds. J. Topol. **13**(1), 308–342 (2020)
99. W. Jaco, J.H. Rubinstein, S. Tillmann, Minimal triangulations for an infinite family of lens spaces. J. Topol. **2**(1), 157–180 (2009)
100. W. Jaco, J.H. Rubinstein, S. Tillmann, $\mathbb{Z}_2$-Thurston norm and complexity of 3-manifolds. Math. Ann. **356**(1), 1–22 (2013)
101. A. Juhász, Holomorphic discs and sutured manifolds. Algebr. Geom. Topol. **6**, 1429–1457 (2006)
102. A. Juhász, Floer homology and surface decompositions. Geom. Topol. **12**(1), 299–350 (2008)
103. A. Juhász, The sutured Floer homology polytope. Geom. Topol. **14**(3), 1303–1354 (2010)
104. A. Juhász, A survey of Heegaard Floer homology, in *New Ideas in Low Dimensional Topology*. Ser. Knots Everything, vol. 56 (World Sci. Publ., Hackensack, NJ, 2015), pp. 237–296
105. M. Kapovich, *Hyperbolic Manifolds and Discrete Groups*. Progress in Mathematics, vol. 183 (Birkhäuser Boston, Inc., Boston, MA, 2001). xxvi+467 pp.
106. G. Katz, Harmonic maps $M^3 \to S^1$ and 2-cycles, realizing the Thurston norm. Trans. Am. Math. Soc. **357**(3), 1177–1224 (2005)
107. D. Kielak, The Bieri-Neumann-Strebel invariants via Newton polytopes. Invent. Math. **219**(3), 1009–1068 (2020)
108. D. Kielak, Residually finite rationally solvable groups and virtual fibring. J. Am. Math. Soc. **33**(2), 451–486 (2020)
109. P. Kirk, C. Livingston, Twisted Alexander invariants, Reidemeister torsion, and Casson-Gordon invariants. Topology **38**(3), 635–661 (1999)

110. T. Kitano, Twisted Alexander polynomial and Reidemeister torsion. Pacific J. Math. **174**(2), 431–442 (1996)
111. T. Kitano, T. Morifuji, Divisibility of twisted Alexander polynomials and fibered knots. Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **4**(1), 179–186 (2005)
112. T. Kitayama, Normalization of twisted Alexander invariants. Int. J. Math. **26**(10), 1550077, 21 pp. (2015)
113. H. Konno, Bounds on genus and configurations of embedded surfaces in 4-manifolds. J. Topol. **9**(4), 1130–1152 (2016)
114. P.B. Kronheimer, Embedded surfaces and gauge theory in three and four dimensions, *Surveys in Differential Geometry*, Cambridge, MA, 1996, vol. III (Int. Press, Boston, MA, 1998), pp. 243–298
115. P.B. Kronheimer, Minimal genus in $S^1 \times M^3$. Invent. Math. **135**(1), 45–61 (1999)
116. P.B. Kronheimer, T.S. Mrowka, The genus of embedded surfaces in the projective plane. Math. Res. Lett. **1**(6), 797–808 (1994)
117. P.B. Kronheimer, T.S. Mrowka, Embedded surfaces and the structure of Donaldson's polynomial invariants. J. Differ. Geom. **41**(3), 573–734 (1995)
118. P.B. Kronheimer, T.S. Mrowka, Monopoles and contact structures. Invent. Math. **130**(2), 209–255 (1997)
119. P.B. Kronheimer, T.S. Mrowka, Scalar curvature and the Thurston norm. Math. Res. Lett. **4**(6), 931–937 (1997)
120. P. Kronheimer, T. Mrowka, *Monopoles and Three-Manifolds*. New Mathematical Monographs, vol. 10 (Cambridge University Press, Cambridge, 2007), xii+796 pp.
121. P. Kronheimer, T. Mrowka, Knots, sutures, and excision. J. Differ. Geom. **84**(2), 301–364 (2010)
122. P. Kronheimer, T. Mrowka, Instanton Floer homology and the Alexander polynomial. Algebr. Geom. Topol. **10**(3), 1715–1738 (2010)
123. P.B. Kronheimer, T.S. Mrowka, Knot homology groups from instantons. J. Topol. **4**(4), 835–918 (2011)
124. P.H. Kropholler, P.A. Linnell, J.A. Moody, Applications of a new K-theoretic theorem to soluble group rings. Proc. Am. Math. Soc. **104**(3), 675–684 (1988)
125. C. Kutluhan, Y.J. Lee, C.H. Taubes, $HF = HM$, I: Heegaard Floer homology and Seiberg-Witten Floer homology. Geom. Topol. **24**(6), 2829–2854 (2020)
126. C. Kutluhan, Y.J. Lee, C.H. Taubes, $HF = HM$, II: Reeb orbits and holomorphic curves for the ech/Heegaard Floer correspondence. Geom. Topol. **24**(6), 2855–3012 (2020)
127. C. Kutluhan, Y.J. Lee, C.H. Taubes, $HF = HM$, III: holomorphic curves and the differential for the ech/Heegaard Floer correspondence. Geom. Topol. **24**(6), 3013–3218 (2020)
128. C. Kutluhan, Y.J. Lee, C.H. Taubes, $HF = HM$, IV: The Sieberg-Witten Floer homology and ech correspondence. Geom. Topol. **24**(7), 3219–3469 (2020)
129. C. Kutluhan, Y.J. Lee, C.H. Taubes, $HF = HM$, V: Seiberg-Witten Floer homology and handle additions. Geom. Topol. **24**(7), 3471–3748 (2020)
130. M. Lackenby, The efficient certification of knottedness and Thurston norm. Adv. Math. **387**, 142 pp. (2021). Paper No. 107796
131. W. Li, W. Zhang, An L^2-Alexander invariant for knots. Commun. Contemp. Math. **8**(2), 167–187 (2006)
132. X.S. Lin, Representations of knot groups and twisted Alexander polynomials. Acta Math. Sin. (Engl. Ser.) **17**(3), 361–380 (2001)
133. F. Lin, Lectures on monopole Floer homology, in *Proceedings of the Gökova Geometry-Topology Conference 2015* (Gökova Geometry/Topology Conference (GGT), Gökova, 2016), pp. 39–80
134. F. Lin, Monopole Floer homology and the spectral geometry of three-manifolds. Comm. Anal. Geom. **28**(5), 1211–1219 (2020)
135. P.A. Linnell, Division rings and group von Neumann algebras. Forum Math. **5**(6), 561–576 (1993)

136. Y. Liu, Virtual cubulation of nonpositively curved graph manifolds. J. Topol. **6**(4), 793–822 (2013)
137. Y. Liu, Degree of L^2-Alexander torsion for 3-manifolds. Invent. Math. **207**(3), 981–1030 (2017)
138. Y. Liu, Finite-volume hyperbolic 3-manifolds are almost determined by their finite quotient groups (2020). arXiv:2011.09412
139. D.D. Long, A.W. Reid, Finding fibre faces in finite covers. Math. Res. Lett. **15**(3), 521–524 (2008)
140. W. Lück, L^2-invariants: theory and applications to geometry and K-theory. *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge*. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics], vol. 44 (Springer, Berlin, 2002). xvi+595 pp.
141. W. Lück, Survey on L^2-invariants and 3-manifolds (2021). arXiv:2102.04549
142. C. Manolescu, Grid diagrams in Heegaard Floer theory, in *European Congress of Mathematics* (Eur. Math. Soc., Zürich, 2013), pp. 643–657
143. S. Matsumoto, Topological entropy and Thurston's norm of atoroidal surface bundles over the circle. J. Fac. Sci. Univ. Tokyo Sect. IA Math. **34**(3), 763–778 (1987)
144. S.V. Matveev, Complexity theory of three-dimensional manifolds. Acta Appl. Math. **19**(2), 101–130 (1990)
145. S. Matveev, Algorithmic topology and classification of 3-manifolds, in *Algorithms and Computation in Mathematics*, vol. 9 (Springer, Berlin, 2003), xii+478 pp.
146. C.T. McMullen, Polynomial invariants for fibered 3-manifolds and Teichmüller geodesics for foliations. Ann. Sci. École Norm. Sup. (4) **33**(4), 519–560 (2000)
147. C.T. McMullen, The Alexander polynomial of a 3-manifold and the Thurston norm on cohomology. Ann. Sci. École Norm. Sup. (4) **35**(2), 153–171 (2002)
148. G. Meng, C.H. Taubes, SW = Milnor torsion. Math. Res. Lett. **3**(5), 661–674(5), 661–674 (1996)
149. J. Milnor, Whitehead torsion. Bull. Am. Math. Soc. **72**, 358–426 (1966)
150. T. Morifuji, On a conjecture of Dunfield, Friedl and Jackson. C. R. Math. Acad. Sci. Paris **350**(19–20), 921–924 (2012)
151. T. Morifuji, Representations of knot groups into $SL(2, C)$ and twisted Alexander polynomials, in *Handbook of Group Actions*, ed. by L. Ji, A. Papadopoulos, S.-T. Yau. Adv. Lect. Math. (ALM), vol. I, 31 (Int. Press, Somerville, MA, 2015), pp. 527–576
152. T. Morifuji, A.T. Tran, Twisted Alexander polynomials of 2-bridge knots for parabolic representations. Pacific J. Math. **269**(2), 433–451 (2014)
153. T. Morifuji, A.T. Tran, Twisted Alexander polynomials of hyperbolic links. Publ. Res. Inst. Math. Sci. **53**(3), 395–418 (2017)
154. L. Mosher, Surfaces and branched surfaces transverse to pseudo-Anosov flows on 3-manifolds. J. Differ. Geom. **34**(1), 1–36 (1991)
155. L. Mosher, Dynamical systems and the homology norm of a 3-manifold, I. Efficient intersection of surfaces and flows. Duke Math. J. **65**(3), 449–500 (1992)
156. L. Mosher, Dynamical systems and the homology norm of a 3-manifold, II. Invent. Math. **107**(2), 243–281 (1992)
157. J.W. Morgan, *The Seiberg-Witten Equations and Applications to the Topology of Smooth Four-Manifolds*. Mathematical Notes, vol. 44 (Princeton University Press, Princeton, NJ, 1996)
158. J.W. Morgan, Z. Szabó, C.H. Taubes, A product formula for the Seiberg-Witten invariants and the generalized Thom conjecture. J. Differ. Geom. **44**(4), 706–788 (1996)
159. M. Nagel, Minimal genus in circle bundles over 3-manifolds. J. Topol. **9**(3), 747–773 (2016)
160. K. Nakamura, The complexity of prime 3-manifolds and the first $\mathbb{Z}/2\mathbb{Z}$-cohomology of small rank(2016). arXiv:1712.02607
161. L.I. Nicolaescu, The Reidemeister torsion of 3-manifolds, in *De Gruyter Studies in Mathematics*, vol. 30 (Walter de Gruyter & Co., Berlin, 2003), xiv+249 pp.
162. L.I. Nicolaescu, Seiberg-Witten invariants of rational homology 3-spheres. Commun. Contemp. Math. **6**(6), 833–866 (2004)

163. Y. Ni, Knot Floer homology detects fibred knots. Invent. Math. **170**(3), 577–608 (2007)
164. Y. Ni, Heegaard Floer homology and fibred 3-manifolds. Am. J. Math. **131**(4), 1047–1063 (2009)
165. Y. Ni, Link Floer homology detects the Thurston norm. Geom. Topol. **13**(5), 2991–3019 (2009)
166. Y. Ni, Z. Wu, Heegaard Floer correction terms and rational genus bounds. Adv. Math. **267**, 360–380 (2014)
167. Y. Ni, Z. Wu, Correction terms, $\mathbb{Z}_2$-Thurston norm, and triangulations. Topol. Appl. **194**, 409–426 (2015)
168. U. Oertel, Homology branched surfaces: Thurston's norm on $H_2(M^3)$, in *Low-Dimensional Topology and Kleinian Groups* (Coventry/Durham, 1984). London Math. Soc. Lecture Note Ser., vol. 112 (Cambridge Univ. Press, Cambridge, 1986), pp. 253–272
169. C. Okonek, A. Teleman, 3-dimensional Seiberg-Witten invariants and non-Kählerian geometry. Math. Ann. **312**(2), 261–288 (1998)
170. J.-P. Otal, *Le théorème d'hyperbolisation pour les variétés fibrées de dimension* 3 (French), Astérisque No. 235 (1996), x+159 pp.
171. J.-P. Otal, *Thurston's Hyperbolization of Haken Manifolds*. Surveys in Differential Geometry, vol. III (Cambridge, MA, 1996) (Int. Press, Boston, MA, 1998), pp. 77–194
172. P. Ozsváth, Z. Szabó, The symplectic Thom conjecture. Ann. Math. (2) **151**(1), 93–124 (2000)
173. P. Ozsváth, Z. Szabó, Higher type adjunction inequalities in Seiberg-Witten theory. J. Differ. Geom. **55**(3), 385–440 (2000)
174. P. Ozsváth, Z. Szabó, Absolutely graded Floer homologies and intersection forms for four-manifolds with boundary. Adv. Math. **173**(2), 179–261 (2003)
175. P. Ozsváth, Z. Szabó, Heegaard Floer homology and alternating knots. Geom. Topol. **7**, 225–254 (2003)
176. P. Ozsváth, Z. Szabó, *Heegaard Diagrams and Holomorphic Disks*. Different Faces of Geometry. Int. Math. Ser. (N. Y.), vol. 3 (Kluwer/Plenum, New York, 2004), pp. 301–348
177. P. Ozsváth, Z. Szabó, Holomorphic disks and genus bounds. Geom. Topol. **8**, 311–334 (2004)
178. P. Ozsváth, Z. Szabó, Holomorphic disks and knot invariants. Adv. Math. **186**(1), 58–116 (2004)
179. P. Ozsváth, Z. Szabó, Holomorphic disks and topological invariants for closed three-manifolds. Ann. of Math. (2) **159**(3), 1027–1158 (2004)
180. P. Ozsváth, Z. Szabó, Holomorphic disks and three-manifold invariants: properties and applications. Ann. Math. (2) **159**(3), 1159–1245 (2004)
181. P. Ozsváth, Z. Szabó, *Lectures on Heegaard Floer Homology*. *Floer Homology, Gauge Theory, and Low-Dimensional Topology*. Clay Math. Proc., vol. 5 (Amer. Math. Soc., Providence, RI, 2006), pp. 29–70
182. P. Ozsváth, Z. Szabó, Link Floer homology and the Thurston norm. J. Am. Math. Soc. **21**(3), 671–709 (2008)
183. P. Ozsváth, Z. Szabó, An overview of knot Floer homology, in *Modern Geometry: A Celebration of the Work of Simon Donaldson*. Proc. Sympos. Pure Math., vol. 99 (Amer. Math. Soc., Providence, RI, 2018), pp. 213–249
184. N. Pacheco-Tallaj, K.V. Schreve, G. Nicholas, Thurston norms of tunnel number-one manifolds. J. Knot Theory Ramif. **28**(9), 1950056, 13 pp. (2019)
185. A.V. Pajitnov, Novikov homology, twisted Alexander polynomials, and Thurston cones. Algebra i Analiz **18**(5), 173–209 (2006); reprinted in St. Petersburg Math. J. **18**(5), 809–835 (2007)
186. L. Person, A piecewise linear proof that the singular norm is the Thurston norm. Topology Appl. **51**(3), 269–289 (1993)
187. P. Przytycki, D.T. Wise, Mixed 3-manifolds are virtually special. J. Am. Math. Soc. **31**(2), 319–347 (2018)
188. J.A. Rasmussen, Floer homology of surgeries on two-bridge knots. Algebr. Geom. Topol. **2**, 757–789 (2002)

189. J.A. Rasmussen, Floer homology and knot complements. Thesis (Ph.D.)-Harvard University, 2003, 126 pp.
190. A.W. Reid, Profinite rigidity, in *Proceedings of the International Congress of Mathematicians–Rio de Janeiro 2018*, vol. II, Invited Lectures (World Sci. Publ., Hackensack, NJ, 2018), pp. 1193–1216
191. D. Rolfsen, *Knots and Links*, vol. 7. Mathematics Lecture Series (Publish or Perish, Inc., Berkeley, CA, 1976), ix+439, pp. 55–01
192. M. Sakuma, A survey of the impact of Thurston's work on knot theory, in *The Tradition of Thurston* (Springer, Cham, 2020), pp. 67–160
193. M. Scharlemann, Sutured manifolds and generalized Thurston norms. J. Differ. Geom. **29**(3), 557–614 (1989)
194. Z. Sela, Dehn fillings that reduce Thurston norm. Israel J. Math. **69**(3), 371–378 (1990)
195. J.-P. Serre, *Galois Cohomology*. Translated from the French by Patrick Ion and revised by the author (Springer, Berlin, 1997), x+210 pp.
196. J.-C. Sikorav, Fibered Cohomology Classes in Dimension Three, Twisted Alexander Polynomials and Novikov Homology (2021). arXiv:1911.03251
197. D.L. Stern, Scalar curvature and harmonic maps to S^1 (2019). arXiv:1908.09754
198. S. Strle, Bounds on genus and geometric intersections from cylindrical end moduli spaces. J. Differ. Geom. **65**(3), 469–511 (2003)
199. H. Sun, A transcendental invariant of pseudo-Anosov maps. J. Topol. **8**(3), 711–743 (2015)
200. W.P. Thurston, Some simple examples of symplectic manifolds. Proc. Am. Math. Soc. **55**(2), 467–468 (1976)
201. W.P. Thurston, *The Geometry and Topology of Three-Manifolds*. Notes (1980) (Electronic version 1.1 - March 2002, http://www.msri.org/publications/books/gt3m/)
202. W.P. Thurston, A norm for the homology of 3-manifolds. Mem. Am. Math. Soc. **59**(339), i–vi and 99–130 (1986)
203. J.L. Tollefson, N. Wang, Branched surfaces and Thurston's norm on homology. Proc. Am. Math. Soc. **122**(2), 635–642 (1994)
204. J.L. Tollefson, N. Wang, Taut normal surfaces. Topology **35**(1), 55–75 (1996)
205. V.G. Turaev, The Alexander polynomial of a three-dimensional manifold (Russian). Mat. Sb. (N.S.) **97**(139) (1975), (3)(7), 341–359, 463
206. V.G. Turaev, Euler structures, nonsingular vector fields, and Reidemeister-type torsions (Russian). Izv. Akad. Nauk SSSR Ser. Mat. **53**(3), 607–643, 672 (1989); translation in Math. USSR-Izv. **34**(3), 627–662 (1990)
207. V. Turaev, Torsion invariants of Spin^c-structures on 3-manifolds. Math. Res. Lett. **4**(5), 679–695 (1997)
208. V. Turaev, A combinatorial formulation for the Seiberg-Witten invariants of 3-manifolds. Math. Res. Lett. **5**(5), 583–598 (1998)
209. V. Turaev, *Introduction to Combinatorial Torsions*. Notes taken by Felix Schlenk. Lectures in Mathematics ETH Zürich (Birkhäuser Verlag, Basel, 2001), viii+123 pp.
210. V. Turaev, A norm for the cohomology of 2-complexes. Algebr. Geom. Topol. **2**, 137–155 (2002)
211. V. Turaev, *Torsions of 3-Dimensional Manifolds*. Progress in Mathematics, vol. 208 (Birkhäuser Verlag, Basel, 2002), x+196 pp.
212. V. Turaev, A function on the homology of 3-manifolds. Algebr. Geom. Topol. **7**, 135–156 (2007)
213. J. Ueki, The profinite completions of knot groups determine the Alexander polynomials. Algebr. Geom. Topol. **18**(5), 3013–3030 (2018)

214. S. Vidussi, Norms on the cohomology of a 3-manifold and SW theory. Pacific J. Math. **208**(1), 169–186 (2003)
215. M. Wada, Twisted Alexander polynomial for finitely presentable groups, Topology **33**(2), 241–256 (1994)
216. E. Witten, Monopoles and four-manifolds. Math. Res. Lett. **1**(6), 769–796 (1994)
217. D.T. Wise, *The Structure of Groups with a Quasiconvex Hierarchy*. Ann. of Math. Studies, vol. 396 (Princeton Univ. Press, Princeton, NJ, 2021)

Chapter 6
From Hyperbolic Dehn Filling to Surgeries in Representation Varieties

Georgios Kydonakis

Abstract Hyperbolic Dehn surgery and the bending procedure provide two ways which can be used to describe hyperbolic deformations of a complete hyperbolic structure on a 3-manifold. Moreover, one can obtain examples of non-Haken manifolds without the use of Thurston's Uniformization Theorem. We review these gluing techniques and present a logical continuity between these ideas and gluing methods for Higgs bundles. We demonstrate how one can construct certain model objects in representation varieties $\mathrm{Hom}\left(\pi_1\left(\Sigma\right), G\right)$ for a topological surface Σ and a semisimple Lie group G. Explicit examples are produced in the case of Θ-positive representations lying in the smooth connected components of the $\mathrm{SO}\left(p, p+1\right)$ representation variety.

Keywords Hyperbolic Dehn surgery · Character variety · Higher Teichmüller space · Higgs bundle · Parabolic structure · Elliptic operator

AMS Classification Primary: 53C07; Secondary: 14H60, 58D27

6.1 Introduction

A Dehn surgery on a 3-manifold M containing a link $L \subset S^3$ is a 2-step process involving the removal of an open tubular neighborhood of the link (drilling) and then gluing back a solid torus using a homeomorphism from the boundary of the solid torus to each of the torus boundary components of M (filling). Of particular interest are the many inequivalent ways one can perform the filling step of the operation, thus providing a way to represent certain examples of 3-dimensional manifolds. In fact, the so-called *fundamental theorem of surgery theory* by Lickorish and Wallace

G. Kydonakis (✉)
Alexander von Humboldt-Stiftung, Bonn, Germany

Ruprecht-Karls-Universität Heidelberg, Heidelberg, Germany
e-mail: gkydonakis@mathi.uni-heidelberg.de

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_6

implies that every closed orientable and connected 3-manifold can be obtained by performing a Dehn surgery on a link in a 3-sphere.

William Thurston introduced hyperbolic geometry into this operation, thus opening the way to certain breakthroughs in 3-manifold theory. His *hyperbolic Dehn filling theorem* implies that the complete hyperbolic structure on the interior of a compact 3-manifold with boundary has a space of hyperbolic deformations parameterized by the generalized Dehn filling coefficients describing the metric completion of the ends of the interior. Among the various and deep advances marked by this result, we highlight here the fact that using hyperbolic Dehn surgery theory one can also obtain examples of non-Haken manifolds, whose hyperbolicity cannot be shown by Thurston's Uniformization Theorem for Haken manifolds. In general, such examples of non-Haken manifolds are not easy to construct otherwise. Deformations of hyperbolic cone structures can, moreover, be better understood when viewed through this prism. In the course of proving Thurston's theorem, one shows not only the existence of a 1-parameter family of cone 3-manifold structures, but can also obtain a path of corresponding holonomies in the representation variety $\mathrm{Hom}(\pi_1(M), \mathrm{SL}(2, \mathbb{C}))$.

Deformations of hyperbolic structures on n-manifolds can be also described by the *bending* procedure. This involves the construction of a family of quasiconformal homeomorphisms of the hyperbolic $(n+1)$-space, which is required to converge under some compatibility conditions. In the case of a surface, the embedded totally geodesic hypersurfaces are simple closed curves along which bending is possible.

Hyperbolic Dehn surgery was originally developed in dimension 3. In this chapter we describe a set of similar ideas of surgery techniques in representation varieties $\mathrm{Hom}(\pi_1(M), G)$, where M this time is a closed connected and oriented topological surface of genus $g \geq 2$ and G is a semisimple Lie group. The Teichmüller space, viewed as the moduli space of marked hyperbolic structures on Σ, can be realized as a connected component of the representation variety $\mathrm{Hom}(\pi_1(M), \mathrm{PSL}(2, \mathbb{R}))$. The recently-emerged field of *higher Teichmüller theory* involves the study of certain connected components of the representation varieties $\mathrm{Hom}(\pi_1(M), G)$, which share essential geometric, topological and dynamical properties with the classical Teichmüller space.

We describe here a gluing construction in $\mathrm{Hom}(\pi_1(M), G)$ "in the tradition" of Thurston's hyperbolic Dehn filling procedure. The parameters involved in this construction are the genus of the surface Σ and the holonomy of a surface group representation along the boundary of Σ.

The non-abelian Hodge correspondence referring to a homeomorphism between representation varieties and moduli spaces of Higgs bundles over a Riemann surface (with underlying topological surface Σ as above) allows us to develop a gluing procedure for the corresponding holomorphic objects, and this makes it easier to determine the connected component where these newly constructed model objects lie, due to an explicit computation of appropriate topological invariants that emerge for their holomorphic counterparts. The deformations involved in the construction are rather expressed in terms of appropriate complex gauge transformations on these holomorphic objects.

In this way, one can construct specific models in certain subsets of representation varieties $\mathrm{Hom}(\pi_1(M), G)$, that are hard to be obtained otherwise; in particular, model representations that do not factor as $\rho : \pi_1(\Sigma) \to \mathrm{SL}(2, \mathbb{R}) \to G$. These models can be used in turn to describe their deformations in the representation variety and use them as a means to study open subsets (or connected components) of objects with certain geometric properties. As an example, we study here model Θ-positive representations that exhaust the smooth $p \cdot (2g-2) - 1$ exceptional components of the $\mathrm{SO}(p, p+1)$-character variety for $p > 2$; similar models have been also constructed for the $2g-3$ exceptional components of the $\mathrm{Sp}(4, \mathbb{R})$-character variety.

This comparison of ideas points towards further ways in developing tools to study certain subsets of representation varieties, quantitative aspects of this holomorphic gluing strategy or universal bounds for the rational parameters involved.

6.2 Hyperbolic Dehn Surgery

In this section we review the basic concepts involved in the hyperbolic Dehn surgery operation. Even though the results of the technique summarized here do not directly apply for the case of character varieties that we study next, these provide a motivation and an interesting counterpoint to the fundamental ideas behind these surgery methods.

6.2.1 Dehn Surgery

Dehn surgery is a method that has found profound relevance in 3-manifold topology and knot theory. It provides a way to represent 3-dimensional manifolds using a “drilling and filling” process. First, a solid torus is removed from a 3-manifold (drilling) and then it is re-attached in many inequivalent ways (filling). This two-stage operation was introduced by Max Dehn in Kapitel II of his 1910 article *Über die Topologie des dreidimensionales Raumes* [27] as a method for constructing *Poincaré spaces*, that is, non-simply connected 3-manifolds with the same topology as the 3-sphere. The texts of Boyer [14], Gordon [40], [41], Luecke [75] offer a broad survey on this construction with numerous references for further study.

The basic parameter of the Dehn surgery operation, in particular referring to the filling stage of the operation, is that of a *slope* on a torus; we briefly introduce this next. Let M be an orientable 3-manifold and $T \subset \partial M$, a toral boundary component of M. Denote by K a knot lying in the interior of M and let $N(K) \subset \mathrm{int}(M)$ be a closed tubular neighborhood of K. For a homeomorphism $f : \partial\left(S^1 \times D^2\right) \to T$, consider the identification space $M(T; f) := \left(S^1 \times D^2\right) \cup_f M$ obtained by identifying the points of $\partial\left(S^1 \times D^2\right)$ with their images by f. We shall call

$M(T; f)$, a *Dehn filling* of M along T. A *Dehn surgery* on a knot K is then a filling of the exterior of the knot K, $M_K := M \backslash \operatorname{int}(N(K))$, along $\partial N(K)$.

Note that a filling $M(T; f)$ depends only on the isotopy class of the attaching homeomorphism $f : \partial\left(S^1 \times D^2\right) \to T$. In fact, the dependence of f is much weaker, for, if $C_0 = \{pt\} \times \partial D^2 \subset \partial\left(S^1 \times D^2\right)$, then $M(T; f)$ depends only on the isotopy class of the curve $f(C_0)$ in T.

Definition 6.1 A *slope* on a torus T is defined as the isotopy class of an essential unoriented simple closed curve on T. If K is a knot in a 3-manifold M, then a slope of K is any slope on $\partial N(K)$.

One has the following proposition:

Proposition 6.1 *A Dehn filling of M along a torus $T \subset \partial M$ is determined up to orientation preserving homeomorphism, by a slope on T. Furthermore, any slope on T arises as the slope of a Dehn filling of M.*

The set of slopes on a torus T is parameterized by the set of $\pm$-pairs of primitive homology classes in $H_1(T)$. In particular, for the 3-sphere S^3 with its usual orientation based on the right-hand rule, the set of slopes of knots in S^3 is canonically identified with $\mathbb{Q} \cup \left\{\frac{1}{0}\right\}$; we may thus realize a slope r of a knot K by a fraction $\frac{p}{q} \in \mathbb{Q} \cup \left\{\frac{1}{0}\right\}$.

Definition 6.2 Let K be a knot in S^3. An *integral slope* of K is a slope corresponding to an integer. We will call *integral surgery* a surgery on K whose slope is integral.

One may now consider the problem of existence and uniqueness of a surgery presentation of a given closed connected orientable 3-manifold by surgery on a finite number of knots in S^3. By a set of *surgery data* $(L; r_1, \ldots, r_n)$ we shall mean a link $L = K_1 \cup \cdots \cup K_n$ lying in the interior of a 3-manifold M, together with a slope r_i for each knot K_i. Let $L(r_1, \ldots, r_n)$ denote the manifold obtained by performing the Dehn surgeries prescribed by the surgery data. In the special case when $M = S^3$ and each r_i is an integral slope, the surgery data $(L; r_1, \ldots, r_n)$ is often called a *framed link*.

The following result is known as the *fundamental theorem of surgery theory*; it was proved using different and independent approaches by Lickorish and Wallace:

Theorem 6.1 (Lickorish [74], Wallace [105]) *Let M be a closed connected orientable 3-manifold. There exists a framed link $(L; r_1, \ldots, r_n)$ in S^3 such that M is homeomorphic to $L(r_1, \ldots, r_n)$.*

For the problem of uniqueness of a surgery presentation of a given manifold, Kirby [59] introduced two moves on (integrally) framed links which do not alter the presented manifold; he also proved that two framed links represent manifolds which are orientation preserving homeomorphic if and only if they are related by a finite sequence of these moves, nowadays called *Kirby moves*. This problem was completely analyzed by Rolfsen in [90].

6.2.2 Hyperbolic Dehn Surgery

A breakthrough in 3-manifold theory as well as in knot theory was signified by the introduction by Thurston of hyperbolic geometry into the Dehn surgery operation. Necessary and sufficient conditions for the complete gluing of a hyperbolic 3-manifold were given by Seifert in [94]. The concept of link of a cusp point of a hyperbolic 3-manifold was introduced by Thurston in his seminal 1979 lecture notes [101].

The celebrated *hyperbolic Dehn filling theorem* of Thurston (Theorem 5.9 in [101]) provides a parameterization of a set of hyperbolic deformations of a complete hyperbolic structure on the interior of a compact 3-manifold with boundary; the parameters, called *generalized Dehn filling coefficients*, describe the metric completion of the ends of the interior.

Among the various and deep advances in 3-manifold theory marked by this result, we will highlight here the fact that using hyperbolic Dehn surgery theory one can also obtain examples of non-Haken manifolds, whose hyperbolicity cannot be shown by Thurston's Uniformization Theorem for Haken manifolds; in fact, the proof of Thurston's theorem does not depend on uniformization. Deformations of hyperbolic cone structures can, moreover, be better understood when viewed through this prism. Another important aspect to be stressed next is the role the generalized Dehn filling coefficients play in the perception of the spaces of hyperbolic deformations parameterized by these coefficients.

The Theorem was first proven in Thurston's notes [101] in the manifold case and has later been extended in the case of orbifolds by Dunbar and Meyerhoff [30]. A detailed review of the proof in both cases can be found in Appendix B of [13] using (in the manifold case) an argument of Zhou [111]. We will follow next the approach of [13] for our purposes.

Let M be a compact 3-manifold with boundary $\partial M = T_1^2 \cup \cdots \cup T_k^2$, a non-empty union of tori, whose interior $\operatorname{int}(M)$ is complete hyperbolic with finite volume. For each boundary component T_j^2 of M, with $j = 1, \ldots, k$, fix two oriented simple closed curves μ_j and λ_j generating the fundamental group $\pi_1\left(T_j^2\right)$. The holonomy of μ_j and λ_j can be viewed as affine transformations of $\mathbb{C} = \partial\mathbb{H}^3 \setminus \{\infty\}$ (∞ being a point fixed by μ_j and λ_j). Then, one can introduce holomorphic parameters u_j and v_j as branches of the logarithm of the linear part of the holonomy around μ_j and λ_j respectively. For $U \subset \mathbb{C}^k$ a neighborhood of the origin, associate to each $u \in U$ a point $\rho_u \in \mathcal{X}(M) = \operatorname{Hom}(\pi_1(M), \mathrm{SL}(2,\mathbb{C}))//\mathrm{SL}(2,\mathbb{C})$ in the $\mathrm{SL}(2,\mathbb{C})$-character variety; this can be done by considering an analytic section

$$s : V \subset \mathcal{X}(M) \to \operatorname{Hom}(\pi_1(M), \mathrm{SL}(2,\mathbb{C})),$$

such that $s(\chi_0) = \rho_0$, where ρ_0 is a lift of the holonomy representation of int (M) and $\chi_0 \in \mathcal{X}(M)$ its character. Then, one has the following important lemma:

Lemma 6.1 (Lemma B.1.6 in [13]) *For $j = 1, \ldots, k$, there is an analytic map $A_j : U \to SL(2, \mathbb{C})$ such that for every $u \in U$:*

$$\rho_u\left(\mu_j\right) = \varepsilon_j A_j(u) \begin{pmatrix} e^{u_j/2} & 1 \\ 0 & e^{-u_j/2} \end{pmatrix} A_j(u)^{-1}, \qquad \text{with } \varepsilon_j = \pm 1,$$

while the commutativity between λ_j and μ_j implies the following:

Lemma 6.2 (Lemma B.1.7 in [13]) *There exist unique analytic functions $v_j, \tau_j : U \to \mathbb{C}$ such that $v_j(0) = 0$ and, for every $u \in U$,*

$$\rho_u\left(\lambda_j\right) = \pm A_j(u) \begin{pmatrix} e^{v_j(u)/2} & \tau_j(u) \\ 0 & e^{-v_j(u)/2} \end{pmatrix} A_j(u)^{-1}.$$

In addition:

1. $\tau_j(0) \in \mathbb{C} - \mathbb{R}$;
2. $\sinh\left(v_j/2\right) = \tau_j \sinh\left(u_j/2\right)$;
3. *v_j is odd in u_j and even in u_l, for $l \neq j$;*
4. $v_j(u) = u_j\left(\tau_j(u) + O\left(|u|^2\right)\right)$.

We are finally set to define the generalized Dehn filling coefficients:

Definition 6.3 (Thurston [101]) For $u \in U$ we define the *generalized Dehn filling coefficients* of the j-th cusp $\left(p_j, q_j\right) \in \mathbb{R}^2 \cup \{\infty\} \cong S^2$ by the formula

$$\begin{cases} \left(p_j, q_j\right) = \infty, & \text{if } u_j = 0 \\ p_j u_j + q_j v_j = 2\pi\sqrt{-1} & \text{if } u_j \neq 0. \end{cases}$$

These coefficients are well-defined and the map

$$U \to S^2 \times \cdots \times S^2$$
$$u \mapsto \left(\left(p_1, q_1\right), \ldots, \left(p_k, q_k\right)\right)$$

defines a homeomorphism between U and a neighborhood of $\{\infty, \ldots, \infty\}$.

Remark 6.1 If $p_j, q_j \in \mathbb{Z}$ are coprime, then the completion at the j-th torus is a non-singular hyperbolic 3-manifold, which topologically is the Dehn filling with surgery meridian $p_j\mu_j + q_j\lambda_j$. One may also perform (p, q)-Dehn surgery also when p and q are not necessarily coprime integers; this refers to *orbifold Dehn surgery*, as in [30]. For instance, $(p, 0)$-Dehn surgery on a knot $K \subset S^3$ provides an orbifold with base S^3 and singular set the knot K with cone angle $2\pi/p$.

The statement of the theorem is the following:

Theorem 6.2 (Hyperbolic Dehn Filling Theorem, Thurston [101]) *Let M be a compact 3-manifold with boundary $\partial M = T_1^2 \cup \cdots \cup T_k^2$, a non-empty union of tori, whose interior int(M) is complete hyperbolic with finite volume. There exists a neighborhood of $\{\infty, \ldots \infty\}$ in $S^2 \times \cdots \times S^2$, such that the complete hyperbolic structure on int(M) has a space of hyperbolic deformations parameterized by the generalized Dehn filling coefficients in this neighborhood.*

The first major step in the proof involves the construction of the algebraic deformation of the holonomies around each boundary component of the manifold M. The second step is to associate generalized Dehn filling coefficients to the aforementioned deformation. The third and final step in the proof involves the construction of the developing maps with the given holonomies. In particular, let $D_0 : \widetilde{\mathrm{int}\,(M)} \to \mathbb{H}^3$ be the developing map for the complete structure on int(M) with holonomy ρ_0. Then, for each $u \in U$, there is a developing map $D_u : \widetilde{\mathrm{int}\,(M)} \to \mathbb{H}^3$ with holonomy ρ_u, such that the completion of int(M) is given by the generalized Dehn filling coefficients of u.

We remark here that the family of maps $\{D_u\}_{u \in U}$ is continuous in u in the compact C^1-topology and that the result above shows not only the existence of a 1-parameter family of cone 3-manifold structures, but also gives a path of corresponding holonomies in the representation variety Hom$(\pi_1(M), \mathrm{SL}(2, \mathbb{C}))$.

6.2.3 Haken Manifolds and Thurston's Uniformization

The notion of *Haken manifold* involves a large class of closed 3-manifolds and play an important role in the study of the topology of 3-manifolds. These were introduced by Wolfgang Haken [48] as a class of compact irreducible 3-manifolds containing incompressible surfaces, for which he showed in [49] that they admit a hierarchy to a union of 3-balls by cutting along essential embedded surfaces. This property allows one to produce certain statements for Haken manifolds using an induction process. Let us next state these definitions more rigorously:

Definition 6.4 Let M be a 3-manifold. A properly embedded surface $\Sigma \subset M$ is called *incompressible* if the map between fundamental groups $\pi_1(\Sigma) \to \pi_1(M)$ is injective. Otherwise, the surface is called *compressible*. A torus in an irreducible 3-manifold is compressible if and only if it bounds a solid torus.

Definition 6.5 A compact orientable 3-manifold M is called a *Haken manifold* if it is irreducible and contains an orientable incompressible surface $\Sigma \subset M$.

In [49], Haken associated a notion of complexity to a Haken manifold, which decreases when one cuts the Haken manifold along an incompressible surface; this can be iterated in order to reduce the complexity until we obtain 3-balls. This approach was a key ingredient in the proof of the Waldhausen theorem showing that closed Haken manifolds are topologically characterized by their fundamental groups:

Theorem 6.3 (Waldhausen, Corollary 6.5 in [103]) *Let M and M' be two Haken manifolds and let $\pi_1(M) \to \pi_1(M')$ be an isomorphism between their fundamental groups. Then M and M' are homeomorphic.*

An algorithm to determine whether a 3-manifold is Haken was given by Jaco and Oertel [56]. Thurston's studies of various examples of 3-manifolds admitting complete hyperbolic metrics lead to his proof of a "uniformization theorem" satisfied by this large class of Haken manifolds:

Theorem 6.4 (Uniformization Theorem for Haken Manifolds, Thurston [101]) *Any atoroidal Haken manifold M admits a hyperbolic structure. By atoroidal here is meant that any embedded incompressible torus is boundary parallel, that is, it can be isotoped into a boundary component of M.*

Thurston's proof uses the hierarchy property of Haken manifolds. By the Waldhausen theorem, (a Haken manifold) M can be decomposed into a finite sum of closed balls B^3 by incompressible surfaces; in other words, there exists a sequence of manifolds with boundary

$$M \mapsto M_1 \mapsto \ldots \mapsto B^3 \cup \ldots \cup B^3.$$

Then, starting with hyperbolic structures on the balls B^3 we may get a hyperbolic structure by gluing at each step in this sequence from these balls back to M. A full proof of this theorem was never published by Thurston; fairly detailed outlines of the proof can be found in the articles by Morgan [81] or Wall [104]. It also follows from Perelman's proof of the more general geometrization conjecture of Thurston constructing the Ricci flow with surgeries on 3-manifolds [84]; see also [7], [82].

The geometrization conjecture evolved from Thurston's considerations that a similar uniformization theorem as for Haken manifolds should hold for all closed 3-manifolds. An important fact considered was that non-Haken manifolds do not contain incompressible surfaces, thus it is impossible to decompose those into simpler pieces. One way by which Thurston proved that non-Haken atoroidal 3-manifolds can be equipped with a hyperbolic structure was by deforming the structure of a cone manifold by increasing its cone angle.

Furthermore, using hyperbolic Dehn surgery it is possible to obtain *non-Haken manifolds, whose hyperbolicity cannot be shown by the uniformization theorem*. Such examples are not easy to construct otherwise; see Reid [89] for explicit examples of non-Haken hyperbolic 3-manifolds with a finite cover which fibers over the circle. Moreover, deformations of hyperbolic structures can be described more concretely using the framework of hyperbolic Dehn surgery.

In [53] Hodgson and Kerckhoff established a universal upper bound on the number of non-hyperbolic Dehn surgeries per boundary torus, thus giving a quantitative version of Thurston's hyperbolic Dehn filling theorem; see also the later article of Lackenby and Meyerhoff [71] on the maximal number of exceptional

Dehn surgeries, providing a proof to Gordon's conjecture [41] on the number of exceptional slopes. For example, Dehn surgeries on the figure-eight knot produce non-Haken, hyperbolic 3-manifolds except in ten cases. For the exterior of the figure-eight knot in S^3 the exceptional surgeries, that is, the ones which do not result in a hyperbolic structure, are

$$\{(1,0)\,,(0,1)\,,\pm(1,1)\,,\pm(2,1)\,,\pm(3,1)\,,\pm(4,1)\}\,.$$

6.3 Deformations of Hyperbolic Structures by Bending

A deformation method of hyperbolic structures on n-manifolds called *bending* is suggested by the famous "Mickey Mouse" example of Thurston (Example 8.7.3 in [101]). Given a hyperbolic structure on a genus two surface, the structure can be considered to arise from the bending of the surface along a simple closed geodesic by an angle $\frac{\pi}{2}$. If the geodesic is short enough, this will give rise to a quasi-Fuchsian group. In order to extend this idea to n dimensions, the manifold is required to contain a totally geodesic submanifold of codimension one along which the bending can take place, thus defining a deformation. That there are compact hyperbolic n-manifolds with arbitrarily many such submanifolds was shown by Millson in [79].

For $n = 3$, hyperbolic structures of infinite volume are related to Kleinian groups which are discrete subgroups of $\mathrm{PSL}(2,\mathbb{C})$ acting discontinuously on part of S^2. In turn, deformations of Kleinian groups can be studied by analyzing the conformal structures on the components of the boundary of the quotient space; a similar phenomenon occurs in higher dimensions (see the works of Apanasov and Tetenov [2, 3]).

Christos Kourouniotis introduced in [62] a deformation technique of hyperbolic structures on n-manifolds via the construction of a family of quasiconformal homeomorphisms of the hyperbolic $(n+1)$-space. His construction of the bending homeomorphism is similar to the construction by Wolpert in [110] of a homeomorphism giving rise to the Fenchel–Nielsen deformation; cf. also the work of Johnson and Millson [57] for an algebraic version of the bending deformation.

The idea in [62] is to construct a quasi-conformal homeomorphism compatible with a subgroup Γ of G_n step by step, as the infinite product of a sequence of homeomorphisms; this product is required to converge and to be compatible with Γ.

In the case of a surface, the embedded totally geodesic hypersurfaces are simple closed curves along which bending is possible. One could also extend in this case the definition of bending to the case of a geodesic lamination, as for instance in the work of Epstein and Marden [31]. Still in this surface case, Kourouniotis has studied in [63] the possibility of bending quasi-Fuchsian structures. Namely, for a closed surface Σ, the space $\mathcal{QF}(\Sigma)$ of quasi-Fuchsian structures on Σ is a quotient of the space of injective homomorphisms $\rho : \pi_1(\Sigma) \to \mathrm{PSL}(2,\mathbb{C})$ with $\mathrm{Im}\rho = \Gamma$ and $\Sigma \times I \cong \mathbb{H}^3/\Gamma$; Fuchsian points are classes of homomorphisms with image in $\mathrm{PSL}(2,\mathbb{R})$ and correspond to hyperbolic structures on Σ. For a simple closed geodesic $\gamma \subset \Sigma$, there is a 1-parameter family of pairs (f_t, ρ_t), where $f_t : \tilde{\Sigma} \to \mathbb{H}^3$ and $\rho_t : \pi_1(\Sigma) \to \mathrm{PSL}(2, \mathbb{C})$, such that f_t is ρ_t-equivariant, for every $t \geq 0$. Note that for $t = 0$, ρ_0 is Fuchsian and f_0 equivariantly embeds $\tilde{\Sigma}$ as a hyperbolic plane in the hyperbolic 3-space $\mathbb{H}^3$. This deformation is induced by a 1-parameter family of isometries from $\mathrm{PSL}(2, \mathbb{C})$. When the bending parameter t is small enough, then f_t is an embedding and ρ_t is an isomorphism of $\pi_1(\Sigma)$ onto a quasi-Fuchsian subgroup of $\mathrm{PSL}(2, \mathbb{C})$.

In [64], Kourouniotis studies some quantitative aspects of this bending construction, while universal bounds on the bending lamination of a quasi-Fuchsian group, hence of the bending deformation, were obtained by Bridgeman [16, 17].

6.4 Higher Teichmüller Theory

The newly-emerged field of higher Teichmüller theory concerns the study of connected components of character varieties for semisimple real Lie groups that entirely consist of discrete and faithful representations. We summarize here some of the very basic topological and geometric properties of these spaces, as well as a recent unified approach to the subject introduced by Olivier Guichard and Anna Wienhard, which seems to be identifying all the cases when such components emerge.

6.4.1 The Teichmüller Space

Let Σ be a closed connected and oriented topological surface with negative Euler characteristic $\chi(\Sigma) = 2 - 2g < 0$, for g the genus of Σ. The *Teichmüller space* $\mathcal{T}(\Sigma)$ of the surface Σ is defined as the space of marked conformal classes of Riemannian metrics on Σ. The Uniformization Theorem of Riemann–Poincaré–Koebe (see [26] for a complete account) guarantees the existence of a unique hyperbolic metric with constant curvature -1 in each conformal class. The Teichmüller space can be thus identified with the moduli space of marked hyperbolic structures. Moreover, the mapping class group $\mathrm{Mod}(\Sigma)$, that is, the group of all orientation-preserving diffeomorphisms of Σ modulo the ones which are isotopic to the identity, acts naturally on $\mathcal{T}(\Sigma)$ by changing the marking; this action is properly discontinuous and the quotient is the moduli space $\mathcal{M}(\Sigma)$ of Riemann surfaces of topological type given by Σ.

A well-known fact about the Teichmüller space is that it is homeomorphic to $\mathbb{R}^{6g-6}$. There are several ways to see this. One direct way is by parameterizing $\mathcal{T}(\Sigma)$ by Fenchel–Nielsen coordinates—a complete proof may be found in [88], Theorem 9.7.4. Another method is to use Teichmüller's theorem to identify $\mathcal{T}(\Sigma)$ with the unit ball in the vector space $Q(M)$ of holomorphic quadratic differentials on a Riemann surface M homeomorphic to Σ—a detailed proof can be found in [55], Theorem 7.2.1. In fact, $\mathcal{T}(\Sigma)$ can be identified with the entire vector space $Q(M)$ using Hopf differentials of harmonic maps from M to a Riemann surface of topological type given by Σ—see [108] for this approach. An application of the Riemann-Roch theorem finally provides that $\dim_{\mathbb{R}} Q(M) = 6g - 6$, for genus $g \geq 2$; we refer, for instance, to Corollary 5.4.2 in [58] for a proof.

However, what opens the way from the classical Teichmüller theory to what is today called *Higher Teichmüller Theory* is the algebraic realization of the space $\mathcal{T}(\Sigma)$ as a subspace of the moduli space of representations of the fundamental group of Σ into the isometry group of the hyperbolic plane. This algebraic realization is conceived through the holonomy representation of a hyperbolic structure. Indeed, for (M, f) a hyperbolic structure over Σ, the orientation preserving homeomorphism $f : \Sigma \to M$ induces an isomorphism of fundamental groups $f_* : \pi_1(\Sigma) \to \pi_1(M)$ and $\pi_1(M)$ acts as the group of deck transformations by isometries on $\tilde{M} \cong \mathbb{H}^2$. But, since $\mathrm{PSL}(2, \mathbb{R}) \cong \mathrm{Isom}^+\left(\mathbb{H}^2\right)$, the orientation preserving isometries, it follows that this action induces a homomorphism $\rho : \pi_1(\Sigma) \to \mathrm{PSL}(2, \mathbb{R})$ which is well-defined up to conjugation by $\mathrm{PSL}(2, \mathbb{R})$. This homomorphism is called the *holonomy* of the hyperbolic structure (M, f). The *representation variety*

$$\mathcal{R}(\mathrm{PSL}(2, \mathbb{R})) := \mathrm{Hom}(\pi_1(\Sigma), \mathrm{PSL}(2, \mathbb{R}))//\mathrm{PSL}(2, \mathbb{R})$$

is the largest Hausdorff quotient of all group homomorphisms $\rho : \pi_1(\Sigma) \to \mathrm{PSL}(2, \mathbb{R})$ modulo conjugation by $\mathrm{PSL}(2, \mathbb{R})$. Furthermore, representations induced by equivalent hyperbolic structures using the above approach are conjugate by an element in $\mathrm{PSL}(2, \mathbb{R})$ and the converse is true.

On the other hand, Weil in [106] (see also Theorem 6.19 in [87]) proved that the set of discrete such embeddings $\{\pi_1(\Sigma) \hookrightarrow \mathrm{PSL}(2, \mathbb{R})\}$ is open in the quotient space $\mathcal{R}(\mathrm{PSL}(2, \mathbb{R}))$. This open subset is called the *Fricke space* $\mathcal{F}(\Sigma)$ of the topological surface Σ. Fricke spaces first appeared in the work of Fricke and Klein [35] defined in terms of Fuchsian groups (see [6] for an expository account).

The connected components of the representation variety $\mathcal{R}(\mathrm{PSL}(2, \mathbb{R}))$ are distinguished in terms of the Euler class $e(\rho)$ of a representation ρ; such a topological invariant for a representation ρ can be considered in the realm of the Riemann–Hilbert correspondence and the associated flat $\mathrm{PSL}(2, \mathbb{R})$-bundle.

In [39], Goldman showed that this Euler class distinguishes the connected components and takes values in $\mathbb{Z} \cap [\chi(\Sigma), -\chi(\Sigma)]$. In particular, the Fricke space $\mathcal{F}(\Sigma)$ is identified with the component maximizing this characteristic class (consisting of representations that correspond to holonomies of hyperbolic structures on Σ).

To conclude this discussion about the Teichmüller space, the Uniformization Theorem implies that $\mathcal{F}(\Sigma)$ and $\mathcal{T}(\Sigma)$ can be identified, therefore the Teichmüller space is a *connected component* of the representation variety $\mathcal{R}(\mathrm{PSL}(2, \mathbb{R}))$. In fact, it is one of the two connected components entirely consisting of discrete and faithful representations $\rho : \pi_1(\Sigma) \to \mathrm{PSL}(2, \mathbb{R})$; the other such component is $\mathcal{T}\left(\bar{\Sigma}\right)$, that is, the Teichmüller space of the surface $\bar{\Sigma}$ with the opposite orientation.

Since the representation variety can be considered for any reductive Lie group G, it is natural to ask whether there are special connected components of it for *higher rank Lie groups* G than $\mathrm{PSL}(2, \mathbb{R})$, which consist entirely of representations related to significant geometric or dynamical structures on the fixed topological surface. This question leads to the introduction of *higher Teichmüller spaces* as we shall see next.

6.4.2 Higher Teichmüller Spaces

Let Σ be a closed oriented (topological) surface of genus g. The fundamental group of Σ is described by

$$\pi_1(\Sigma) = \left\langle a_1, b_1, \ldots, a_g, b_g \,\middle|\, \prod [a_i, b_i] = 1 \right\rangle,$$

where $[a_i, b_i] = a_i b_i a_i^{-1} b_i^{-1}$ is the commutator. The set of all representations of $\pi_1(\Sigma)$ into a connected reductive real Lie group G, $\mathrm{Hom}(\pi_1(\Sigma), G)$, can be naturally identified with the subset of G^{2g} consisting of $2g$-tuples $(A_1, B_1, \ldots, A_g, B_g)$ satisfying the algebraic equation $\prod [A_i, B_i] = 1$. The group G acts on the space $\mathrm{Hom}(\pi_1(\Sigma), G)$ by conjugation

$$(g \cdot \rho) = g\rho(\gamma) g^{-1},$$

where $g \in G$, $\rho \in \mathrm{Hom}(\pi_1(\Sigma), G)$ and $\gamma \in \pi_1(\Sigma)$, and the restriction of this action to the subspace $\mathrm{Hom}^{\mathrm{red}}(\pi_1(\Sigma), G)$ of reductive representations provides that the orbit space is Hausdorff. Here, by a reductive representation we mean one that composed with the adjoint representation in the Lie algebra of G can be decomposed as a sum of irreducible representations. When G is algebraic, this is equivalent to the Zariski closure of the image of $\pi_1(\Sigma)$ in G being a reductive group. Define the *moduli space of reductive representations of* $\pi_1(\Sigma)$ *into* G to be the orbit space

$$\mathcal{R}(G) = \mathrm{Hom}^{\mathrm{red}}(\pi_1(\Sigma), G)/G.$$

The following theorem of Goldman [38] shows that this space is a real analytic variety and so $\mathcal{R}(G)$ is usually called the *character variety*:

Theorem 6.5 (Goldman [38]) *The moduli space $\mathcal{R}(G)$ has the structure of a real analytic variety, which is algebraic if G is algebraic and is a complex variety if G is complex.*

Higher Teichmüller Theory is concerned with the study of the properties of fundamental group representations lying in certain subsets of the character variety $\mathcal{R}(G)$, for simple real groups G. An abundance of methods from geometry, gauge theory, algebraic geometry and dynamics is used to approach these subsets, many methods of which provided by the non-abelian Hodge theory for the moduli space $\mathcal{R}(G)$. The term *higher Teichmüller space* originates in the work of Vladimir Fock and Alexander Goncharov [33], who developed a more algebro-geometric approach to Lusztig's notion of total positivity in the context of general split real semisimple reductive Lie groups (see [76]) and defined positive representations of the fundamental group $\pi_1(\Sigma)$ into these groups; among establishing significant geometric properties, Fock and Goncharov construct in [33] all positive representations and show that they are faithful, discrete and positive hyperbolic. Today, the term refers to connected components of the character variety in a broader sense:

Definition 6.6 Let Σ be a closed connected oriented topological surface of genus $g \geq 2$ and G a semisimple real Lie group. A *higher Teichmüller space* is a connected component of the character variety $\mathcal{R}(G)$ that entirely consists of faithful representations with discrete image.

Several essential features of higher Teichmüller spaces can be traced back to the ideas and work of Thurston. For instance, Thurston's shear coordinates have been extended in this setting by Fock and Goncharov [33], and are sometimes called *Fock–Goncharov coordinates*; noncommutative coordinates on the spaces of framed and decorated fundamental group representations for a surface with boundary into the group $\mathrm{Sp}(2n, \mathbb{R})$ have been introduced by Alessandrini, Guichard, Rogozinnikov and Wienhard in [1]. Labourie and McShane [70] studied cross ratios and McShane–Mirzakhani identities in the case $G = \mathrm{PSL}(n, \mathbb{R})$ and gave explicit expressions of these generalized identities in terms of a suitable choice of Fock–Goncharov coordinates; see also the work of Vlamis and Yarmola [102] for a generalization of Basmajian's identity for Hitchin representations into $\mathrm{PSL}(n, \mathbb{R})$, as well as the article of Fanoni and Pozzetti [32] for Basmajian-type inequalities for maximal representations $\rho : \pi_1(\Sigma) \to \mathrm{Sp}(2n, \mathbb{R})$. Hitchin and maximal representations, in particular, lie in higher Teichmüller spaces and will be briefly reviewed below. Generalizations of the McShane identities for higher Teichmüller spaces were obtained by Huang and Sun in [54]; these are expressed in terms of simple root lengths, triple ratios and edge functions. Le in [72] gave a definition of a higher lamination in the spirit of Thurston for the space of framed G-local systems over Σ and showed that this coincides with the approach of Fock and Goncharov [33] as the tropical points of a higher Teichmüller space. Another example is the pressure metric for Anosov representations from [18, 19], which can be viewed as a generalization of the Weil–Peterson metric on the Teichmüller space as seen by Thurston. Moreover, generalizations of the Collar Lemma from hyperbolic geometry to Hitchin representations and to maximal representations have been also considered in [73] and [22] respectively (see also [8]).

Examples, however, of higher Teichmüller spaces appeared long before the term was invented. For an adjoint split real semisimple Lie group G, there exists a unique embedding $\pi : \mathrm{SL}(2, \mathbb{R}) \to G$, which is the associated Lie group homomorphism to a principal 3-dimensional subalgebra of $\mathfrak{g}$, Kostant's principal subalgebra $\mathfrak{sl}(2, \mathbb{R}) \subset \mathfrak{g}$ (see [61]). For a fixed discrete embedding $\iota : \pi_1(\Sigma) \to$

$\mathrm{SL}(2,\mathbb{R})$, Nigel Hitchin in [52] showed that the subspace containing $\pi \circ \iota : \pi_1(\Sigma) \to G$ is a connected component and, in fact, topologically trivial of dimension $(2g-2)\dim G$. In the special case when the group is $G = \mathrm{PSL}(2,\mathbb{R})$, this component is the Teichmüller space.

Following the work of Hitchin, it became apparent that the spaces identified, now called *Hitchin components*, include representations with important geometric features. For instance, Labourie introduced in [68] the notion of an *Anosov representation* and used techniques from dynamical systems to prove (among other essential geometric properties) that representations lying inside the Hitchin component for $G = \mathrm{PSL}(n,\mathbb{R})$, $\mathrm{PSp}(2n,\mathbb{R})$ or $\mathrm{PO}(n,n+1)$ are faithful with discrete image; we refer the reader to [18, 43, 44, 69, 70, 73, 85] for subsequent works on the geometric and dynamical properties of representations in the Hitchin components.

The second family of Lie groups G where components of discrete and faithful representations have been detected, is the family of Hermitian Lie groups of non-compact type, that is, the symmetric space associated to G is an irreducible Hermitian symmetric space of non-compact type. In this case, a characteristic number called the *Toledo invariant* of a representation $\rho : \pi_1(\Sigma) \to G$ can be defined as the integer

$$T_\rho := \left\langle \rho^*(\kappa_G), [\Sigma] \right\rangle,$$

where $\rho^*(\kappa_G)$ is the pullback of the Kähler class $\kappa_G \in H_c^2(G,\mathbb{R})$ of G and $[\Sigma] \in H_2(\Sigma,\mathbb{R})$ is the orientation class. The absolute value of the Toledo invariant has an upper bound of Milnor–Wood type

$$\left|T_\rho\right| \leq (2g-2)\,\mathrm{rk}(G) \tag{6.1}$$

and a representation $\rho : \pi_1(\Sigma) \to G$ is called *maximal* when this upper bound is achieved. Subspaces of maximal representations also have interesting geometric and dynamical properties and, in particular, consist entirely of discrete and faithful representations, as seen in [20] and [21].

It is also interesting to note at this point that in the case when the group G is the group $\mathrm{PSL}(2,\mathbb{R})$, the Toledo invariant is actually the Euler class, Inequality (6.1) is the Milnor–Wood inequality for the Euler class and the space of maximal representations in this case is identified with the Teichmüller space, as in [39].

We refer the reader to the survey articles of Wienhard [107] and Pozzetti [86] for a broader presentation of the geometric properties of higher Teichmüller spaces, as well as for an overview of the similarities and differences between these spaces and the classical Teichmüller space.

6.4.3 Θ-Positive Representations

The special connected components introduced for the two families of Lie groups above, namely the adjoint split real semisimple Lie groups and the Hermitian Lie groups on non-compact type share (among many other fundamental properties) a common characterization that relates to the existence of a continuous equivariant map sending positive triples in $\mathbb{RP}^1$ to positive triples in certain flag varieties associated with the Lie group G. This property was identified by Labourie [68], Guichard [43] and Fock–Goncharov [33] in the case of split semisimple real Lie groups, and by Burger–Iozzi–Wienhard [21] for Hermitian Lie groups of non-compact type.

This in turn provided the motivation to propose in [45] that the characterization above in terms of positivity can, in fact, distinguish *all* higher Teichmüller spaces. We next include more details about this general conjectural picture; for complete references the reader is directed to the original article of Guichard and Wienhard [45].

The definition of a Θ-positive structure for a real semisimple Lie group G is a generalization of Lusztig's total positivity condition in [76] and is given in regards to properties of the Lie algebra of parabolic subgroups $P_\Theta < G$ defined by a subset of simple positive roots $\Theta \subset \Delta$. In these terms, let $\mathfrak{u}_\Theta := \sum_{\alpha \in \Sigma_\Theta^+} \mathfrak{g}_\alpha$, for $\Sigma_\Theta^+ = \Sigma^+ \backslash \mathrm{Span}\,(\Delta - \Theta)$, where Σ^+ denotes the set of positive roots, and then the standard parabolic subgroup P_Θ associated to $\Theta \subset \Delta$ is the normalizer in G of $\mathfrak{u}_\Theta$. The group P_Θ is the semidirect product of its unipotent radical $U_\Theta := \exp(\mathfrak{u}_\Theta)$. Consider the Levi subgroup $L_\Theta := P_\Theta \cap P_\Theta^{opp}$, where P_Θ^{opp} is the normalizer in G of $\mathfrak{u}_\Theta^{opp} := \sum_{\alpha \in \Sigma_\Theta^+} \mathfrak{g}_{-\alpha}$. The Levi factor L_Θ acts on $\mathfrak{u}_\Theta$ via the adjoint action. Denote by L_Θ^0 the component of L_Θ containing the identity.

For $\mathfrak{z}_\Theta$, the center of the Lie algebra $\mathfrak{l}_\Theta := \mathrm{Lie}\,(L_\Theta)$, $\mathfrak{u}_\Theta$ can be decomposed into weight spaces

$$\mathfrak{u}_\Theta = \sum_{\beta \in \mathfrak{z}_\Theta^*} \mathfrak{u}_\beta,$$

where $\mathfrak{u}_\beta := \{N \in \mathfrak{u}_\Theta \,|\mathrm{ad}\,(Z)\,N = \beta\,(Z)\,N,\ \text{for every } Z \in \mathfrak{z}_\Theta\}$.

Definition 6.7 (Guichard–Wienhard, Definition 4.2 in [45]) Let G be a semisimple Lie group with finite center and $\Theta \subset \Delta$ a subset of simple roots. The group G admits a Θ-*positive structure* if for all $\beta \in \Theta$, there exists an L_Θ^0-invariant sharp convex cone in $\mathfrak{u}_\beta$.

A central result in [45] provides that the semisimple Lie groups G that can admit a Θ-positive structure are classified as follows:

Theorem 6.6 (Guichard–Wienhard, Theorem 4.3 in [45]) *A semisimple Lie group G admits a Θ-positive structure if and only if the pair (G, Θ) belongs to one of the following four cases:*

1. *G is a split real form and $\Theta = \Delta$.*
2. *G is a Hermitian symmetric Lie group of tube type and $\Theta = \{\alpha_r\}$.*
3. *G is a Lie group locally isomorphic to a group $SO(p, q)$, for $p \neq q$, and $\Theta = \{\alpha_1, \ldots, \alpha_{p-1}\}$.*
4. *G is a real form of the groups F_4, E_6, E_7, E_8 with restricted root system of type F_4, and $\Theta = \{\alpha_1, \alpha_2\}$.*

In order to define the notion of a positive triple in the generalized flag variety G/P_Θ for a semisimple Lie group G with a Θ-positive structure, one needs to introduce the notion of a Θ-positive semigroup. First, associate to Θ a subgroup $W(\Theta)$ of the Weyl group W as follows: The group W is generated by the reflections s_α, for $\alpha \in \Delta$; set $\sigma_\beta = s_\beta$ for all $\beta \in \Theta - \{\beta_\Theta\}$ and define σ_{β_Θ} to be the longest element of the Weyl group $W_{\{\beta_\Theta\}\cup(\Delta-\Theta)}$ of the sub-root system generated by $\{\beta_\Theta\}\cup(\Delta - \Theta)$. Define now the subgroup of W,

$$W(\Theta) = \langle \sigma_\beta \rangle_{\beta\in\Theta}.$$

The group $W(\Theta)$ acts on the weight spaces $\mathfrak{u}_\Theta$, for $\beta \in \operatorname{span}(\Theta)$. Denote by w_Θ^0, the longest element in $W(\Theta)$ and consider a reduced expression $w_\Theta^0 = \sigma_{i_1} \cdots \sigma_{i_l}$. Then, for $c_\beta^0 \subset \mathfrak{u}_\Theta$, the interior of the L_Θ^0-invariant closed convex cone, there is a map for every $\beta \in \Theta$ defined by

$$\begin{aligned} F_{\sigma_{i_1}\cdots\sigma_{i_l}} : c^0_{\beta_{i_1}} \times \cdots \times c^0_{\beta_{i_l}} &\to U_\Theta \\ \left(v_{i_1}, \ldots, v_{i_l}\right) &\mapsto \chi_{\beta_{i_1}}\left(v_{i_1}\right) \cdot \ldots \cdot \chi_{\beta_{i_l}}\left(v_{i_l}\right), \end{aligned}$$

where, for any $\beta \in \Theta$, the map $\chi_\beta : \mathfrak{u}_\Theta \to U_\beta \subset U_\Theta$ with $v \mapsto \exp(v)$ is considered. The Θ-positive semigroup of U_Θ is now defined as follows:

Theorem 6.7 (Guichard–Wienhard, Theorem 4.5 in [45]) *The image $U_\Theta^{>0}$ of the map $F_{\sigma_{i_1}\cdots\sigma_{i_l}}$ defined above is independent of the reduced expression of w_Θ^0.*

One may now define positive triples in the generalized flag variety:

Definition 6.8 Fix E_Θ and F_Θ to be the standard flags in G/P_Θ such that $\operatorname{Stab}_G(F_\Theta) = P_\Theta$ and $\operatorname{Stab}_G(E_\Theta) = P_\Theta^{opp}$. For any $S_\Theta \in G/P_\Theta$ transverse to F_Θ, there exists $u_{S_\Theta} \subset U_\Theta$ such that $S_\Theta = u_{S_\Theta} E_\Theta$. The triple $(E_\Theta, S_\Theta, F_\Theta)$ in the generalized flag variety G/P_Θ will be called Θ-*positive*, if $u_{S_\Theta} \in U_\Theta^{>0}$, for $U_\Theta^{>0}$ the Θ-positive semigroup of U_Θ.

The definition of a Θ-positive fundamental group representation is now the following:

Definition 6.9 (Guichard–Wienhard, Definition 5.3 in [45]) Let Σ be a closed connected and oriented topological surface of genus $g \geq 2$ and let G be a semisimple Lie group admitting a Θ-positive structure. A representation of the fundamental group of Σ into G will be called Θ-*positive*, if there exists a ρ-equivariant positive map $\xi : \partial\pi_1(\Sigma) = \mathbb{RP}^1 \to G/P_\Theta$ sending positive triples in $\mathbb{RP}^1$ to Θ-positive triples in G/P_Θ.

In their recent article [47], Guichard, Labourie and Wienhard show that Θ-positive representations are Θ-Anosov, thus discrete and faithful, and that, in fact, for the four families of semisimple Lie groups G listed in Theorem 6.6 above, there are higher Teichmüller spaces in the character variety:

Theorem 6.8 (Guichard–Labourie–Wienhard, Theorem A in [47]) *Let G be a semisimple Lie group that admits a Θ-positive structure. Then there exists a connected component of the representation variety $\mathcal{R}(G)$ that consists solely of discrete and faithful representations.*

6.5 Non-abelian Hodge Theory

A major contribution to the various methods available in order to study higher Teichmüller spaces involves fixing a complex structure J on the topological surface Σ, thus transforming Σ into a Riemann surface $X = (\Sigma, J)$, therefore opening the way to holomorphic techniques and the theory of *Higgs bundles*, as initiated by Nigel Hitchin in his article *The self duality equations on a Riemann surface* published in 1987 [51]. The non-abelian Hodge theory correspondence provides a real-analytic isomorphism between the character variety $\mathcal{R}(G)$ and the moduli space of polystable G-Higgs bundles, which we briefly introduce next.

6.5.1 Moduli Spaces of G-Higgs Bundles

Let X be a compact Riemann surface and let G be a real reductive group. The latter involves considering *Cartan data* (G, H, θ, B), where $H \subset G$ is a maximal compact subgroup, $\theta : \mathfrak{g} \to \mathfrak{g}$ is a Cartan involution and B is a non-degenerate bilinear form on $\mathfrak{g}$ which is $\mathrm{Ad}(G)$-invariant and θ-invariant. The Cartan involution θ gives a decomposition (called the *Cartan decomposition*)

$$\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{m}$$

into its ± 1-eigenspaces, where $\mathfrak{h}$ is the Lie algebra of H.

Let $H^{\mathbb{C}}$ be the complexification of H and let $\mathfrak{g}^{\mathbb{C}} = \mathfrak{h}^{\mathbb{C}} \oplus \mathfrak{m}^{\mathbb{C}}$ be the complexification of the Cartan decomposition. The adjoint action of G on $\mathfrak{g}$ restricts to give a representation (the isotropy representation) of H on $\mathfrak{m}$. This is independent of the choice of Cartan decomposition, since any two Cartan decompositions of G are related by a conjugation using also that $[\mathfrak{h}, \mathfrak{m}] \subseteq \mathfrak{m}$. The action of H extends to a linear holomorphic action of $H^{\mathbb{C}}$ on $\mathfrak{m}^{\mathbb{C}}$, thus providing the complexified isotropy representation $\iota : H^{\mathbb{C}} \to \mathrm{GL}(\mathfrak{m}^{\mathbb{C}})$. This introduces the following definition:

Definition 6.10 Let $K \cong T^*X$ be the canonical line bundle over a compact Riemann surface X. A *G-Higgs bundle* is a pair (E, φ) where

- E is a principal holomorphic $H^{\mathbb{C}}$-bundle over X and
- φ is a holomorphic section of the vector bundle $E\left(\mathfrak{m}^{\mathbb{C}}\right) \otimes K = \left(E \times_{\iota} \mathfrak{m}^{\mathbb{C}}\right) \otimes K$.

The section φ is called the *Higgs field*. Two G-Higgs bundles (E, φ) and $\left(E', \varphi'\right)$ are said to be *isomorphic* if there is a principal bundle isomorphism $E \cong E'$ which takes φ to φ' under the induced isomorphism $E\left(\mathfrak{m}^{\mathbb{C}}\right) \cong E'\left(\mathfrak{m}^{\mathbb{C}}\right)$.

To define a moduli space of G-Higgs bundles we need to consider a notion of semistability, stability and polystability. These notions are defined in terms of an antidominant character for a parabolic subgroup $P \subseteq H^{\mathbb{C}}$ and a holomorphic reduction σ of the structure group of the bundle E from $H^{\mathbb{C}}$ to P (see [37] for the precise definitions).

When the group G is connected, principal $H^{\mathbb{C}}$-bundles E are topologically classified by a characteristic class $c(E) \in H^2\left(X, \pi_1\left(H^{\mathbb{C}}\right)\right) \cong \pi_1\left(H^{\mathbb{C}}\right) \cong \pi_1(H) \cong \pi_1(G)$.

Definition 6.11 For a fixed class $d \in \pi_1(G)$, the *moduli space of polystable G-Higgs bundles* of fixed topological class d with respect to the group of complex gauge transformations is defined as the set of isomorphism classes of polystable G-Higgs bundles (E, φ) such that $c(E) = d$. We will denote this set by $\mathcal{M}_d(G)$.

Using the general GIT constructions of Schmitt for decorated principal bundles in the case of a real form of a complex reductive algebraic Lie group, it is shown that the moduli space $\mathcal{M}_d(G)$ is an algebraic variety. The expected dimension of the moduli space of G-Higgs bundles is $(g-1)\dim G^{\mathbb{C}}$, in the case when G is a connected semisimple real Lie group; see [37, 92, 93] for details.

6.5.2 G-Hitchin Equations

Let (E, φ) be a G-Higgs bundle over a compact Riemann surface X. By a slight abuse of notation we shall denote the underlying smooth objects of E and φ by the same symbols. The Higgs field can be thus viewed as a $(1,0)$-form $\varphi \in \Omega^{1,0}\left(E\left(\mathfrak{m}^{\mathbb{C}}\right)\right)$. Given a reduction h of structure group to H in the smooth $H^{\mathbb{C}}$-bundle E, we denote by F_h the curvature of the unique connection compatible with

h and the holomorphic structure on E. Let $\tau_h : \Omega^{1,0}\left(E\left(\mathfrak{g}^{\mathbb{C}}\right)\right) \to \Omega^{0,1}\left(E\left(\mathfrak{g}^{\mathbb{C}}\right)\right)$ be defined by the compact conjugation of $\mathfrak{g}^{\mathbb{C}}$ which is given fiberwise by the reduction h, combined with complex conjugation on complex 1-forms. The next theorem was proved in [37] for an arbitrary reductive real Lie group G.

Theorem 6.9 (Hitchin–Kobayashi Correspondence, Theorem 3.21 in [37]) *There exists a reduction h of the structure group of E from $H^{\mathbb{C}}$ to H satisfying the Hitchin equation*

$$F_h - [\varphi, \tau_h(\varphi)] = 0$$

if and only if (E, φ) is polystable.

From the point of view of moduli spaces it is convenient to fix a C^∞ principal H-bundle $\mathbf{E}_H$ with fixed topological class $d \in \pi_1(H)$ and study the moduli space of solutions to Hitchin's equations for a pair (A, φ) consisting of an H-connection A and $\varphi \in \Omega^{1,0}\left(X, \mathbf{E}_H\left(\mathfrak{m}^{\mathbb{C}}\right)\right)$ with

$$\begin{aligned} F_A - [\varphi, \tau(\varphi)] &= 0 \qquad (*) \\ \bar{\partial}_A \varphi &= 0 \end{aligned}$$

where d_A is the covariant derivative associated with A and $\bar{\partial}_A$ is the (0, 1)-part of d_A, defining the holomorphic structure on $\mathbf{E}_H$. Also, τ is defined by the fixed reduction of structure group $\mathbf{E}_H \hookrightarrow \mathbf{E}_H\left(H^{\mathbb{C}}\right)$. The gauge group $\mathcal{G}_H$ of $\mathbf{E}_H$ acts on the space of solutions by conjugation and the moduli space of solutions is defined by

$$\mathcal{M}_d^{\text{gauge}}(G) := \{(A, \varphi) \text{ satisfying equations } (*)\}/\mathcal{G}_H\ .$$

Now, Theorem 6.9 implies that there is a homeomorphism

$$\mathcal{M}_d(G) \cong \mathcal{M}_d^{\text{gauge}}(G)\,.$$

Using the one-to-one correspondence between H-connections on $\mathbf{E}_H$ and $\bar{\partial}$-operators on $\mathbf{E}_{H^{\mathbb{C}}}$, the homeomorphism in the above theorem can be interpreted as saying that in the $\mathcal{G}_H^{\mathbb{C}}$-orbit of a polystable G-Higgs bundle $\left(\bar{\partial}_{E_0}, \varphi_0\right)$ we can find another Higgs bundle $\left(\bar{\partial}_E, \varphi\right)$ whose corresponding pair (d_A, φ) satisfies the equation $F_A - [\varphi, \tau(\varphi)] = 0$, and this is unique up to H-gauge transformations.

6.5.3 The Non-abelian Hodge Correspondence

We can assign a topological invariant to a representation $\rho \in \mathcal{R}(G)$ by considering its corresponding flat G-bundle on Σ defined as $E_\rho = \tilde{\Sigma} \times_\rho G$. Here $\tilde{\Sigma} \to \Sigma$ is the universal cover and $\pi_1(\Sigma)$ acts on G via ρ. A topological invariant is then given by the characteristic class $c(\rho) := c(E_\rho) \in \pi_1(G) \simeq \pi_1(H)$, for $H \subseteq G$ a maximal compact subgroup of G. For a fixed $d \in \pi_1(G)$ the moduli space of reductive representations with fixed topological invariant d is now defined as the subvariety

$$\mathcal{R}_d(G) := \{[\rho] \in \mathcal{R}(G) \,|\, c(\rho) = d\}.$$

A reductive fundamental group representation corresponds to a solution to the Hitchin equations. This is seen using that any solution h to Hitchin's equations defines a flat reductive G-connection

$$D = D_h + \varphi - \tau(\varphi), \tag{6.2}$$

where D_h is the unique H-connection on E compatible with its holomorphic structure. Conversely, given a flat reductive connection D on a G-bundle E_G, there exists a harmonic metric, in other words, a reduction of structure group to $H \subset G$ corresponding to a harmonic section of $E_G/H \to X$. This reduction produces a solution to Hitchin's equations such that Eq. (6.2) holds.

In summary, equipping the surface Σ with a complex structure J, a reductive representation of $\pi_1(\Sigma)$ into G corresponds to a polystable G-Higgs bundle over the Riemann surface $X = (\Sigma, J)$; this is the content of *non-abelian Hodge correspondence*; its proof is based on combined work by Hitchin [51], Simpson [96], [98], Donaldson [28] and Corlette [25]:

Theorem 6.10 (Non-abelian Hodge Correspondence) *Let G be a connected semisimple real Lie group with maximal compact subgroup $H \subseteq G$ and let $d \in \pi_1(G) \simeq \pi_1(H)$. Then there exists a homeomorphism*

$$\mathcal{R}_d(G) \cong \mathcal{M}_d(G).$$

The introduction of holomorphic techniques via the non-abelian Hodge correspondence allows the description of a theory of higher Teichmüller spaces from the Higgs bundle point of view. In [15], Bradlow, Collier, García-Prada, Gothen and Oliviera obtain a parameterization of special components of the moduli space of Higgs bundles on a compact Riemann surface using the decomposition data for a complex simple Lie algebra $\mathfrak{g}$. The possible decompositions of $\mathfrak{g}$ are defined by a newly introduced class of $\mathfrak{sl}(2,\mathbb{R})$-triples, and the classification of these triples is shown to be in bijection with the classification of the Θ-positive structures of Guichard and Wienhard (Theorem 6.6). We refer to [15] for the precise statements; see also the survey article of García-Prada [36] for a broader description of the results for higher Teichmüller spaces that can be obtained using the theory of Higgs bundles.

6.6 Surgeries in Representation Varieties-General Theory

We next describe a gluing construction for points of the moduli spaces appearing in the non-abelian Hodge correspondence. In particular, this technique can be used to obtain specific model objects of the moduli spaces which are hard to be constructed otherwise and can be used to improve our understanding of the geometric properties of the subsets of the character variety they live in.

6.6.1 Topological Gluing Construction

For a closed oriented surface Σ of genus g, let $\Sigma = \Sigma_l \cup_\gamma \Sigma_r$ be a decomposition of Σ along one simple closed oriented separating geodesic γ into two subsurfaces, say Σ_l and Σ_r. Let now $\rho_l : \pi_1(\Sigma) \to G$ and $\rho_r : \pi_1(\Sigma) \to G$ be two representations into a semisimple Lie group G.

One could amalgamate the restriction of ρ_l to Σ_l with the restriction of ρ_r to Σ_r, however the holonomies of those along γ do not have to agree a priori. If the holonomies do agree (possibly after applying a deformation of at least one of the two representations for the holonomies to match up), then one can introduce new representations by gluing with a use of the van Kampen theorem at the level of topological surfaces, as follows.

Definition 6.12 A *hybrid representation* is defined as the amalgamated representation

$$\rho := \rho_l \left|_{\pi_1(\Sigma_l)} * \rho_r \right|_{\pi_1(\Sigma_r)} : \pi_1(\Sigma) \simeq \pi_1(\Sigma_l) *_{\langle\gamma\rangle} \pi_1(\Sigma_r) \to G.$$

Remark 6.2 The assumption that the holonomies agree over the boundary is crucial. In §3.3.1 of [46], Guichard and Wienhard provide an explicit example of hybrid representations in the case when the group is the symplectic group $\mathrm{Sp}(4, \mathbb{R})$. Special attention is paid there in order to establish this assumption via an appropriate deformation argument.

The above construction/definition can be generalized to the case when the subsurfaces Σ_l and Σ_r are not necessarily connected. For Σ as earlier, let $\Sigma_1 \subset \Sigma$ denote a subsurface with Euler characteristic $\chi(\Sigma_1) \leq -1$. The (nonempty) boundary of Σ_1 is a union of disjoint circles

$$\partial \Sigma_1 = \coprod_{d \in \pi_0(\partial \Sigma_1)} \gamma_d.$$

The circles γ_d are oriented so that for each d, the surface Σ_1 lies on the left of γ_d. Now, write

$$\Sigma \backslash \partial \Sigma_1 = \bigcup_{c \in \pi_0(\Sigma \backslash \partial \Sigma_1)} \Sigma_c.$$

Then, for any $d \in \pi_0(\partial \Sigma_1)$, the curve γ_d bounds exactly two connected components of $\Sigma \backslash \partial \Sigma_1$, namely, one is included in Σ_1 and denoted by $\Sigma_{l(d)}$ with $l(d) \in \pi_0(\Sigma_1)$, while the other is included in the complement of Σ_1 and is denoted by $\Sigma_{r(d)}$ with $r(d) \in \pi_0(\Sigma \backslash \Sigma_1)$. In this way, we have $l(d), r(d) \in \pi_0(\Sigma \backslash \partial \Sigma_1)$, but it can be that $l(d) = l(d')$ or that $r(d) = r(d')$, for $d \neq d'$.

Assume now that the graph with vertex set $\pi_0(\Sigma \backslash \Sigma_1)$ and edges given by the pairs $\{l(d), r(d)\}_{d \in \pi_0(\partial \Sigma_1)}$ is a tree. This allows us to apply a generalized van Kampen theorem argument and write the fundamental group $\pi_1(\Sigma)$ as the amalgamated product of the groups $\pi_1(\Sigma_c)$, for all $c \in \pi_0(\Sigma \backslash \partial \Sigma_1)$ over the groups $\pi_1(\gamma_d)$, for all $d \in \pi_0(\partial \Sigma_1)$.

Pick a family of representations $\{\rho_c : \pi_1(\Sigma_c) \to G\}_{c \in \pi_0(\Sigma \backslash \partial \Sigma_1)}$ subordinate to the following condition: there exist elements $g_c \in G$ for each $c \in \pi_0(\Sigma \backslash \partial \Sigma_1)$, such that for any $d \in \pi_0(\partial \Sigma_1)$ it holds that

$$g_{l(d)} \rho_{l(d)}(\gamma_d) g_{l(d)}^{-1} = g_{r(d)} \rho_{r(d)}(\gamma_d) g_{r(d)}^{-1}.$$

Then one may construct a hybrid representation $\rho : \pi_1(\Sigma) \to G$ by amalgamating the representations $g_c \rho_c g_c^{-1}$, for each $c \in \pi_0(\Sigma \backslash \partial \Sigma_1)$. An explicit example of amalgamation of a family of representations that satisfy the condition above is provided in §3.3.2 of [46] in the case when $G = \mathrm{Sp}(4, \mathbb{R})$.

6.6.2 Gluing in Exceptional Components of the Moduli Space

Motivated by the amalgamation construction for representations and in the realm of the non-abelian Hodge correspondence, one may seek for an analogous gluing construction from a holomorphic point of view. The benefit from establishing this method in the Higgs bundle moduli space is that it is easier to compute the Higgs bundle invariants for any models constructed in order to identify in which connected component these new objects lie. Indeed, for the cases when the Lie group is the group $\mathrm{Sp}(4, \mathbb{R})$ or $\mathrm{SO}(p, p+1)$ the moduli space has a number of exceptional components in terms of their topological and geometric properties; these exceptional components do, in fact, fall in the class of higher Teichmüller spaces. It is for such components that a gluing construction for Higgs bundles can provide good models that are not easily obtained otherwise, thus allowing us to study more closely the components themselves. Examples of models in the case of the group $\mathrm{Sp}(4, \mathbb{R})$ were obtained in [66] (see also [65]), while for $G = \mathrm{SO}(p, p+1)$ we will demonstrate some examples in Sect. 6.7 later on.

6.6.2.1 Parabolic GL $(n, \mathbb{C})$-Higgs Bundles

Remember that the amalgamation method involved fundamental group representations defined over a surface with boundary. The appropriate analog to a surface group representation into a reductive Lie group G for a surface with boundary is a *parabolic G-Higgs bundle* over a Riemann surface with a divisor. This involves an extra layer of structure encoded by a weighted filtration on each fiber of the bundle over a collection of finitely many distinct points of the surface. We include next basic definitions for a parabolic GL $(n, \mathbb{C})$-Higgs bundle; concrete examples of such pairs will be studied later on in Sect. 6.7.

Parabolic vector bundles over Riemann surfaces with marked points were introduced by Conjeeveram S. Seshadri in [95] and similar to the Narasimhan–Seshadri correspondence, there is an analogous correspondence between stable parabolic bundles and unitary representations of the fundamental group of the punctured surface with fixed holonomy class around each puncture [78]. Later on, Carlos Simpson in [97] proved a non-abelian Hodge correspondence over a *non-compact* curve.

Definition 6.13 Let X be a closed, connected, smooth Riemann surface of genus $g \geq 2$ with s-many marked points $x_1, \ldots, x_s$ and let a divisor $D = \{x_1, \ldots, x_s\}$. A *parabolic vector bundle* E over X is a holomorphic vector bundle $E \to X$ of rank n with *parabolic structure* at each $x \in D$ (*weighted flag* on each fiber E_x):

$$E_x = E_{x,1} \supset E_{x,2} \supset \ldots \supset E_{x,r(x)+1} = \{0\}$$
$$0 \leq \alpha_1(x) < \ldots < \alpha_{r(x)}(x) < 1.$$

The real numbers $\alpha_i(x) \in [0, 1)$ for $1 \leq i \leq r(x)$ are called the *weights* of the subspaces E_x and we usually write (E, α) to denote a parabolic vector bundle equipped with a parabolic structure determined by a system of weights $\alpha(x) = \left(\alpha_1(x), \ldots, \alpha_{r(x)}(x)\right)$ at each $x \in D$; whenever the system of weights is not discussed in the context, we will be omitting the notation α to ease exposition. Moreover, let $k_i(x) = \dim\left(E_{x,i}/E_{x,i+1}\right)$ denote the *multiplicity* of the weight $\alpha_i(x)$ and notice that $\sum_i k_i(x) = n$. A weighted flag shall be called *full*, if $k_i(x) = 1$ for every $1 \leq i \leq r(x)$ and every $x \in D$.

The *parabolic degree* and *parabolic slope* of a vector bundle equipped with a parabolic structure are the real numbers

$$\operatorname{par}\deg(E) = \deg E + \sum_{x \in D} \sum_{i=1}^{r(x)} k_i(x)\, \alpha_i(x),$$

$$\operatorname{par}\mu\,(\mathrm{E}) = \frac{\operatorname{pardeg}(E)}{\operatorname{rk}(E)}.$$

Definition 6.14 Let K be the canonical bundle over X and E a parabolic vector bundle. The bundle morphism $\Phi : E \to E \otimes K(D)$ will be called a *parabolic Higgs field* if it preserves the parabolic structure at each point $x \in D$:

$$\Phi\left|_x \left(E_{x,i}\right) \subset E_{x,i} \otimes K(D)\right|_x .$$

In particular, we call Φ *strongly parabolic* if

$$\Phi\left|_x \left(E_{x,i}\right) \subset E_{x,i+1} \otimes K(D)\right|_x ,$$

that is, $\Phi \in H^0(X, \mathrm{End}(E) \otimes K(D))$ is an element with simple poles along the divisor D, whose residue at $x \in D$ is nilpotent with respect to the filtration.

After these considerations we define parabolic Higgs bundles as follows.

Definition 6.15 Let K be the canonical bundle over X and E a parabolic vector bundle over X. A *parabolic Higgs bundle* is a pair (E, Φ), where E is a parabolic vector bundle and $\Phi : E \to E \otimes K(D)$ is a strongly parabolic Higgs field.

Analogously to the non-parabolic case, we may define a notion of stability as follows:

Definition 6.16 A parabolic Higgs bundle will be called *stable* (resp. *semistable*) if for every Φ-invariant parabolic subbundle $F \leq E$ we have $\mathrm{par}\mu(F) < \mathrm{par}\mu(E)$ (resp. $\leq$). Furthermore, it will be called *polystable* if it is the direct sum of stable parabolic Higgs bundles of the same parabolic slope.

6.6.3 Complex Connected Sum of Riemann Surfaces

In order to describe how two parabolic Higgs bundles can be glued to a (non-parabolic) Higgs bundle, the first step is to glue their underlying surfaces with boundary as follows.

Take annuli $\mathbb{A}_1 = \{z \in \mathbb{C} \,|\, r_1 < |z| < R_1\}$ and $\mathbb{A}_2 = \{z \in \mathbb{C} \,|\, r_2 < |z| < R_2\}$ on two copies of the complex plane, and consider the Möbius transformation $f_\lambda : \mathbb{A}_1 \to \mathbb{A}_2$ with $f_\lambda(z) = \frac{\lambda}{z}$, where $\lambda \in \mathbb{C}$ with $|\lambda| = r_2 R_1 = r_1 R_2$. This is a conformal biholomorphism (equivalently bijective, angle-preserving and orientation-preserving) between the two annuli and such that the continuous extension of the function $z \mapsto |f_\lambda(z)|$ to the closure of $\mathbb{A}_1$ reverses the order of the boundary components.

Consider two compact Riemann surfaces X_1, X_2 of respective genera g_1, g_2. Choose points $p \in X_1$, $q \in X_2$ and local charts around these points $\psi_i : U_i \to \Delta(0, \varepsilon_i)$ on X_i, for $i = 1, 2$. Now fix positive real numbers $r_i < R_i < \varepsilon_i$ such that the following two conditions are satisfied:

- $\psi_i^{-1}\left(\overline{\Delta(0, R_i)}\right) \cap U_j \neq \emptyset$, for every $U_j \neq U_i$ from the complex atlas of X_i. In other words, we are considering an annulus around each of the p and q contained entirely in the neighborhood of a single chart, and
- $\frac{R_2}{r_2} = \frac{R_1}{r_1}$.

Set now

$$X_i^* = X_i \backslash \psi_i^{-1}\left(\overline{\Delta(0, r_i)}\right).$$

Choosing the biholomorphism $f_\lambda : \mathbb{A}_1 \to \mathbb{A}_2$ as above, f_λ is used to glue the two Riemann surfaces X_1, X_2 along the inverse image of the annuli $\mathbb{A}_1, \mathbb{A}_2$ on the surfaces, via the biholomorphism

$$g_\lambda : \Omega_1 = \psi_1^{-1}(\mathbb{A}_1) \to \Omega_2 = \psi_2^{-1}(\mathbb{A}_2)$$

with $g_\lambda = \psi_2^{-1} \circ f_\lambda \circ \psi_1$.

Define $X_\lambda = X_1 \#_\lambda X_2 = X_1^* \coprod X_2^* / \sim$, where the gluing of Ω_1 and Ω_2 is performed through the equivalence relation which identifies $y \in \Omega_1$ with $w \in \Omega_2$ iff $w = g_\lambda(y)$. For collections of s-many distinct points D_1 on X_1 and D_2 on X_2, this procedure is assumed to be taking place for annuli around each pair of points (p, q) for $p \in D_1$ and $q \in D_2$.

If X_1, X_2 are orientable and orientations are chosen for both, since f_λ is orientation preserving we obtain a natural orientation on the connected sum $X_1 \# X_2$ which coincides with the given ones on X_1^* and X_2^*.

Therefore, $X_\# = X_1 \# X_2$ is a Riemann surface of genus $g_1 + g_2 + s - 1$, the *complex connected sum*, where g_i is the genus of the X_i and s is the number of points in D_1 and D_2. Its complex structure however is heavily dependent on the parameters p_i, q_i, λ.

6.6.4 Gluing at the Level of Solutions to Hitchin's Equations

For gluing two parabolic G-Higgs bundles over a complex connected sum $X_\#$ of Riemann surfaces, we choose to switch to the language of solutions to Hitchin's equations and make use of the analytic techniques of Clifford Taubes for gluing instantons over 4-manifolds [100] in order to control the stability condition. These techniques have been applied to establish similar gluing constructions for solutions to gauge-theoretic equations, as for instance in [29, 34, 50, 91], and they pertain first to finding good *local model solutions* of the gauge-theoretic equations. Then one has to put, using appropriate gauge transformations, the initial data into these model forms, which are identified locally over annuli around the marked points, thus allowing a construction of a new pair over $X_\#$ that combines the original data from X_1 and X_2. This produces, however, an *approximate solution* of the equations, which then has to be corrected to an exact solution via a gauge transformation. The argument providing the existence of such a gauge is translated into a Banach fixed point theorem argument and involves the study of the linearization of a relevant elliptic operator. We briefly describe these steps in the sequel; for complete proofs we refer to [65] and [66].

6.6.4.1 The Local Model

Local SL(2,ℝ)-model solutions to the Hitchin equations can be obtained by studying the behavior of the harmonic map between a surface X with a given complex structure and the surface X with the corresponding Riemannian metric of constant curvature -4, under degeneration of the domain Riemann surface X to a nodal surface; a Riemann surface with nodes arises from an unnoded surface by pinching off one or more simple closed curves (see [99, 109] for a detailed description).

Let (E, Φ) be an SL(2,ℝ)-Higgs bundle over X with $E = L \oplus L^{-1}$ for L a holomorphic square root of the canonical line bundle over X endowed with an auxiliary Hermitian metric h_0, and $\Phi = \begin{pmatrix} 0 & q \\ 1 & 0 \end{pmatrix} \in H^0(X, \mathfrak{sl}(E))$ for q a holomorphic quadratic differential. If (E, Φ) is stable, then there is an induced Hermitian metric $H_0 = h_0 \oplus h_0^{-1}$ on E and an associated Chern connection A with respect to h, such that $A = A_L \oplus A_L^{-1}$, where A_L denotes the restriction of the connection A to the line bundle L. The stability condition implies that there exists a complex gauge transformation g unique up to unitary gauge transformations, such that $\left(A_{1,s}, \Phi_{1,s}\right) := g^*(A, \Phi)$ is a solution to the Hitchin equations. Calculations in [99] considering the Hermitian metric on L and a complex gauge giving rise to an exact solution $\left(A_{1,s}, \Phi_{1,s}\right)$ of the self-duality equations imply that

$$A_{1,s} = O\left(|\zeta|^s\right)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\left(\frac{d\zeta}{\zeta} - \frac{d\bar{\zeta}}{\bar{\zeta}}\right), \ \Phi_{1,s} = \left(1 + O\left(|\zeta|^s\right)\right)\begin{pmatrix} 0 & \frac{s}{2} \\ \frac{s}{2} & 0 \end{pmatrix}\frac{d\zeta}{i\zeta}$$

for local coordinates ζ. Therefore, after a unitary change of frame, the Higgs field $\Phi_{1,s}$ is asymptotic to the model Higgs field $\Phi_s^{\text{mod}} = \begin{pmatrix} \frac{s}{2} & 0 \\ 0 & -\frac{s}{2} \end{pmatrix}\frac{d\zeta}{i\zeta}$, while the connection $A_{1,s}$ is asymptotic to the trivial flat connection.

In conclusion, the *model solution* to the SL(2,ℝ)-Hitchin equations we will be considering is described by

$$A^{\text{mod}} = 0, \ \Phi^{\text{mod}} = \begin{pmatrix} C & 0 \\ 0 & -C \end{pmatrix}\frac{dz}{z}$$

over a punctured disk with z-coordinates around the puncture with the condition that $C \in \mathbb{R}$ with $C \neq 0$ and that the meromorphic quadratic differential $q := \det \Phi^{\text{mod}}$ has at least one simple zero. That this is indeed the generic case, is discussed in [77].

6.6.4.2 Approximate Solutions of the SL(2,$\mathbb{R}$)-Hitchin Equations

Let X be a compact Riemann surface and $D := \{p_1, \ldots, p_s\}$ a collection of s distinct points on X. Moreover, let (E, h) be a Hermitian vector bundle on E. Choose an initial pair $\left(A^{\text{mod}}, \Phi^{\text{mod}}\right)$ on E, such that in some unitary trivialization of E around each point $p \in D$, the pair coincides with the local model from Sect. 6.6.4.1; of course, on the interior of each region $X\setminus\{p\}$ the pair $\left(A^{\text{mod}}, \Phi^{\text{mod}}\right)$ need not satisfy the Hitchin equations.

One can then define *global Sobolev spaces on* X as the spaces of admissible deformations of the model unitary connection and the model Higgs field $\left(A^{\text{mod}}, \Phi^{\text{mod}}\right)$ and introduce the moduli space $\mathcal{M}\left(X^{\times}\right)$ of solutions to the Hitchin equations modulo unitary gauge transformation, which are close to the model solution over a punctured Riemann surface $X^{\times} := X - D$ for some fixed parameter $C \in \mathbb{R}$; this moduli space was explicitly constructed by Konno in [60] as a hyperkähler quotient.

In fact, as was shown by Biquard and Boalch (Lemma 5.3 in [10]) and later improved by Swoboda (Lemma 3.2 in [99]), a pair $(A, \Phi) \in \mathcal{M}\left(X^{\times}\right)$ is asymptotically close to the model $\left(A^{\text{mod}}, \Phi^{\text{mod}}\right)$ near each puncture in D. In particular, there exists a complex gauge transformation $g = \exp(\gamma)$ such that $g^*(A, \Phi)$ coincides with $\left(A_p^{\text{mod}}, \Phi_p^{\text{mod}}\right)$ on a sufficiently small neighborhood of the point p, for each $p \in D$.

We shall now use this complex gauge transformation as well as a smooth cut-off function to obtain an approximate solution to the SL(2,$\mathbb{R}$)-Hitchin equations. For fixed local coordinates z around each puncture p and given the positive function $r = |z|$ around the puncture, fix a constant $0 < R < 1$ and choose a smooth cut-off function $\chi_R : [0, \infty) \to [0, 1]$ with $\operatorname{supp}\chi \subseteq [0, R]$ and $\chi_R(r) = 1$ for $r \leq \frac{3R}{4}$. We impose the further requirement on the growth rate of this cut-off function:

$$|r\partial_r \chi_R| + \left|(r\partial r)^2 \chi_R\right| \leq k \tag{6.3}$$

for some constant k not depending on R.

The map $x \mapsto \chi_R(r(x)) : X^{\times} \to \mathbb{R}$ gives rise to a smooth cut-off function on the punctured surface $X^{\times}$ which by a slight abuse of notation we shall still denote by χ_R. We may use this function χ_R to glue the two pairs (A, Φ) and $\left(A_p^{\text{mod}}, \Phi_p^{\text{mod}}\right)$ into an *approximate solution* (Fig. 6.1)

$$\left(A_R^{app}, \Phi_R^{app}\right) := \exp(\chi_R \gamma)^* (A, \Phi).$$

The pair $\left(A_R^{app}, \Phi_R^{app}\right)$ is a smooth pair and is by construction an exact solution of the Hitchin equations away from each punctured neighborhood $\mathcal{U}_p$, while it coincides

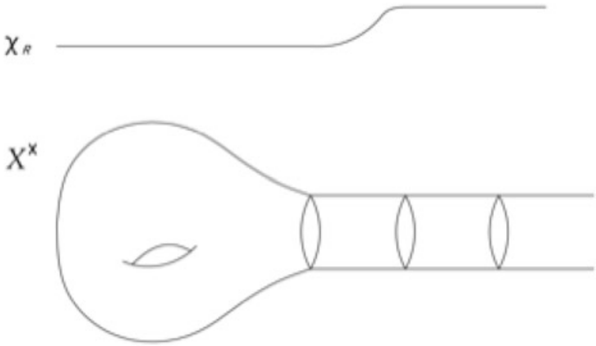


Fig. 6.1 Constructing an approximate solution over the punctured surface $X^\times$

with the model pair $\left(A_p^{\mathrm{mod}}, \Phi_p^{\mathrm{mod}}\right)$ near each puncture. More precisely, we have:

$$\left(A_R^{app}, \Phi_R^{app}\right) = \begin{cases} (A, \Phi), \text{ over } X \setminus \bigcup\limits_{p \in D} \left\{ z \in \mathcal{U}_p \,\middle|\, \frac{3R}{4} \leq |z| \leq R \right\} \\ \left(A_p^{\mathrm{mod}}, \Phi_p^{\mathrm{mod}}\right), \text{ over } \left\{ z \in \mathcal{U}_p \,\middle|\, 0 < |z| \leq \frac{3R}{4} \right\}, \text{ for each } p \in D. \end{cases}$$

Since $\left(A_R^{app}, \Phi_R^{app}\right)$ is complex gauge equivalent to an exact solution (A, Φ) of the Hitchin equations, the Higgs field Φ_R^{app} is holomorphic with respect to the holomorphic structure $\bar{\partial}_{A_R^{app}}$, in other words, one has $\bar{\partial}_{A_R^{app}} \Phi_R^{app} = 0$. Moreover, assumption (6.3) on the growth rate of the bump function χ_R provides us with a good estimate of the error up to which $\left(A_R^{app}, \Phi_R^{app}\right)$ satisfies the first among the Hitchin equations, $F(A) + \left[\Phi, \Phi^*\right] = 0$.

6.6.5 Approximate Solutions to the G-Hitchin Equations

We now wish to obtain an approximate G-Higgs pair by extending the SL(2,$\mathbb{C}$)-data via an embedding

$$\phi : \mathrm{SL}(2,\mathbb{R}) \hookrightarrow G,$$

for a reductive Lie group G. It is important that copies of a maximal compact subgroup of SL(2,$\mathbb{R}$) are mapped via ϕ into copies of a maximal compact subgroup of G and that the norm of the infinitesimal deformation ϕ_* on the complexified Lie algebra $\mathfrak{g}^{\mathbb{C}}$ satisfies a Lipschitz condition. Assuming that this is indeed the case for an embedding ϕ (examples can be found in [66] and will be demonstrated in Sect. 6.7), one gets by extension via the embedding ϕ a $G^{\mathbb{C}}$-pair satisfying the G-Hitchin equations up to an error, which we have good control of.

For $i = 1, 2$, let X_i be a closed Riemann surface of genus g_i and let $D_1 = \{p_1, \ldots, p_s\}$, $D_2 = \{q_1, \ldots, q_s\}$ a divisor of s-many distinct points on X_1, X_2 respectively. Choose local coordinates z near the points in D_1 and local coordinates w near the points in D_2. Assume that we get via an embedding as was described above approximate solutions (A_1, Φ_1), (A_2, Φ_2), which agree over neighborhoods around the points in the divisors D_1 and D_2, with $A_1 = A_2 = 0$ and with $\Phi_1(z) = -\Phi_2(w)$. Then, there is a suitable frame for the connections over which the Hermitian metrics are both described by the identity matrix and so they are constant in particular. Set $\left(A_{p,q}^{\text{mod}}, \Phi_{p,q}^{\text{mod}}\right) := \left(A_{1,p}^{\text{mod}}, \Phi_{1,p}^{\text{mod}}\right) = -\left(A_{2,q}^{\text{mod}}, \Phi_{2,q}^{\text{mod}}\right)$. We can glue the pairs (A_1, Φ_1), (A_2, Φ_2) together to get an *approximate solution* of the G-Hitchin equations over the complex connected sum $X_\# := X_1 \# X_2$ (Figs. 6.2 and 6.3):

$$\left(A_R^{app}, \Phi_R^{app}\right) := \begin{cases} (A_1, \Phi_1), & \text{over } X_1 \backslash X_2; \\ \left(A_{p,q}^{\text{mod}}, \Phi_{p,q}^{\text{mod}}\right), & \text{over } \Omega \text{ around each pair of points } (p, q); \\ (A_2, \Phi_2), & \text{over } X_2 \backslash X_1. \end{cases}$$

By construction, $\left(A_R^{app}, \Phi_R^{app}\right)$ is a smooth pair on $X_\#$, complex gauge equivalent to an exact solution of the Hitchin equations by a smooth gauge transformation defined over all of $X_\#$. It satisfies the second Hitchin equation (holomorphicity), while the first equation is satisfied up to an error which we have good control of.

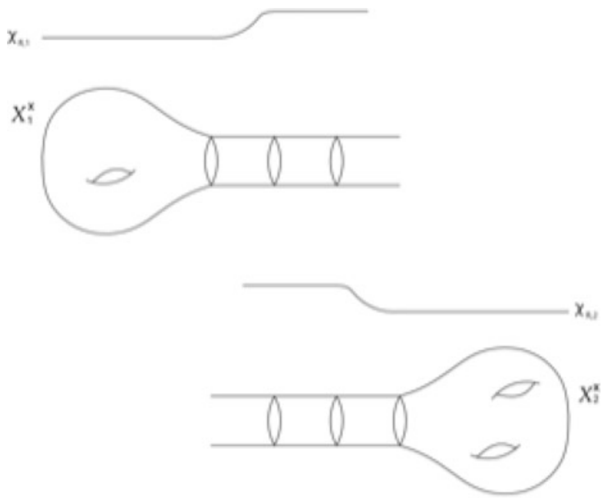


Fig. 6.2 Constructing approximate solutions over $X_1^\times$ and $X_2^\times$

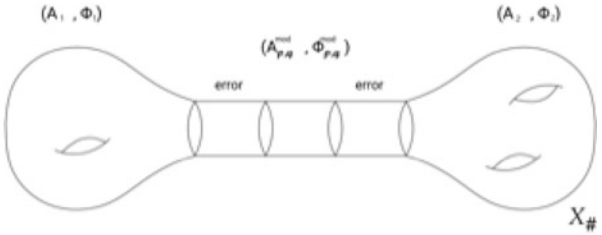


Fig. 6.3 The approximate solution $\left(A_R^{app}, \Phi_R^{app}\right)$ over the complex connected sum $X_\#$

6.6.6 The Contraction Mapping Argument

A standard strategy, due largely to Taubes [100], for correcting an approximate solution to an exact solution of gauge-theoretic equations involves studying the linearization of a relevant elliptic operator. In the Higgs bundle setting, the linearization of the Hitchin operator was first described in [77] and furthermore in [99] for solutions to the SL(2,$\mathbb{C}$)-self duality equations over a nodal surface. We are going to use this analytic machinery to correct our approximate solution to an exact solution over the complex connected sum of Riemann surfaces. We summarize this strategy below.

Let G be a connected, semisimple Lie group. For the complex connected sum $X_\#$, consider the nonlinear G-Hitchin operator at a pair $(A, \Phi) \in \Omega^1\left(X_\#, E_H\left(\mathfrak{h}^{\mathbb{C}}\right)\right) \oplus \Omega^{1,0}\left(X_\#, E_H\left(\mathfrak{m}^{\mathbb{C}}\right)\right)$:

$$\mathcal{H}(A, \Phi) = \left(F(A) - [\Phi, \tau(\Phi)], \bar{\partial}_A \Phi\right). \tag{6.4}$$

Moreover, consider the orbit map

$$\gamma \mapsto \mathcal{O}_{(A,\Phi)}(\gamma) = g^*(A, \Phi) = \left(g^*A, g^{-1}\Phi g\right),$$

for $g = \exp(\gamma)$ and $\gamma \in \Omega^0\left(X_\#, E_H\left(\mathfrak{h}^{\mathbb{C}}\right)\right)$, where $H \subset G$ is a maximal compact subgroup.

Therefore, correcting the approximate solution $\left(A_R^{app}, \Phi_R^{app}\right)$ to an exact solution of the G-Hitchin equations accounts to finding a point γ in the complex gauge orbit of $\left(A_R^{app}, \Phi_R^{app}\right)$, for which $\mathcal{H}\left(g^*\left(A_R^{app}, \Phi_R^{app}\right)\right) = 0$. However, since we have seen that the second equation is satisfied by the pair $\left(A_R^{app}, \Phi_R^{app}\right)$ and since the condition $\bar{\partial}_A \Phi = 0$ is preserved under the action of the complex gauge group $\mathcal{G}_H^{\mathbb{C}}$, we actually seek a solution γ to the following equation

$$\mathcal{F}_R(\gamma) := pr_1 \circ \mathcal{H} \circ \mathcal{O}_{\left(A_R^{app}, \Phi_R^{app}\right)}(\exp(\gamma)) = 0.$$

For a Taylor series expansion of this operator

$$\mathcal{F}_R(\gamma) = pr_1\mathcal{H}\left(A_R^{app}, \Phi_R^{app}\right) + L_{\left(A_R^{app}, \Phi_R^{app}\right)}(\gamma) + Q_R(\gamma),$$

where Q_R includes the quadratic and higher order terms in γ, we can then see that $\mathcal{F}_R(\gamma) = 0$ if and only if γ is a fixed point of the map

$$\begin{aligned} T : H_B^2(X_\#) &\to H_B^2(X_\#) \\ \gamma &\mapsto -G_R\left(\mathcal{H}\left(A_R^{app}, \Phi_R^{app}\right) + Q_R(\gamma)\right), \end{aligned}$$

where we denoted $G_R := L^{-1}_{\left(A_R^{app}, \Phi_R^{app}\right)}$ and $H_B^2(X_\#)$ is the Hilbert space defined by

$$H_B^2(X_\#) := \left\{\gamma \in L^2(X_\#) \,\middle|\, \nabla_B\gamma, \nabla_B^2\gamma \in L^2(X_\#)\right\},$$

for a fixed background connection ∇_B defined as a smooth extension to $X_\#$ of the model connection $A_{p,q}^{\mathrm{mod}}$ over the cylinder for each pair of points (p, q).

The problem then reduces to showing that the mapping T is a contraction of the open ball B_{ρ_R} of radius ρ_R in $H_B^2(X_\#)$, since then from Banach's fixed point theorem there will exist a unique γ such that $T(\gamma) = \gamma$, in other words, such that $\mathcal{F}_R(\gamma) = 0$. In particular, one needs to show that:

1. T is a contraction defined on B_{ρ_R} for some ρ_R, and
2. T maps B_{ρ_R} to B_{ρ_R}.

In order to complete the above described contraction mapping argument, we need to show the following:

1. The linearized operator at the approximate solution $L_{\left(A_R^{app}, \Phi_R^{app}\right)}$ is invertible.
2. There is an upper bound for the inverse operator $G_R = L^{-1}_{\left(A_R^{app}, \Phi_R^{app}\right)}$ as an operator $L^2(rdrd\theta) \to L^2(rdrd\theta)$.
3. There is an upper bound for the inverse operator $G_R = L^{-1}_{\left(A_R^{app}, \Phi_R^{app}\right)}$ also when viewed as an operator $L^2(rdrd\theta) \to H_B^2(X_\#, rdrd\theta)$.
4. We can control a Lipschitz constant for Q_R, that means there exists a constant $C > 0$ such that

$$\|Q_R(\gamma_1) - Q_R(\gamma_0)\|_{L^2} \leq C\rho\|\gamma_1 - \gamma_0\|_{H_B^2}$$

for all $0 < \rho \leq 1$ and $\gamma_0, \gamma_1 \in B_\rho$, the closed ball of radius ρ around 0 in $H_B^2(X_\#)$.

6.6.7 Correcting an Approximate Solution to an Exact Solution

We shall focus on the linear term in the Taylor series expansion. The linearization operator $L_{(A,\Phi)}$ at a pair $(A,\Phi) \in \Omega^1\left(X_\#, E_H\left(\mathfrak{h}^{\mathbb{C}}\right)\right) \oplus \Omega^{1,0}\left(X_\#, E_H\left(\mathfrak{m}^{\mathbb{C}}\right)\right)$ is defined by

$$L_{(A,\Phi)} := -i * D\mathcal{F}(\gamma) : \Omega^0\left(X_\#, E_H\left(\mathfrak{h}^{\mathbb{C}}\right)\right) \to \Omega^0\left(X_\#, E_H\left(\mathfrak{h}^{\mathbb{C}}\right)\right),$$

where $D\mathcal{F}(\gamma)$ denotes the differential

$$D\mathcal{F}(\gamma) = \partial_A \bar{\partial}_A \gamma - \bar{\partial}_A \partial_A \gamma^* + [\Phi, -\tau([\Phi,\gamma])] + [[\Phi,\gamma], -\tau(\Phi)],$$

for $H \subset G$ a maximal compact subgroup and τ the compact real form of $\mathfrak{g}^{\mathbb{C}}$. It satisfies the following lemma; for a proof, see Lemma 5.1 in [66]:

Lemma 6.3 *For $\gamma \in \Omega^0(X_\#, E_H(\mathfrak{h}))$, the linearization operator satisfies*

$$\left\langle L_{(A,\Phi)}\gamma, \gamma \right\rangle_{L^2} = \|d_A\gamma\|_{L^2}^2 + 2\,\|[\Phi,\gamma]\|_{L^2}^2 \geq 0.$$

In particular, $L_{(A,\Phi)}\gamma = 0$ if and only if $d_A\gamma = [\Phi,\gamma] = 0$.

In order to prove the existence of the inverse operator $G_R := L^{-1}_{\left(A_R^{app},\Phi_R^{app}\right)}$ and obtain an upper bound for its L^2-norm, we apply a version of the Cappell–Lee–Miller gluing theorem for a pair of cylindrical $\mathbb{Z}_2$-graded Dirac-type operators (see [23] and [83, §5.B]).

For our approximate solution $\left(A_R^{app}, \Phi_R^{app}\right)$ constructed over $X_\#$ with $0 < R < 1$ and $T = -\log R$, consider the elliptic complex

$$0 \to \Omega^0\left(X_\#, E_H\left(\mathfrak{h}^{\mathbb{C}}\right)\right) \xrightarrow{L_{1,T}} \Omega^1\left(X_\#, E_H\left(\mathfrak{h}^{\mathbb{C}}\right)\right) \oplus \Omega^{1,0}\left(X_\#, E_H\left(\mathfrak{g}^{\mathbb{C}}\right)\right)$$
$$\xrightarrow{L_{2,T}} \Omega^2\left(X_\#, E_H\left(\mathfrak{h}^{\mathbb{C}}\right)\right) \oplus \Omega^2\left(X_\#, E_H\left(\mathfrak{g}^{\mathbb{C}}\right)\right) \to 0,$$

where

$$L_{1,T}\gamma = \left(d_{A_R^{app}}\gamma, \left[\Phi_R^{app}, \gamma\right]\right)$$

is the linearization of the complex gauge group action and

$$L_{2,T}(\alpha,\varphi) = D\mathcal{H}(\alpha,\varphi) = \begin{pmatrix} d_{A_R^{app}}\alpha + \left[\Phi_R^{app}, -\tau(\varphi)\right] + \left[\varphi, -\tau\left(\Phi_R^{app}\right)\right] \\ \bar{\partial}_{A_R^{app}}\varphi + \left[\alpha, \Phi_R^{app}\right] \end{pmatrix}$$

is the differential of the Hitchin operator from (6.4). Note that, in general, it does not hold that

$$L_{2,T}L_{1,T} = \left[F_{A_R^{app}}, \gamma\right] + \left[\left[\Phi_R^{app}, -\tau\left(\Phi_R^{app}\right)\right], \gamma\right] = 0,$$

since $\left(A_R^{app}, \Phi_R^{app}\right)$ need not be an exact solution. Decomposing $\Omega^*\left(X_\#, E_H\left(\mathfrak{g}^{\mathbb{C}}\right)\right)$ into forms of even, respectively odd total degree, we may introduce the $\mathbb{Z}_2$-graded Dirac-type operator

$$\mathfrak{D}_T := \begin{pmatrix} 0 & L_{1,T}^* + L_{2,T} \\ L_{1,T} + L_{2,T}^* & 0 \end{pmatrix}$$

on the closed surface $X_\#$.

For applying the Cappell–Lee–Miller theorem, one has to study the kernel $\ker\left(L_1 + L_2^*\right)$ on the extended L^2-space $L^2_{\text{ext}}\left(X_\#^\times\right)$ for the nodal surface $X_\#^\times$ obtained by extending the cylindrical neck of $X_\#$ infinitely (see Definition 6.2 and §6.2 in [66] for the precise definitions).

As $R \searrow 0$, the curve $X_\#$ degenerates to a nodal surface $X_\#^\times$ (equivalently, the cylindrical neck of $X_\#$ extends infinitely). For the cut-off functions χ_R that we considered in obtaining the approximate pair $\left(A_R^{app}, \Phi_R^{app}\right)$, their support will tend to be empty as $R \searrow 0$, therefore the "error regions" disappear along with the neck Ω, thus $\left(A_R^{app}, \Phi_R^{app}\right) \to (A_0, \Phi_0)$ uniformly on compact subsets with

$$\left(A_0^{app}, \Phi_0^{app}\right) = \begin{cases} (A_1, \Phi_1), & X_1 \backslash \Omega \\ (A_2, \Phi_2), & X_2 \backslash \Omega \end{cases}$$

an exact solution with the holonomy of the associated flat connection in G.

For trivial kernel $\ker\left(L_1 + L_2^*\right)$, and computing the upper bound for the inverse operator and a Lipschitz constant for the quadratic or higher order terms in the Taylor series expansion, one can correct the approximate solution constructed into an exact solution of the G-Hitchin equations. The contraction mapping argument described above then provides the following:

Theorem 6.11 *There exists a constant* $0 < R_0 < 1$*, and for every* $0 < R < R_0$ *there exist a constant* $\sigma_R > 0$ *and a unique section* $\gamma \in H_B^2\left(X_\#, E_H\left(\mathfrak{h}^{\mathbb{C}}\right)\right)$ *satisfying* $\|\gamma\|_{H_B^2(X_\#)} \leq \sigma_R$*, so that, for* $g = \exp(\gamma)$*,*

$$(A_\#, \Phi_\#) = g^*\left(A_R^{app}, \Phi_R^{app}\right)$$

is an exact solution of the G*-Hitchin equations over the closed surface* $X_\#$*.*

Theorem 6.11 now implies that for $\bar{\partial} := A_{\#}^{0,1}$, the Higgs bundle $\left(E_{\#} := \left(\mathbb{E}_{\#}, \bar{\partial}\right), \Phi_{\#}\right)$ is a polystable G-Higgs bundle over the complex connected sum $X_{\#}$. Collecting the steps from the previous subsections one has the following:

Theorem 6.12 *Let X_1 be a closed Riemann surface of genus g_1 and $D_1 = \{p_1, \ldots, p_s\}$ be a collection of s distinct points on X_1. Let also G be a subgroup of $GL(n, \mathbb{C})$. Consider respectively a closed Riemann surface X_2 of genus g_2 and a collection of also s distinct points $D_2 = \{q_1, \ldots, q_s\}$ on X_2. Let $(E_1, \Phi_1) \to X_1$ and $(E_2, \Phi_2) \to X_2$ be parabolic polystable G-Higgs bundles with corresponding solutions to the Hitchin equations (A_1, Φ_1) and (A_2, Φ_2). Assume that these solutions agree with model solutions $\left(A_{1,p_i}^{\text{mod}}, \Phi_{1,p_i}^{\text{mod}}\right)$ and $\left(A_{2,q_j}^{\text{mod}}, \Phi_{2,q_j}^{\text{mod}}\right)$ near the points $p_i \in D_1$ and $q_j \in D_2$, and that the model solutions satisfy $\left(A_{1,p_i}^{\text{mod}}, \Phi_{1,p_i}^{\text{mod}}\right) = -\left(A_{2,q_j}^{\text{mod}}, \Phi_{2,q_j}^{\text{mod}}\right)$, for s pairs of points $\left(p_i, q_j\right)$. Then there is a polystable G-Higgs bundle $(E_{\#}, \Phi_{\#}) \to X_{\#}$, constructed over the complex connected sum of Riemann surfaces $X_{\#} = X_1 \# X_2$, which agrees with the initial data over $X_{\#} \backslash X_1$ and $X_{\#} \backslash X_2$.*

Definition 6.17 We call a G-Higgs bundle constructed by the procedure developed above a *hybrid G-Higgs bundle*.

6.6.8 Topological Invariants

The connected component of the moduli space $\mathcal{M}(G)$ that a hybrid Higgs bundle lies, can be determined by Higgs bundle topological invariants, and one needs to understand how these invariants behave under the complex connected sum operation. The next two propositions show that there is an additivity property for topological invariants over the connected sum operation, both from the Higgs bundle and the surface group representation point of view.

When the group G is a subgroup of $\mathrm{GL}(n, \mathbb{C})$, the data of a parabolic G-Higgs bundle (defined in full generality in [11]) reduce to the data of a parabolic Higgs bundle as seen in Sect. 6.6.2.1. Moreover, the basic topological invariant of a parabolic (resp. non-parabolic) pair is the parabolic degree (resp. degree) of some underlying parabolic (resp. non-parabolic) bundle in the Higgs bundle data. We refer to [67] for a detailed description of this data and the corresponding topological invariants for a number of cases of parabolic G-Higgs bundles.

The following proposition now describes an additivity property for the degrees:

Proposition 6.2 (Proposition 8.1 in [66]) *Let $X_{\#} = X_1 \# X_2$ be the complex connected sum of two closed Riemann surfaces X_1 and X_2 with divisors D_1 and D_2 of s distinct points on each surface, and let V_1, V_2 be parabolic vector bundles over X_1 and X_2 respectively. Then, if the parabolic bundles V_1, V_2 glue to a bundle $V_1 \# V_2$ over $X_{\#}$, the following identity holds*

$$\deg\left(V_1 \# V_2\right) = par\deg\left(V_1\right) + par\deg\left(V_2\right).$$

Considering the connected sum of the underlying topological surfaces $\Sigma = \Sigma_1 \cup_\gamma \Sigma_2$ along a loop γ, a notion of Toledo invariant is defined for representations over these subsurfaces with boundary; see [21] for a detailed definition in this context. Moreover, the authors in [21] have established an additivity property for the Toledo invariant over a connected sum of surfaces. In particular:

Proposition 6.3 (Proposition 3.2 in [21]) *If $\Sigma = \Sigma_1 \cup_\gamma \Sigma_2$ is the connected sum of two subsurfaces Σ_i along a simple closed separating loop γ, then*

$$T_\rho = T_{\rho_1} + T_{\rho_2},$$

where $\rho_i = \rho \left|_{\pi_1(\Sigma_i)}\right.$, for $i = 1, 2$.

The two propositions above allow one to determine the topological invariants of the hybrid Higgs bundles, respectively fundamental group representations, from the topological invariants of the underlying objects that were deformed and glued together. Note, in particular, that this property implies that the amalgamated product of two maximal representations is again a maximal representation defined over the compact surface Σ.

6.7 Examples: Model Higgs Bundles in Exceptional Components of Orthogonal Groups

We now exhibit specific examples where the previous gluing construction can provide model objects lying inside higher Teichmüller spaces of particular geometric importance.

When the Lie group is $G = \mathrm{Sp}(4, \mathbb{R})$, hybrid Higgs bundles in the exceptional connected components of the maximal $G = \mathrm{Sp}(4, \mathbb{R})$-Higgs bundles identified by Gothen in [42] were obtained in [66]. We next provide such examples in the case of the group $G = \mathrm{SO}(p, p+1)$, which involves an extra parameter compared to the $\mathrm{Sp}(4, \mathbb{R})$-case. Note, however, that a maximality property is not apparent in this case apart from when $p = 2$, since the group $G = \mathrm{SO}(p, p+1)$ for $p \neq 2$ is not Hermitian of noncompact type; cf. the discussion on maximality in Sect. 6.4.2.

6.7.1 SO (p, q)-Higgs Bundle Data

The connected components of the $\mathrm{SO}(p, q)$-character variety $\mathcal{R}(\mathrm{SO}(p, q))$ can be more explicitly described using the theory of Higgs bundles. Let X be a compact Riemann surface with underlying topological surface Σ. Under the non-abelian Hodge correspondence, fundamental group representations into the group $\mathrm{SO}(p, q)$

correspond to holomorphic tuples (V, Q_V, W, Q_W, η) over X, where:

- (V, Q_V) and (W, Q_W) are holomorphic orthogonal bundles of rank p and q respectively with the additional condition that $\wedge^p (V) \cong \wedge^q (W)$.
- $\eta : W \to V \otimes K$ is a holomorphic section of $\mathrm{Hom}(W, V) \otimes K$.

Using Higgs bundle methods, in particular a real valued proper function defined by the L^2-norm of the Higgs field and a natural holomorphic $\mathbb{C}^*$-action, the authors in [5] classify *all* polystable local minima of the Hitchin function in $\mathcal{M}(\mathrm{SO}(p, q))$, for $2 < p \leq q$. For these moduli spaces, not all local minima occur at fixed points of the $\mathbb{C}^*$-action and additional connected components of $\mathcal{M}(\mathrm{SO}(p, q))$ emerge by constructing a map

$$\Psi : \mathcal{M}_{K^p}(\mathrm{SO}(1, q-p+1)) \times \bigoplus_{j=1}^{p-1} H^0\left(X, K^{2j}\right) \to \mathcal{M}(\mathrm{SO}(p, q)),$$

which is an isomorphism onto its image, open and closed. In the description above, $\mathcal{M}_{K^p}(\mathrm{SO}(1, q-p+1))$ denotes the moduli space of K^p-twisted $\mathrm{SO}(1, q-p+1)$-Higgs bundles on the Riemann surface X, where K is the canonical line bundle over X, and $\bigoplus_{j=1}^{p-1} H^0\left(X, K^{2j}\right)$ denotes the vector space of holomorphic differentials of degree $2j$. Note that a K^p-twisted $\mathrm{SO}(1, n)$-Higgs bundle is defined by a triple $\left(I, \hat{W}, \hat{\eta}\right)$, where $\left(\hat{W}, Q_{\hat{W}}\right)$ is a rank n orthogonal bundle, $I = \wedge^n \hat{W}$ and $\hat{\eta} \in H^0\left(\mathrm{Hom}\left(\hat{W}, I\right) \otimes K^p\right)$. A point in the image of the map Ψ is then described by

$$\Psi\left(\left(I, \hat{W}, \hat{\eta}\right), q_2, \ldots, q_{2p-2}\right) = (V, W, \eta), \tag{6.5}$$

where

$$V := I \otimes \left(K^{p-1} \oplus K^{p-3} \oplus \ldots \oplus K^{3-p} \oplus K^{1-p}\right);$$

$$W := \hat{W} \oplus I \otimes \left(K^{p-2} \oplus K^{p-4} \oplus \ldots \oplus K^{4-p} \oplus K^{2-p}\right);$$

$$\eta := \begin{pmatrix} \hat{\eta} & q_2 & q_4 & \cdots & q_{2p-2} \\ 0 & 1 & q_2 & \cdots & q_{2p-4} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & 1 & q_2 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}. \tag{6.6}$$

Moreover, an $\mathrm{SO}(p,q)$-Higgs bundle (V, W, η) is (poly)stable if and only if the K^p-twisted $\mathrm{SO}(1,n)$-Higgs bundle $\left(I, \hat{W}, \hat{\eta}\right)$ is (poly)stable (see Lemma 4.4 in [5]).

The case when $q = p + 1$ is even more special, because the relevant K^p-twisted $\mathrm{O}(q - p + 1)$-Higgs bundles in the pre-image of Ψ are now rank 2 orthogonal bundles. In this case, when the first Stiefel-Whitney class $w_1\left(\hat{W}, Q_{\hat{W}}\right)$ vanishes, then the structure group of $\hat{W}$ reduces to $\mathrm{SO}(2, \mathbb{C}) \cong \mathbb{C}^*$ and thus

$$\left(\hat{W}, Q_{\hat{W}}\right) \cong \left(M \oplus M^{-1}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right),$$

for a degree d holomorphic line bundle $M \in \mathrm{Pic}^d(X)$, while for stability reasons d is an integer in the interval $[0, p(2g - 2)]$. This degree is a new topological invariant, which distinguishes extra components of the moduli space $\mathcal{M}(\mathrm{SO}(p, p + 1))$, and in [24] is proven the following:

Theorem 6.13 (Theorem 4.1 in [24]) *For each integer $d \in (0, p(2g - 2) - 1]$ there is a smooth connected component $\mathcal{R}_d(SO(p, p + 1))$ of the moduli space $\mathcal{R}(SO(p, p + 1))$, which does not contain representations with compact Zariski closure.*

Since all points in these $p(2g - 2) - 1$ many components are smooth, all corresponding fundamental group representations are *irreducible representations*. In fact, these representations are conjectured in [24] to have Zariski dense image. For this reason we shall call these components *exceptional* to distinguish them among the rest of the components of the character varieties $\mathcal{R}(\mathrm{SO}(p, q))$ that are not detected by the fixed points of the $\mathbb{C}^*$-action.

Definition 6.18 The connected components of the moduli space $\mathcal{M}(\mathrm{SO}(p, p + 1))$, which are smooth, will be called the *exceptional components* of the moduli space $\mathcal{M}(\mathrm{SO}(p, p + 1))$.

For each integer $0 < d \leq p(2g - 2) - 1$, the Higgs bundles (V, W, η) in the exceptional components are described by the map Ψ from (6.5) as follows:

$$(V, Q_V) = \left(K^{p-1} \oplus K^{p-3} \oplus \ldots \oplus K^{3-p} \oplus K^{1-p}, \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}\right),$$

$$(W, Q_W) = \left(M \oplus K^{p-2} \oplus K^{p-4} \oplus \ldots \oplus K^{4-p} \oplus K^{2-p} \oplus M^{-1}, \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}\right), \tag{6.7}$$

$$\eta = \begin{pmatrix} 0 & 0 & \dots & 0 & \nu \\ 1 & q_2 & q_4 & \cdots & q_{2p-2} \\ 0 & 1 & q_2 & & q_{2p-4} \\ 0 & 0 & \ddots & & \vdots \\ \vdots & & \ddots & 1 & q_2 \\ 0 & 0 & \dots & 0 & \mu \end{pmatrix} : V \to W \otimes K,$$

for $M \in \mathrm{Pic}^d(X)$, and sections $\mu \in H^0\left(M^{-1}K^p\right) \setminus \{0\}$ and $\nu \in H^0(MK^p)$ with $0 \neq \mu \neq \lambda\nu$.

In the case when $d = p(2g-2)$, then (V, W, η) lies in the *Hitchin component* of $\mathcal{M}(\mathrm{SO}(p, p+1))$ with data

$$(V, Q_V) = \left(K^{p-1} \oplus K^{p-3} \oplus \ldots \oplus K^{3-p} \oplus K^{1-p}, \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \right),$$

$$(W, Q_W) = \left(K^p \oplus K^{p-2} \oplus K^{p-4} \oplus \ldots \oplus K^{4-p} \oplus K^{2-p} \oplus K^{-p}, \begin{pmatrix} & 1 \\ 1 & \end{pmatrix} \right),$$

$$\eta = \begin{pmatrix} q_2 & q_4 & \dots & q_{2p-2} & q_{2p} \\ 1 & q_2 & q_4 & \cdots & q_{2p-2} \\ 0 & 1 & q_2 & & q_{2p-4} \\ 0 & 0 & \ddots & & \vdots \\ \vdots & & \ddots & 1 & q_2 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} : V \to W \otimes K.$$

6.7.2 Hitchin Equations for Orthogonal Groups

The moduli space $\mathcal{M}(\mathrm{SO}(p,q))$ of polystable $\mathrm{SO}(p,q)$-Higgs bundles is alternatively viewed as the moduli space of polystable pairs $\left(\bar{\partial}_E, \Phi\right)$ modulo the gauge group $\mathcal{G}(\mathbf{E})$, where $\bar{\partial}_E$ is a Dolbeault operator on a principal $\mathrm{SO}(p,\mathbb{C}) \times \mathrm{SO}(q,\mathbb{C})$-bundle $\mathbf{E}$ and $\Phi \in \Omega^{1,0}\left(E\left(\mathfrak{m}^{\mathbb{C}}\right)\right)$ satisfying $\bar{\partial}_E(\Phi) = 0$, for the (-1)-eigenspace $\mathfrak{m}$ in the Cartan decomposition of the Lie algebra of the group $\mathrm{SO}(p,q)$.

For the principal $\mathrm{SO}(p,\mathbb{C}) \times \mathrm{SO}(q,\mathbb{C})$-bundle $\mathbf{E}$ equipped with a Dolbeault operator $\bar{\partial}_E$, the gauge group

$$\mathcal{G}(\mathbf{E}) \cong \Omega^0\left(E_{\mathrm{SO}(p,\mathbb{C})}(\mathrm{SO}(p,\mathbb{C}))\right) \times \Omega^0\left(E_{\mathrm{SO}(q,\mathbb{C})}(\mathrm{SO}(q,\mathbb{C}))\right)$$

acts on the operators $\bar{\partial}_E$ by conjugation, where $\mathbf{E} = E_{\mathrm{SO}(p,\mathbb{C})} \times E_{\mathrm{SO}(q,\mathbb{C})}$. Now a Dolbeault operator on $\mathbf{E}$ corresponds to a connection A on the reduction V of $\mathbf{E}$ to $\mathrm{SO}(p,\mathbb{C}) \times \mathrm{SO}(q,\mathbb{C})$ and consider a Higgs field $\Phi \in \Omega^{1,0}\left(V\left(\mathfrak{m}^{\mathbb{C}}\right)\right)$.

The group $G = \mathrm{SO}(p,q)$ is a real form of $\mathrm{SO}(p+q,\mathbb{C})$. It coincides with the compact real form when $p = q = 0$ and with the split real form when $p = q$ for $p+q$ even, or when $q = p+1$ for $p+q$ odd. Matrix conjugation $\tau(X) = \bar{X}$ defines the compact real form; indeed, we check

$$\begin{aligned}\mathfrak{so}(p+q) &= \left\{X \in \mathfrak{so}(p+q,\mathbb{C}) \,\middle|\, X = \bar{X}\right\}\\ &= \left\{X \in \mathfrak{so}(p+q,\mathbb{R}) \,\middle|\, X + X^T = 0\right\}.\end{aligned}$$

If we locally write $\Phi = \varphi dz$, then a calculation shows that

$$[\Phi, \tau(\Phi)] = \begin{pmatrix} -\varphi\varphi^* - \bar{\varphi}\varphi^T & \\ & -\varphi^T\bar{\varphi} - \varphi^*\varphi \end{pmatrix}.$$

The Hitchin-Kobayashi correspondence for $G = \mathrm{SO}(p,q)$ provides that if an $\mathrm{SO}(p,q)$-Higgs bundle (V, Q_V, W, Q_W, η) is polystable, then and only then the pair (A, Φ) as considered above satisfies the Hitchin equation

$$\begin{cases} F_A - [\Phi, \tau(\Phi)] = 0 \\ \quad \bar{\partial}_A(\Phi) = 0, \end{cases}$$

where F_A denotes the curvature of the unique connection compatible with the structure group reduction and the holomorphic structure. For a local description of the connection $A = (A_1, A_2)$ the equation $F_A - [\Phi, \tau(\Phi)] = 0$ becomes the pair

$$\begin{aligned} F_{A_1} + \varphi\varphi^* + \bar{\varphi}\varphi^T &= F_{A_1} + 2\mathrm{Re}\left(\varphi\varphi^*\right) = 0\\ F_{A_2} + \varphi^T\bar{\varphi} + \varphi^*\varphi &= F_{A_2} + 2\mathrm{Re}\left(\varphi^T\bar{\varphi}\right) = 0. \end{aligned}$$

6.7.3 Model Parabolic SL $(2, \mathbb{R})$-Higgs Bundles

Parabolic $\mathrm{SL}(2,\mathbb{R})$-Higgs bundles corresponding via the non-abelian Hodge correspondence to Fuchsian representations of the fundamental group of a punctured surface into the group $\mathrm{PSL}(2,\mathbb{R})$ were first identified by Biswas, Arés-Gastesi and Govindarajan in [12]; see also the article of Mondello [80] for a complete topological description of the relevant representation space. We next investigate these pairs more closely.

Let $D = \{x_1, \ldots, x_s\}$ be a finite collection of s-many points on a closed genus g Riemann surface X, such that $2g - 2 + s > 0$. Let K denote the canonical line bundle over the Riemann surface X. Consider the pair (E, Φ), where:

1. $E := (L \otimes \iota)^* \oplus L$,
   where L is a line bundle with $L^2 = K$ and $\iota := \mathcal{O}_X(D)$ denotes the line bundle over the divisor D; we equip the bundle E with a parabolic structure given by a trivial flag $E_{x_i} \supset \{0\}$ and weight $\frac{1}{2}$ for every $1 \leq i \leq s$.
2. $\Phi := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in H^0(X, \mathrm{End}(E) \otimes K \otimes \iota)$.

Then, the pair (E, Φ) is a stable parabolic $\mathrm{SL}(2, \mathbb{R})$-Higgs bundle with parabolic degree $\mathrm{pardeg}(E) = 0$. Therefore, from the non-abelian Hodge correspondence on non-compact curves [97], the vector bundle E supports a tame harmonic metric; the local estimate for this Hermitian metric on E restricted to the line bundle L is

$$r^{\frac{1}{2}} |\log r|^{\frac{1}{2}},$$

for $r = |z|$. Indeed, if $\beta \in \mathbb{R}$ denotes in general the weights in the filtration of the filtered local system F corresponding to a parabolic Higgs bundle with weights α, for $0 \leq \alpha < 1$, then, if W_k is the span of vectors of weights $\leq k$, the weight filtration of $\mathrm{Res}_x(F)$ describes the behavior of the tame harmonic map under the local estimate

$$C r^{\beta} |\log r|^{\frac{k}{2}}.$$

In our case, the weight is $\alpha = \frac{1}{2} = \beta$ and the residue at each point $x_i \in D$ is $N = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, an upper triangular 2×2 nilpotent matrix. Thus, its weight filtration is $W_{-2} = 0$, $W_{-1} = W_0 = \mathrm{Im}(N) = \ker(N)$, and $W_1 =$ the whole space. Therefore, in the notation of Simpson from [97] we have $L \subset W_1$ and $L \not\subset W_0 = W_{-1}$, while the Hermitian metric on the line bundle L is locally

$$r^{\alpha} |\log r|^{\frac{k}{2}} = r^{\frac{1}{2}} |\log r|^{\frac{1}{2}}.$$

For the parabolic dual $(L \otimes \iota)^*$, the weight is by construction equal to $1 - \frac{1}{2}$ and in the weight filtration for the residue it is $(L \otimes \iota)^* \subset W_{-1}$ and $L \not\subset W_1$. Thus, the Hermitian metric on $(L \otimes \iota)^*$ is locally

$$r^{\alpha} |\log r|^{\frac{k}{2}} = r^{1-\frac{1}{2}} |\log r|^{-\frac{1}{2}} = r^{\frac{1}{2}} |\log r|^{-\frac{1}{2}}.$$

In conclusion, the metric on Hom $\left(L, (L \otimes \iota)^*\right)$ is induced by the restricted tame harmonic metric of E on the line bundles L and $(L \otimes \iota)^*$, as a section of $L^* \otimes (L \otimes \iota)^*$ and is locally described by

$$r^{-\frac{1}{2}}|\log r|^{-\frac{1}{2}} \cdot r^{\frac{1}{2}}|\log r|^{-\frac{1}{2}} = |\log r|^{-1},$$

for $r = |z|$. Subsequently, the metric on the tangent bundle L^{-2} is locally

$$r^{-\frac{1}{2}}|\log r|^{-\frac{1}{2}} \cdot r^{-\frac{1}{2}}|\log r|^{-\frac{1}{2}} = r^{-1}|\log r|^{-1}$$

and is therefore the Poincaré metric of the punctured disk on $\mathbb{C}$; we refer the interested reader to [12] and [97] for further information.

6.7.4 *Parabolic SO* $(p, p+1)$*-Models*

In this subsection we construct model parabolic SO $(p, p+1)$-Higgs bundles which shall be later on used in providing the desired (non-parabolic) SO $(p, p+1)$-models in the exceptional components over the complex connected sum of Riemann surfaces. Of critical importance to this construction are the parabolic SL $(2, \mathbb{R})$-Higgs bundles (E, Φ) of Biswas, Arés-Gastesi and Govindarajan from [12] described earlier. As we have seen in Sect. 6.6.4.1, from the gauge theoretic viewpoint, a model solution to the SL $(2, \mathbb{C})$-Hitchin equations that corresponds to the polystable pair (E, Φ) is given by a pair $\left(A^{\text{ mod}}, \Phi^{\text{ mod}}\right)$, where

$$A^{\text{ mod}} = 0, \;\; \Phi^{\text{ mod}} = \begin{pmatrix} C & 0 \\ 0 & -C \end{pmatrix} \frac{dz}{z}$$

over a punctured disk with z-coordinates around the puncture with the condition that $C \in \mathbb{R}$ with $C \neq 0$, and that the meromorphic quadratic differential $q := \det \Phi^{\text{ mod}}$ has at least one simple zero.

6.7.4.1 Models via the Irreducible Representation SL $(2, \mathbb{R}) \hookrightarrow$ SO $(p, p+1)$

We next construct model parabolic SO $(p, p+1)$-Higgs bundles lying inside the parabolic Teichmüller component for SO $(p, p+1)$. The general construction of this component was carried out in [67], while in the non-parabolic case, a detailed construction of models can be found in [4].

The connected component of the special orthogonal group containing the identity $\mathrm{SO}_0(p, p+1)$ is a split real form of $\mathrm{SO}(2p+1, \mathbb{C})$. The Lie algebra of $\mathrm{SO}(p, p+1)$ is

$$\begin{aligned}\mathfrak{so}(p, p+1) &= \left\{X \in \mathfrak{sl}(2p+1, \mathbb{R}) \,\middle|\, X^t I_{p,p+1} + I_{p,p+1} X = 0\right\} \\ &= \left\{\begin{pmatrix} X_1 & X_2 \\ X_2^t & X_3 \end{pmatrix} \middle| X_1, X_3 \text{ real skew-sym. of rank } p, p+1 \text{ resp.}; \right. \\ &\qquad \left. X_2 \text{ real } (p \times (p+1))\text{-matrix}\right\}.\end{aligned}$$

The Lie algebra $\mathfrak{so}(p, p+1)$ admits a Cartan decomposition $\mathfrak{so}(p, p+1) = \mathfrak{h} \oplus \mathfrak{m}$ into its (± 1)-eigenspaces, where

$$\mathfrak{h} = \mathfrak{so}(p) \times \mathfrak{so}(p+1) = \left\{\begin{pmatrix} X_1 & 0 \\ 0 & X_3 \end{pmatrix} \middle| X_1 \in \mathfrak{so}(p), X_3 \in \mathfrak{so}(p+1)\right\},$$

$$\mathfrak{m} = \left\{\begin{pmatrix} 0 & X_2 \\ X_2^t & 0 \end{pmatrix} \middle| X_2 \text{ real } (p \times (p+1))\text{-matrix}\right\}.$$

The Cartan decomposition of the complex Lie algebra is

$$\mathfrak{so}(2p+1, \mathbb{C}) = (\mathfrak{so}(p, \mathbb{C}) \times \mathfrak{so}(p+1, \mathbb{C})) \oplus \mathfrak{m}^{\mathbb{C}},$$

where

$$\mathfrak{m}^{\mathbb{C}} = \left\{\begin{pmatrix} 0 & X_2 \\ -X_2^t & 0 \end{pmatrix} \middle| X_2 \text{ complex } (p \times (p+1))\text{-matrix}\right\}.$$

If $\mathfrak{c}$ is a Cartan subalgebra of $\mathfrak{so}(p, p+1)$ and Δ is the set of the corresponding roots, then the element

$$\sum_{\alpha \in \Delta} c_\alpha X_\alpha \in \mathfrak{so}(2p+1, \mathbb{C}),$$

is regular nilpotent, for $c_\alpha \neq 0$, $\alpha \in \Pi$ and X_α a root vector for α, where

$$\Delta^+ = \left\{e_i \pm e_j, \text{ with } 1 \leq i < j \leq p\right\} \cup \{e_i, 1 \leq i \leq p\},$$

$$\Pi = \{a_i = e_i - e_{i+1}, \ 1 \leq i \leq p-1\} \cup \left\{a_p = e_p\right\}.$$

The corresponding root vectors are

$$X_{e_i-e_j} = E_{i,j} - E_{p+j,p+i}$$
$$X_{e_i+e_j} = E_{i,p+j} - E_{j,p+i}$$
$$X_{e_i} = E_{i,2p+1} - E_{2p+1,p+i}$$
$$X_{-e_i} = E_{p+i,2p+1} - E_{2p+1,i}.$$

Now, let $x := \sum_{i=1}^{p} 2\,(p+1-i)\left(E_{i,i} - E_{p+i,p+i}\right)$ and take $e := \sum_{a\in\Pi} X_a$. From this choice it is then satisfied that $[x, e] = 2e$, for the semisimple element x and the regular nilpotent element e. Moreover, the conditions $\left[x, \tilde{e}\right] = -2\tilde{e}$ and $\left[e, \tilde{e}\right] = x$ determine another nilpotent element $\tilde{e}$, thus the triple $\langle x, e, \tilde{e}\rangle \cong \mathfrak{sl}\,(2, \mathbb{C})$ defines a principal 3-dimensional Lie subalgebra of $\mathfrak{so}\,(p, p+1)$.

The adjoint action $\langle x, e, \tilde{e}\rangle \cong \mathfrak{so}\,(2, \mathbb{C}) \rightarrow \mathrm{End}\,(\mathfrak{so}\,(2p+1, \mathbb{C}))$ of this subalgebra decomposes $\mathfrak{so}\,(p, p+1)$ as a direct sum of irreducible representations

$$(2p+1, \mathbb{C}) = \bigoplus_{i=1}^{p} V_i,$$

with $\dim V_i = 4i-1$, for $1 \leq i \leq p$. Therefore, $V_i = S^{4i-2}\mathbb{C}^2$, $1 \leq i \leq p$ with eigenvalues $4i-2$, $4i-4$,..., $-4i+4$, $-4i+2$ for the action of $\mathrm{ad}x$, and the highest weight vectors are $e_1, \ldots, e_p$, where e_i has eigenvalue $4i-2$, for $1 \leq i \leq p$.

Considering the representation

$$\mathfrak{sl}\,(2, \mathbb{C}) \rightarrow \mathfrak{so}\,(2p+1, \mathbb{C}),$$

for $\mathfrak{so}\,(2p+1, \mathbb{C}) = S^2\mathbb{C}^2 + S^6\mathbb{C}^2 + \ldots + S^{4p-2}\mathbb{C}^2 = \Lambda^2\left(S^{2p}\mathbb{C}^2\right)$, we may next deduce the defining data (E_1, Φ_1) for a parabolic $\mathrm{SO}\,(p, p+1)$-Higgs bundle inside the parabolic Teichmüller component for the split real form $G^r = \mathrm{SO}_0\,(p, p+1)$. The parabolic vector bundle is obtained from the $(2p)$-th symmetric power of the parabolic $\mathrm{SL}\,(2, \mathbb{R})$-bundle in the Teichmüller component, as follows.

Let X_1 be a compact Riemann surface of genus g_1, $D_1 = \{p_1, \ldots, p_s\}$ a collection of s distinct points on X_1 and let $L_1 \rightarrow X_1$ with $L_1^2 \cong K_{X_1}$ and $\iota_1 = \mathcal{O}_{X_1}(D_1)$. Consider the parabolic vector bundle $(L_1 \otimes \iota_1)^* \oplus L_1$ over (X_1, D_1), equipped with a trivial flag and weight $\frac{1}{2}$. Then, the vector bundle E_1 of a model parabolic $\mathrm{SO}\,(p, p+1)$-Higgs bundle in the parabolic Teichmüller component is

$$\begin{aligned}
E_1 :&= S^{2p}\left((L_1 \otimes \iota_1)^* \oplus L_1\right)\\
&= L_1^{-2p} \otimes \mathcal{O}\,(-pD_1) \oplus L_1^{-2p+2} \otimes \mathcal{O}\,((1-p)\,D_1) \oplus \ldots\\
&\ldots \oplus L_1^{2p-2} \otimes \mathcal{O}\,((p-1)\,D_1) \oplus L_1^{2p} \otimes \mathcal{O}\,(pD_1)\\
&= K_1^{-p} \otimes \mathcal{O}\,(-pD_1) \oplus K_1^{-(p-1)} \otimes \mathcal{O}\,((1-p)\,D_1) \oplus \ldots\\
&\ldots \oplus K_1^{p-1} \otimes \mathcal{O}\,((p-1)\,D_1) \oplus K_1^{p} \otimes \mathcal{O}\,(pD_1),
\end{aligned}$$

equipped with a trivial parabolic flag and weight 0.

Remark 6.3 Note that in the above description we have included the consideration for the parabolic structure in a symmetric power of a parabolic bundle. In fact, restricting attention on the first original term $(L_1 \otimes \iota_1)^*$ with weight $\frac{1}{2}$, the symmetric power $S^{2p}\left((L_1 \otimes \iota_1)^*\right)$ is the line bundle $L_1^{-2p} \otimes \mathcal{O}(-2pD_1)$ with weight $2p \cdot \frac{1}{2} = p$. However, we obtain a well-defined parabolic bundle by reducing the weight to a number within the interval $[0, 1)$, this means, by tensoring $L_1^{-2p} \otimes \mathcal{O}(-2pD_1)$ by $\mathcal{O}(pD_1)$. We thus get $K_1^{-p} \otimes \mathcal{O}(-pD_1)$ with weight 0, as appears in the first term of the parabolic bundle E_1 above.

The Higgs field in the parabolic $\mathrm{SO}(p, p+1)$-Teichmüller component is given by

$$\tilde{e} + q_1 e_1 + \ldots + q_p e_p,$$

for $\left(q_1, \ldots, q_p\right) \in \bigoplus_{i=1}^{p} H^0\left(K_1^{2i} \otimes \iota_1^{2i-1}\right)$ and $e_1, \ldots, e_p$ are the highest weight vectors. From the set of simple roots of $\mathfrak{so}(p, p+1)$,

$$\Pi = \{e_i - e_{i+1},\ 1 \leq i \leq p-1\} \cup \left\{e_p\right\},$$

we obtain the 3-dimensional subalgebra $\langle x, e, \tilde{e}\rangle \cong \mathfrak{sl}(2, \mathbb{C}) \hookrightarrow \mathfrak{so}(p, p+1)$, with

$$x = \left(\begin{array}{cccc|cccc|c} 2p & & & & & & & & \\ & 2(p-1) & & & & & & & \\ & & \ddots & & & & & & \\ & & & 2 & & & & & \\ \hline & & & & -2p & & & & \\ & & & & & -2(p-1) & & & \\ & & & & & & \ddots & & \\ & & & & & & & -2 & \\ \hline & & & & & & & & 0 \end{array}\right), \tag{6.8}$$

$$e = \left(\begin{array}{cccc|cccc|c} 0 & 1 & & & & & & & \\ & \ddots & \ddots & & & & & & \\ & & & 1 & & & & & \\ & & & 0 & & & & & 1 \\ \hline & & & & 0 & & & & \\ & & & & -1 & \ddots & & & \\ & & & & & \ddots & & & \\ & & & & & & -1 & 0 & \\ \hline & & & & & & & -1 & 0 \end{array}\right), \tag{6.9}$$

the semisimple and regular nilpotent element respectively; from these we may also determine the third element in the principal 3-dimensional subalgebra of $\mathfrak{so}(p, p+1)$:

$$\tilde{e} = \left(\begin{array}{c|c|c} A & & \\ \hline & B & D \\ \hline C & & \end{array}\right), \tag{6.10}$$

where

$$A = \begin{pmatrix} 0 & & & & \\ 2p & \ddots & & & \\ & 2p+2(p-1) & & & \\ & & \ddots & & \\ & & & 2p+2(p-1)+\cdots+2\cdot 2 & 0 \end{pmatrix}$$

is a $p \times p$ block with zeros on the main diagonal,

$$B = \begin{pmatrix} 0 & -2p & & & \\ & \ddots & -2p-2(p-1) & & \\ & & & \ddots & \\ & & & & -2p-2(p-1)-\cdots-2\cdot 2 \\ & & & & 0 \end{pmatrix}$$

is a $p \times p$ block with zeros on the main diagonal, and

$$C = \begin{pmatrix} 0 \cdots 0 \; 2p+2(p-1)+\cdots+2 \end{pmatrix} \text{ is a } 1 \times p \text{ block},$$

$$D = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -2p-2(p-1)-\cdots-2 \end{pmatrix} \text{ is a } p \times 1 \text{ block.}$$

From the analysis above we deduce that a model parabolic Higgs pair lying inside the parabolic $\mathrm{SO}_0(p, p+1)$-Hitchin component which is a local minimum of the Hitchin functional, when viewed as an $\mathrm{SL}(2p+1, \mathbb{C})$-pair, is a pair (E_1, Φ_1) with

- $E_1 = K_1^{-p} \otimes \mathcal{O}(-pD_1) \oplus K_1^{-(p-1)} \otimes \mathcal{O}((1-p)D_1) \oplus \ldots \oplus K_1^{(p-1)} \otimes \mathcal{O}((p-1)D_1) \oplus K_1^{p} \otimes \mathcal{O}(pD_1)$
  a parabolic vector bundle of rank $2p+1$ over (X_1, D_1) equipped with a parabolic structure given by a trivial flag and weight 0,

- $\Phi_1 = \begin{pmatrix} 0 & 1 & 0 \cdots & & 0 \\ 0 & 0 & 1 & 0 & \cdots 0 \\ \vdots & & & & \\ 0 \cdots & & & 0 & 1 \end{pmatrix} : E_1 \to E_1 \otimes K_1 \otimes \iota_1$
  as a $p \times (p+1)$-matrix.

The next lemma is analogous to Lemma 2.1 in [12].

Lemma 6.4 *The parabolic Higgs bundle* (E_1, Φ_1) *above is a parabolic stable Higgs bundle of parabolic degree zero.*

Proof The proof that $\text{pardeg}\,(E_1) = 0$ is immediate, following the properties of the parabolic degree on a direct sum and the dual of a parabolic bundle. The Φ_1-invariant proper subbundles of E_1 are of the form

$$K_1^{-p} \otimes \mathcal{O}(-pD_1) \oplus K_1^{-(p-1)} \otimes \mathcal{O}(-(p-1)\,D_1) \oplus \ldots \oplus K_1^{m-p} \otimes \mathcal{O}((m-p)\,D_1)\,,$$

for $0 \leq m \leq 2p-1$. One now checks that these all have negative parabolic degree, that is, strictly less than $\text{pardeg}\,(E_1)$.

Therefore, from the punctured-surface version of the non-abelian Hodge correspondence [97], there is a tame harmonic metric on the vector bundle E_1. Let A_1 denote the associated Chern connection. Parabolic stability implies the existence of a complex gauge transformation, unique up to modification by a unitary gauge, such that (A_1, Φ_1) solves the Hitchin equations.

In a suitably chosen local holomorphic trivialization of E_1, the pair (A_1, Φ_1) is asymptotic to a model solution, which after a unitary change of frame can be written locally over a punctured neighborhood around a point $p_i \in D_1$ as

$$A_1^{\text{mod}} = 0, \quad \Phi_1^{\text{mod}} = Cx\frac{dz}{z},$$

where x denotes the semisimple element from (6.8) and z the local coordinate around the point $p_i \in D_1$.

6.7.4.2 Models via the General Map Ψ

Let X_2 be a compact Riemann surface of genus g_2 and $D_2 = \{q_1, \ldots, q_s\}$ a collection of s points on X_2. Let $\iota_2 = \mathcal{O}_{X_2}(D_2)$. The second family of model parabolic $\text{SO}\,(p, p+1)$-Higgs bundles is obtained via the more general map

$$\Psi^{par} : \mathcal{M}^{par}_{K_2^p \otimes \iota_2^{p-1}}(\text{SO}\,(1,2)) \times \bigoplus_{j=1}^{p-1} H^0\left(X_2, K_2^{2j} \otimes \iota_2^{2j-1}\right) \to \mathcal{M}^{par}(\text{SO}\,(p, p+1))$$

defined as in (6.5), but considering also the relevant parabolic structures. Take $\left(I, \hat{W}, \hat{\eta}\right) \in \mathcal{M}^{par}_{K_2^p \otimes \iota_2^{p-1}}(\mathrm{SO}(1,2))$, the moduli space of K_2^p-twisted parabolic $\mathrm{SO}(1,2)$-Higgs bundles, for

- $\hat{W} := \tilde{M} \oplus \tilde{M}^\vee$, for $\tilde{M} \cong \mathcal{O}((2k-1-p)\,D_2)$ with $k = 1, \ldots, p$ an integer;
- $I := \wedge^2_{par} \hat{W} \cong \wedge \tilde{M} \otimes \wedge \tilde{M}^\vee \cong \tilde{M} \otimes \tilde{M}^\vee \cong \mathcal{O}$;
- $\hat{\eta} = 0$.

Then, one gets by the definition of the map Ψ^{par} the triple $\Psi^{par}\left(\left(I, \hat{W}, \hat{\eta}\right), (0, \ldots, 0)\right) =: (V, W, \eta)$, where

- $V = K_2^{p-1} \otimes \mathcal{O}((p-1)\,D_2) \oplus \ldots \oplus K_2^{1-p} \otimes \mathcal{O}((1-p)\,D_2)$;
- $W = \tilde{M} \oplus \tilde{M}^\vee \oplus K_2^{p-2} \otimes \mathcal{O}((p-2)\,D_2) \oplus \ldots \oplus K_2^{2-p} \otimes \mathcal{O}((2-p)\,D_2)$;
- $\eta = \begin{pmatrix} \hat{\eta} = 0 & 0 & \cdots & & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \\ 0 & \cdots & & 0 & 1 \end{pmatrix}$.

From the description of the Higgs bundle data we see that since $\hat{\eta} = 0$, the triple (V, W, η) reduces to an $\mathrm{SO}(p, p-1) \times \mathrm{SO}(2)$-Higgs bundle whose $\mathrm{SO}(p, p-1)$-factor lies in the parabolic Hitchin component. We rather define this as an $\mathrm{SL}(2p+1, \mathbb{C})$-pair (E_2, Φ_2), where

- $E_2 = V \oplus W = \tilde{M} \oplus \tilde{M}^\vee \oplus K_2^{-(p-1)} \otimes \mathcal{O}((1-p)\,D_2) \oplus \ldots \oplus K_2^{p-1} \otimes \mathcal{O}((p-1)\,D_2)$;
- $\Phi_2 = \begin{pmatrix} 0 & 0 & 0 & \cdots & & & & 0 \\ 0 & 0 & 0 & \cdots & & & & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & & 0 \\ \vdots & \vdots & \vdots & 0 & 1 & 0 & \cdots & 0 \\ & & & & & \ddots & & \\ & & & & & & & 1 \\ 0 & 0 & 0 & \cdots & & & & 0 \end{pmatrix} : E_2 \to E_2 \otimes K_2 \otimes \iota_2$.

The Φ_2-invariant proper subbundles of E_2 are

$$\tilde{M} \oplus \tilde{M}^\vee \oplus K_2^{-(p-1)} \otimes \mathcal{O}((1-p)\,D_2)$$
$$\tilde{M} \oplus \tilde{M}^\vee \oplus K_2^{-(p-1)} \otimes \mathcal{O}((1-p)\,D_2) \oplus K_2^{-(p-2)} \otimes \mathcal{O}((2-p)\,D_2)$$
$$\vdots$$
$$\tilde{M} \oplus \tilde{M}^\vee \oplus K_2^{-(p-1)} \otimes \mathcal{O}((1-p)\,D_2) \oplus \ldots \oplus K_2^{(p-2)} \otimes \mathcal{O}((p-2)\,D_2)\,,$$

or, in general, these are of the form

$$\tilde{M} \oplus \tilde{M}^{\vee} \oplus K_2^{-(p-1)} \otimes \mathcal{O}\left((1-p)\, D_2\right) \oplus \ldots \oplus K_2^{(l-p)} \otimes \mathcal{O}\left((l-p)\, D_2\right),$$

for each $1 \leq l \leq 2p-2$. As in the previous lemma, one sees that all proper Φ_2-invariant subbundles of E_2 have negative parabolic degree, while $\text{pardeg}\,(E_2) = 0$. Therefore, the models (E_2, Φ_2) for every $k = 1, \ldots, p$ are all parabolic stable. For A_2 be the Chern connection with respect to a tame harmonic metric on E_2, in a suitably chosen local holomorphic trivialization of E_2, the pair (A_2, Φ_2) is, after conjugation by a unitary gauge, asymptotic to a model solution which locally over a punctured neighborhood around a point $q_j \in D_2$ is written as

$$A_2^{\text{mod}} = 0, \;\; \Phi_2^{\text{mod}} = \begin{pmatrix} 0 &&&&&&&& \\ & 0 &&&&&&& \\ && 2\,(p-1)\,C &&&&&& \\ &&& \ddots &&&&& \\ &&&& 2C &&&& \\ &&&&& -2(p-1)C &&& \\ &&&&&& \ddots && \\ &&&&&&& -2C & \\ &&&&&&&& 0 \end{pmatrix} \frac{dw}{w},$$

for coordinates w around each puncture $q_j \in D_2$.

6.7.5 Gauge-Theoretic Gluing of Parabolic SO $(p, p+1)$-Higgs Bundles

We have described parabolic SO $(p, p+1)$-models (E_i, Φ_i), $i = 1, 2$, which are parabolic stable. Model solutions to the Hitchin equations over punctured disks corresponding to the pairs (E_i, Φ_i) are respectively of the form

$$A_1^{\text{mod}} = 0, \;\; \Phi_1^{\text{mod}} = \begin{pmatrix} 2pC &&&&&&& \\ & 2(p-1)C &&&&&& \\ && \ddots &&&&& \\ &&& 2C &&&& \\ &&&& -2pC &&& \\ &&&&& \ddots && \\ &&&&&& -2C & \\ &&&&&&& 0 \end{pmatrix} \frac{dz}{z},$$

$$A_2^{\text{mod}} = 0, \ \Phi_2^{\text{mod}} = \begin{pmatrix} 0 & & & & & & & & \\ & 0 & & & & & & & \\ & & 2(p-1)C & & & & & & \\ & & & \ddots & & & & & \\ & & & & 2C & & & & \\ & & & & & -2(p-1)C & & & \\ & & & & & & \ddots & & \\ & & & & & & & -2C & \\ & & & & & & & & 0 \end{pmatrix} \frac{dw}{w}.$$

In order to glue the above parabolic $\mathrm{SO}(p, p+1)$-Higgs bundles over the complex connected sum of Riemann surfaces $X_\# := X_1 \# X_2$ of genus $g = g_1 + g_2 + s - 1$ we shall use the gauge-theoretic gluing construction summarized in Sect. 6.6. To this end, the initial model data $\left(A_i^{\text{mod}}, \Phi_i^{\text{mod}}\right)$ should be identified locally over the annuli around the points in the divisors of s-many points D_i, for $i = 1, 2$. This is achieved using the perturbation argument described next.

Consider the embedding

$$\Psi_i^{par} : \mathcal{M}^{par}_{K_i^p \otimes \iota_i^{p-1}}(\mathrm{SO}(1,2)) \times \bigoplus_{j=1}^{p-1} H^0\left(X_i, K_i^{2j} \otimes \iota_i^{2j-1}\right) \to \mathcal{M}^{par}(\mathrm{SO}(p, p+1)),$$

for $i = 1, 2$. Over the pair (X_1, D_1), take a parabolic $\mathrm{SO}(1,2)$-Higgs bundle defined by the triple $\left(\hat{W}_1, I_1, \hat{\eta}_1\right)$ with

$$\begin{aligned} \hat{W}_1 &:= K_1^\vee \oplus K_1; \\ I_1 &\cong \mathcal{O}; \\ \hat{\eta}_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Let $\left(\tilde{A}_1, \tilde{\Phi}_1\right)$ be the corresponding solution to the Hitchin equations. There is a complex gauge transformation which locally puts $\left(\tilde{A}_1, \tilde{\Phi}_1\right)$ into the model form

$$\tilde{A}_1^{\text{mod}} = 0, \ \tilde{\Phi}_1^{\text{mod}} = \begin{pmatrix} 2C & 0 & 0 \\ 0 & -2C & 0 \\ 0 & 0 & 0 \end{pmatrix} \frac{dz}{z}$$

over a disk centered at the points in D_1 with coordinates z.

Remark 6.4 The existence of this complex gauge transformation is provided for the local $\mathrm{SL}(2,\mathbb{R})$-model solution $\left(A^{\text{mod}}=0,\ \Phi^{\text{mod}}=\begin{pmatrix} C & 0 \\ 0 & -C \end{pmatrix}\frac{dz}{z}\right)$; then we embed into $\mathrm{SO}(1,2)$.

Using the map Ψ_1^{par} above, the parabolic stable $\mathrm{SO}(p,p+1)$-pair (E_1,Φ_1) over (X_1,D_1) corresponds to an *approximate solution* (A_1,Φ_1) of the Hitchin equations, which near each point of D_1 has the form $\left(A_1^{\text{mod}},\Phi_1^{\text{mod}}\right)$ with

$$A_1^{\text{mod}}=0,\quad \Phi_1^{\text{mod}}=\begin{pmatrix} 2pC & & & & & & & \\ & 2(p-1)C & & & & & & \\ & & \ddots & & & & & \\ & & & 2C & & & & \\ & & & & -2pC & & & \\ & & & & & \ddots & & \\ & & & & & & -2C & \\ & & & & & & & 0 \end{pmatrix}\frac{dz}{z},$$

for $p>2$ and $C\in\mathbb{R}$ nonzero.

Over the pair (X_2,D_2), take the triple $\left(\hat{W}_2,I_2,\hat{\eta}_2\right)$ with

$$\hat{W}_2:=\tilde{M}\oplus\tilde{M}^{\vee},\ \text{where } \tilde{M}\cong\mathcal{O}\left((2k-1-p)\,D_2\right),\ \text{for } k=1,\dots,p$$
$$I_2\cong\mathcal{O}$$
$$\hat{\eta}_2\in H^0\left(\mathrm{Hom}\left(\hat{W}_2,I_2\right)\otimes K_2^p\otimes\iota_2^{p-1}\right).$$

Applying a similar argument as above, we may perturb the relevant $\mathrm{SL}(2,\mathbb{R})$-pair and extend our data to $\mathrm{SO}(p,p+1)$ to finally get an *approximate* solution (A_2,Φ_2), which near each point of D_2 has the form $\left(A_2^{\text{mod}},\Phi_2^{\text{mod}}\right)$ with

$$A_2^{\text{mod}}=0,\quad \Phi_2^{\text{mod}}=\begin{pmatrix} -2pC & & & & & & & \\ & -2(p-1)C & & & & & & \\ & & \ddots & & & & & \\ & & & -2C & & & & \\ & & & & 2pC & & & \\ & & & & & \ddots & & \\ & & & & & & 2C & \\ & & & & & & & 0 \end{pmatrix}\frac{dw}{w},$$

for $p>2$ and $C\in\mathbb{R}$ nonzero, as above.

The complex connected sum of Riemann surfaces $X_\# = X_1 \# X_2$ is realized along the curve $zw = \lambda$ for a parameter $\lambda \in \mathbb{C}$, and so $\frac{dz}{z} = -\frac{dw}{w}$ for coordinates on annuli around each puncture which are glued using a biholomorphism for each pair of points $\left(p_i, q_j\right)$ from the divisors D_1 and D_2. Let $\Omega \subset X_\#$ denote the result of gluing these pairs of annuli and set $\left(A^{\text{mod}}_{p_i,q_j}, \Phi^{\text{mod}}_{p_i,q_j}\right) := \left(A_1^{\text{mod}}, \Phi_1^{\text{mod}}\right) = -\left(A_2^{\text{mod}}, \Phi_2^{\text{mod}}\right)$. We can glue the pairs (A_1, Φ_1), (A_2, Φ_2) together to get an *approximate solution* of the $\mathrm{SO}\,(p, p+1)$-Hitchin equations:

$$\left(A^{app}, \Phi^{app}\right) := \begin{cases} (A_1, \Phi_1)\,, & \text{over } X_1 \backslash X_2 \\ \left(A^{\text{mod}}_{p_i,q_j}, \Phi^{\text{mod}}_{p_i,q_j}\right), & \text{over } \Omega \text{ around each pair of points } \left(p_i, q_j\right) \\ (A_2, \Phi_2) & \text{over } X_2 \backslash X_1, \end{cases}$$

over the connected sum bundle over $X_\#$.

By construction, (A^{app}, Φ^{app}) is a smooth pair on $X_\#$, complex gauge equivalent to an exact solution of the Hitchin equations by a smooth gauge transformation defined over all of $X_\#$. The next step is to correct the approximate solution (A^{app}, Φ^{app}) to an exact solution of the $\mathrm{SO}\,(p, p+1)$-Hitchin equations. We follow the contraction mapping argument for the nonlinear G-Hitchin operator from Sects. 6.6.6 and 6.6.7 developed for a general connected semisimple Lie group G. We next describe how the general theory for showing that the linearization operator is invertible adapts to the case when $G = \mathrm{SO}\,(p, p+1)$; the computation of the necessary analytic estimates for an approximate solution does not depend on the semisimple Lie group G and can be found in §6 of [66].

For the group $G = \mathrm{SO}(p, p+1)$, a maximal compact subgroup is $H = \mathrm{SO}(p, \mathbb{C}) \times \mathrm{SO}(p+1, \mathbb{C})$ with Lie algebra $\mathfrak{h} = \mathfrak{so}(p) \times \mathfrak{so}(p+1)$. Moreover, for a Higgs field $\Phi = \varphi dz$, the compact real form $\tau : \mathfrak{g}^{\mathbb{C}} \to \mathfrak{g}^{\mathbb{C}}$ is giving $\tau(\Phi) = \bar{\varphi} dz$. For the notation introduced in Sect. 6.6.7, we have the following:

Lemma 6.5 *For an element* $(\psi_1, \psi_2) \in \ker\left(L_1 + L_2^*\right) \cap L^2_{ext}\left(X_\#^\times\right)$, *we have*

$$d_{A_0^{app}} \psi_i = \left[\psi_i, \Phi_0^{app}\right] = \left[\psi_i, \left(\Phi_0^{app}\right)^*\right] = 0,$$

for $i = 1, 2$.

Proof The proof follows exactly the same steps as that of Lemma 6.5 in [66]. There are no nontrivial off-diagonal elements in the kernel of the operator

$$D\,(\psi_1, \psi_2) := 2 \begin{pmatrix} \frac{i}{2} \partial_\theta \psi_1 + [\psi_2, \tau\,(\varphi)] \\ -\frac{i}{2} \partial_\theta \psi_2 - [\psi_1, \varphi] \end{pmatrix},$$

since we have taken a diagonal and traceless model Higgs field $\Phi^{\text{mod}} = \varphi^{\text{mod}} \frac{dz}{z}$, where

$$\varphi^{\text{mod}} = \begin{pmatrix} 2pC & & & & & & & \\ & 2(p-1)C & & & & & & \\ & & \ddots & & & & & \\ & & & 2C & & & & \\ & & & & -2pC & & & \\ & & & & & \ddots & & \\ & & & & & & -2C & \\ & & & & & & & 0 \end{pmatrix},$$

with $p > 2$ and $C \in \mathbb{R}$ nonzero.

The next proposition is now providing invertibility:

Proposition 6.4 *The operator $L_1 + L_2^*$ considered as a densely defined operator on $L^2_{ext}\left(X^\times_\#\right)$ has trivial kernel.*

Proof Let $(\psi_1, \psi_2) \in \ker\left(L_1 + L_2^*\right) \cap L^2_{\text{ext}}\left(X^\times_\#\right)$. From Lemma 6.5 we have

$$d_{A_0^{app}} \psi_i = \left[\psi_i, \Phi_0^{app}\right] = \left[\psi_i, \left(\Phi_0^{app}\right)^*\right] = 0,$$

for $i = 1, 2$. We show that $\psi_1 = 0$ by showing separately that $\gamma := \psi_1 + \psi_1^* \in \Omega^0\left(X^\times_\#, E\left(\mathfrak{h}\right)\right)$ and $\delta := i\left(\psi_1 - \psi_1^*\right) \in \Omega^0\left(X^\times_\#, E\left(\mathfrak{h}\right)\right)$ both vanish.

For holomorphic coordinate z centered at the node of $X^\times_\#$, the Higgs field Φ_0^{app} in our exact solution is written $\Phi_0^{app} = \varphi \frac{dz}{z}$ with

$$\varphi \in \mathfrak{m}^{\mathbb{C}}\left(\text{SO}\left(p, p+1\right)\right) = \left\{\begin{pmatrix} 0 & Q \\ -Q^T & 0 \end{pmatrix}, \text{ for } M \text{ a } p \times (p+1) \text{ complex matrix}\right\}.$$

We get that $d|\gamma|^2 = 2\left\langle d_{A_0^{app}}\gamma, \gamma\right\rangle = 0$, that is, $|\gamma|$ is constant on $X^\times_\#$, as well as that $\gamma(x)$ lies in the kernel of the linearization operator (see Lemma 6.3). Moreover,

$$\gamma\left(x\right) \in \mathfrak{h} = \left\{\begin{pmatrix} M & 0 \\ 0 & N \end{pmatrix} | M \in \mathfrak{so}\left(p\right), N \in \mathfrak{so}\left(p+1\right)\right\},$$

thus $\gamma(x)$ has orthogonal eigenvectors for distinct eigenvalues, but even if there are degenerate eigenvalues, it is still possible to find an orthonormal basis consisting of eigenvectors of $\gamma\left(x\right)$.

Now, if $\gamma\left(x\right)$ is non-zero, since $\left[\varphi\left(x\right), \gamma\left(x\right)\right] = 0$ it follows that $\varphi\left(x\right)$ preserves the eigenspaces of $\gamma\left(x\right)$ for all $x \in X^\times_\#$ and so $\left\langle\varphi\left(x\right)v, \varphi\left(x\right)w\right\rangle = \left\langle v, w\right\rangle$ for

$v, w \in \mathbb{C}^{2p+1}$. In other words, $\varphi(x)$ ought to be an isometry with respect to the usual norm in $\mathbb{C}^{2p+1}$. Equivalently, $\varphi(x)$ is unitary for all $x \in X_\#^\times$. The determinant of the Higgs field $\det \Phi_0^{app}$ generically has a simple zero in at least one point in $X_\#^\times$. For a zero x_0 chosen, say, on the left hand side surface X_1 of $X_\#^\times$ where we embed via the irreducible representation described in Sect. 6.7.4.1—let us denote it at present ϕ_{irr}—we see that

$$\varphi(x_0) = \phi_{irr*}\begin{pmatrix} 0 & z \\ 1 & 0 \end{pmatrix} = \tilde{e} + ze$$

which is not unitary, for the matrices e and $\tilde{e}$ as in (6.9) and (6.10) respectively. This is a contradiction and therefore, $\gamma = 0$ everywhere.

That δ vanishes, as well as that $\psi_2 = 0$, is proven entirely similarly.

Remark 6.5 The assumption of the existence of at least one simple zero of a generic meromorphic quadratic differential allows us to show that the linear operator $L_{(A,\Phi)}$ is injective and thus assure absence of small eigenvalues of this linear operator governing the gluing construction (cf. Swoboda [99] for a similar application). That a generic solution of the rank 2 Hitchin equations has only simple zeroes is proven in [77].

Theorem 6.12 adapts in the case $G = \mathrm{SO}(p, p+1)$ to provide the following:

Theorem 6.14 *Let X_1 be a closed Riemann surface of genus g_1 and $D_1 = \{p_1, \ldots, p_s\}$ a collection of s distinct points on X_1. Consider respectively a closed Riemann surface X_2 of genus g_2 and a collection of also s distinct points $D_2 = \{q_1, \ldots, q_s\}$ on X_2. Let $(E_1, \Phi_1) \to X_1$ and $(E_2, \Phi_2) \to X_2$ be parabolic polystable $SO(p, p+1)$-Higgs bundles, one from each of the families described in Sects. 6.7.4.1 and 6.7.4.2 with corresponding solutions to the Hitchin equations (A_1, Φ_1) and (A_2, Φ_2). Then there is a polystable $SO(p, p+1)$-Higgs bundle $(E_\#, \Phi_\#) \to X_\#$ over the complex connected sum of Riemann surfaces $X_\# = X_1 \# X_2$, which agrees with the initial data over $X_\# \setminus X_1$ and $X_\# \setminus X_2$.*

Definition 6.19 We call such an $\mathrm{SO}(p, p+1)$-Higgs bundle constructed by the theorem above a *hybrid* $\mathrm{SO}(p, p+1)$*-Higgs bundle*.

6.7.6 Model Representations in the Exceptional Components of $\mathcal{R}(SO(p, p+1))$

We now show that the specific hybrid $\mathrm{SO}(p, p+1)$-Higgs bundles constructed in the previous section lie inside the $p(2g-2) - 1$ exceptional components of the character variety $\mathcal{R}(\mathrm{SO}(p, p+1))$. In fact, by varying the parameters in the construction, namely, the genera g_1, g_2 of the Riemann surfaces X_1, X_2, the number of points s in the divisors D_1, D_2, and the weight $\alpha = 2k - 1 - p$ for the line bundle

$\tilde{M} \cong \mathcal{O}((2k-1-p)D_2)$, one obtains models in *all exceptional components*. This is seen by an explicit computation of the degree of the line bundle M appearing in the description (6.7) of the Higgs bundle data; the exceptional components are fully distinguished by the degree of this line bundle. We have considered:

$$E_1 = K_1^{-p} \otimes \mathcal{O}(-pD_1) \oplus K_1^{-(p-1)} \otimes \mathcal{O}((1-p)D_1) \oplus \dots$$
$$\dots \oplus K_1^{(p-1)} \otimes \mathcal{O}((p-1)D_1) \oplus K_1^{p} \otimes \mathcal{O}(pD_1), \text{ and}$$

$$E_2 = V \oplus W = \tilde{M}^{\vee} \oplus \tilde{M} \oplus K_2^{-(p-1)} \otimes \mathcal{O}((1-p)D_2) \oplus \dots$$
$$\dots \oplus K_2^{p-1} \otimes \mathcal{O}((p-1)D_2),$$

with $\tilde{M} \cong \mathcal{O}((2k-1-p)D_2)$ and $\operatorname{pardeg}\left(\tilde{M}\right) = (2k-1-p)s$, for $k = 1, \dots, p$. We now use Proposition 6.2, which asserts an additivity property for the parabolic degree of the bundle over the connected sum operation. We thus have that for each $j \in \{1-p, \dots, p-1\}$ the bundle $K_1^{\otimes par\, j} \# K_2^{\otimes par - j}$ has degree

$$\begin{aligned}
\deg\left(K_1^{\otimes par\, j} \# K_2^{\otimes par\, j}\right) &= \operatorname{pardeg}\left(K_1^{j} \otimes \mathcal{O}(jD_1)\right) + \operatorname{pardeg}\left(K_2^{j} \otimes \mathcal{O}(jD_2)\right) \\
&= j\,(2g_1 - 2 + s) + j\,(2g_2 - 2 + s) \\
&= 2j\,(g_1 + g_2 + s - 1 - 1) \\
&= 2j\left(g_{X_\#} - 1\right) \\
&= \deg K_{X_\#}^{\otimes j}.
\end{aligned}$$

It is thus a line bundle isomorphic to $K_{X_\#}^{\otimes j}$.

Moreover, gluing the parabolic line bundles $K_1^{p} \otimes \mathcal{O}(pD_1)$ and $\tilde{M}$ provides a line bundle $M \in \operatorname{Pic}(X_\#)$ with degree

$$\begin{aligned}
\deg(M) &= \operatorname{pardeg}\left(K_1^{p} \otimes \mathcal{O}(pD_1)\right) + \operatorname{pardeg}\left(\tilde{M}\right) \\
&= p\,(2g_1 - 2 + s) + (2k-1-p)s \\
&= 2p\,(g_1 - 1) + (2k-1)\,s.
\end{aligned}$$

We deduce that the result of the construction is a Higgs bundle (V, W_k, η) with data V and η as in (6.7) and

$$W_k := M \oplus K_{X_\#}^{p-2} \oplus \dots \oplus K_{X_\#}^{2-p} \oplus M^{-1}$$

with $d = \deg(M) = 2p(g_1 - 1) + (2k - 1)s$, for $k = 1, \ldots, p$. One can now check that varying the values of the parameters g_1, s and k, we can obtain model $\mathrm{SO}(p, p+1)$-Higgs bundles by gluing, which exhaust all the exceptional smooth $p(2g-2) - 1$ components of $\mathcal{M}(\mathrm{SO}(p, p+1))$.

Remark 6.6 Notice that the case when $p = 1$ actually describes the $\mathrm{Sp}(4, \mathbb{R})$-case from [66]. Indeed, we then have $k = 1$ and so $\tilde{M} \cong \mathcal{O}$ with $d = \deg(M) = 2(g_1 - 1) + s = -\chi(\Sigma_l)$. The case $p > 2$ thus involves an *extra parameter* on the non-trivial line bundle $\tilde{M}$ given by the parabolic structure on a trivial flag.

6.7.7 Model Representations and Positivity

The model $\mathrm{SO}(p, p+1)$-Higgs bundles obtained above are now all Θ-positive. This follows directly from the recent work of Beyrer and Pozzetti [9], who showed that the set of Θ-positive representations is closed in the character variety $\mathcal{R}(\mathrm{SO}(p, q))$, for $p \leq q$. Moreover, Theorem 6.8 asserts that the connected components parameterized by using Higgs bundle methods in [15] consist solely of Θ-positive representations; the exceptional components of Definition 6.18 do, indeed, fall in these cases (see [5]).

A more direct way to show that the models in the exceptional $p(2g-2) - 1$ smooth components of $\mathcal{R}(\mathrm{SO}(p, p+1))$ are Θ-positive, is by gluing the positivity condition at the level of infinity of the fundamental group. In fact, a Hitchin representation into $\mathrm{SO}(p, p+1)$, and a representation which factors through $\mathrm{SO}(p-1, p) \times \mathrm{SO}(2)$ with $\mathrm{SO}(p-1, p)$-factor in the relative Hitchin component, that is, like the ones we chose, are both Θ-positive (see [5, 24]). On the other hand, in [33, pp. 95–100], Fock and Goncharov provide a gluing method for positive local systems on a pair of Riemann surfaces with boundary for the case of split real Lie groups, and so for the group $\mathrm{SO}(p, p+1)$ in particular. This involves the requirement that the monodromies along the two boundary components, as well as the assigned configurations of positive flags coincide; see p. 99 of (loc. cit.) for more details.

Acknowledgements I would like to warmly thank the editors Ken'ichi Ohshika and Athanase Papadopoulos for their kind invitation to contribute to this collective volume. The work included in Sect. 6.7 was supported by the Labex IRMIA of the Université de Strasbourg and was undertaken in collaboration with Olivier Guichard. I acknowledge the Max-Planck-Institut für Mathematik in Bonn for its hospitality and support during the production of this chapter. I am also grateful to an anonymous referee as well as the editors for their careful reading of the manuscript and a number of useful suggestions.

References

1. D. Alessandrini, O. Guichard, E. Rogozinnikov, A. Wienhard, Noncommutative coordinates for symplectic representations. arXiv: 1911.08014 (2019)
2. B.N. Apanasov, Non-triviality of Teichmüller space for Kleinian groups in space, in *Riemann Surfaces and Related Topics*, *Proceedings of the 1978 Stony Brook Conference* (Princeton University Press, 1980), pp. 21–31
3. B.N. Apanasov, A.V. Tetenov, On the existence of non-trivial quasi-conformal deformations of Kleinian groups in space. Soviet Math. Dokl. **19**, 242–245 (1978)
4. M. Aparicio Arroyo, The geometry of $\mathrm{SO}(p, q)$-Higgs bundles. Ph.D. thesis, Universidad de Salamanca, Consejo Superior de Investigaciones Científicas, 2009
5. M. Aparicio-Arroyo, S. Bradlow, B. Collier, O. García-Prada, P.B. Gothen, A. Oliveira, $\mathrm{SO}\,(p, q)$-Higgs bundles and higher Teichmüller components. Invent. Math. **218**(1), 197–299 (2019)
6. L. Bers, F. Gadiner, Fricke spaces. Adv. Math. **62**(3), 249–284 (1986)
7. L. Bessières, G. Besson, M. Boileau, S. Maillot, J. Porti, Geometrization of 3-manifolds, in *EMS Tracts in Mathematics 13* (European Mathematical Society, Zürich, 2010), x+237 pp.
8. J. Beyrer, B. Pozzetti, A collar lemma for partially hyperconvex surface group representations. Trans. Amer. Math. Soc. **374**(10), 6927–6961 (2021)
9. J. Beyrer, B. Pozzetti, Positive surface group representations in $\mathrm{PO}(p, q)$. arXiv: 2106.14725 (2021)
10. O. Biquard, P. Boalch, Wild non-abelian Hodge theory on curves. Compos. Math. **140**(1), 179–204 (2004)
11. O. Biquard, O. García-Prada, I. Mundet i Riera, Parabolic Higgs bundles and representations of the fundamental group of a punctured surface into a real group. Adv. Math. **372**, 107305 (2020)
12. I. Biswas, P. Arés-Gastesi, S. Govindarajan, Parabolic Higgs bundles and Teichmüller spaces for punctured surfaces. Trans. Amer. Math. Soc. **349**(4), 1551–1560 (1997)
13. M. Boileau, J. Porti, Geometrization of 3-orbifolds of cyclic type. Astérisque No. 272, 208 pp. (2001)
14. S. Boyer, Dehn surgery on knots, in *Handbook of Geometric Topology*, ed. by R.J. Daverman, R.B. Sher (North-Holland, 2002), pp. 165–218
15. S. Bradlow, B. Collier, O. García-Prada, P. Gothen, A. Oliveira, A general Cayley correspondence and higher Teichmüller spaces, arXiv: 2101.09377 (2021)
16. M. Bridgeman, Average bending of convex pleated planes in hyperbolic three-space. Invent. Math. **132**, 381–391 (1998)
17. M. Bridgeman, Bounds on the average bending of the convex hull of a Kleinian group. Mich. Math. J. **51**, 363–378 (2003)
18. M. Bridgeman, R. Canary, F. Labourie, A. Sambarino, The pressure metric for Anosov representations. Geom. Funct. Anal. **25**(4), 1089–1179 (2015)
19. M. Bridgeman, R. Canary, A. Sambarino, An introduction to pressure metrics for higher Teichmüller spaces. Ergodic Theory Dyn. Syst. **38**(6), 2001–2035 (2018)
20. M. Burger, A. Iozzi, F. Labourie, A. Wienhard, Maximal representations of surface groups: Symplectic Anosov structures. Pure Appl. Math. Q. **1**(3), Special issue In memory of Armand Borel. Part 2, 543–590 (2005)
21. M. Burger, A. Iozzi, A. Wienhard, Surface group representations with maximal Toledo invariant. Ann. of Math. (2) **172**(1), 517–566 (2010)
22. M. Burger, M.B. Pozzetti, Maximal representations, non-Archimedean Siegel spaces, and buildings. Geom. Topol. **21**(6), 3539–3599 (2019)
23. S. Cappell, R. Lee, E. Miller, Self-adjoint elliptic operators and manifold decompositions. Part I: Low eigenmodes and stretching. Commun. Pure Appl. Math. **49**, 825–866 (1996)
24. B. Collier, $\mathrm{SO}\,(n, n + 1)$-surface group representations and their Higgs bundles. Ann. Sci. Éc. Norm. Supér. (4) **63**(6), 1561–1616 (2020)

25. K. Corlette, Flat G-bundles with canonical metrics. J. Diff. Geom. **28**, 361–382 (1988)
26. H.P. de Saint-Gervais, Uniformisation des surfaces de Riemann, ENS Editions (Lyon, 2010), 544 pp.
27. M. Dehn, Über die Topologie des dreidimensionales Raumes. Math. Ann. **69**, 137–168 (1910)
28. S.K. Donaldson, Twisted harmonic maps and the self-duality equations. Proc. London Math. Soc. (3) **55**, 127–131 (1987)
29. S. Donaldson, P. Kronheimer, *The Geometry of Four-manifolds*. Oxford Math. Monographs (Oxford Science Publications, 1990)
30. W.D. Dunbar, R.G. Meyerhoff, Volumes of hyperbolic 3-orbifolds. Indiana Univ. Math. J. **43**, 611–637 (1994)
31. D.B.A. Epstein, A. Marden, Convex hulls in hyperbolic space, a theorem of Sullivan, and measured pleated surfaces, in *Analytical and Geometric Aspects of Hyperbolic Space*, ed. by D.B.A. Epstein, Warwick and Durham 1984, London Mathematical Society Lecture Note Series 111 (Cambridge University Press, 1987), pp. 113–253
32. F. Fanoni, B. Pozzetti, Basmajian-type inequalities for maximal representations. J. Differential Geom. **116**, 405–458 (2020)
33. V.V. Fock, A. Goncharov, Moduli spaces of local systems and higher Teichmüller theory. Publ. Math. Inst. Hautes Études Sci. **103**, 1–211 (2006)
34. L. Foscolo, A gluing construction for periodic monopoles. Int. Math. Res. Not. IMRN, no. 24, 7504–7550 (2017)
35. R. Fricke, F. Klein, *Vorlesungen über die Theorie der automorphen Funktionen*, Vols. 1,2 (Teubner, Stuttgart, 1986, 1912)
36. O. García-Prada, Higgs bundles and higher Teichmüller spaces, in *Handbook of Teichmúller Theory. Vol. VII*, ed. by A. Papadopoulos. IRMA Lectures in Mathematics and Theoretical Physics, vol. 30 (2020), pp. 239–285
37. O. García-Prada, P.B. Gothen, I. Mundet i Riera, The Hitchin-Kobayashi correspondence, Higgs pairs and surface group representations. arXiv:0909.4487 (2009)
38. W.M. Goldman, The symplectic nature of fundamental groups of surfaces. Adv. Math. **54**(2), 200–225 (1984)
39. W.M. Goldman, Topological components of spaces of representations. Invent. Math. **93**(3), 557–607 (1988)
40. C.McA. Gordon, Dehn surgery on knots, in *Proceedings of the International Congress of Mathematicians, Kyoto 1990* (Springer, Tokyo, 1991), pp. 631–642
41. C.McA. Gordon, Dehn filling: A survey, in *Proceedings of the Mini Semester in Knot Theory, Banach Center, Warsaw 1995*. Banach Center Publications, vol. 42 (Institute of Mathematics, Polish Academy of Sciences, Warszawa, 1998), pp. 129–144
42. P.B. Gothen, Components of spaces of representations and stable triples. Topology **40**(4), 823–850 (2001)
43. O. Guichard, Composantes de Hitchin et répresentations hyperconvexes de groupes de surface. J. Differential Geom. **80**(3), 391–431 (2008)
44. O. Guichard, A. Wienhard, Anosov representations: domains of discontinuity and applications. Invent. Math. **190**(2), 357–438 (2012)
45. O. Guichard, A. Wienhard, Positivity and higher Teichmüller theory, in *Proceedings of the 7th European Congress of Mathematics* (2016)
46. O. Guichard, A. Wienhard, Topological invariants of Anosov representations. J. Topol. **3**(3), 578–642 (2010)
47. O. Guichard, F. Labourie, A. Wienhard, Positivity and representations of surface groups. arXiv: 2106.14584 (2021)
48. W. Haken, Theorie der Normalflächen. Ein Isotopiekriterium fúr den Kreisknoten. Acta Math. **105**, 245–375 (1961)
49. W. Haken, Über das Homöomorphieproblem der 3-Mannigfaltigkeiten. I. Math. Z. **80**, 89–120 (1962)
50. S. He, A gluing theorem for the Kapustin-Witten equations with a Nahm pole. J. Topol. **12**(3), 855–915 (2019)

51. N.J. Hitchin, The self-duality equations on a Riemann surface. Proc. London Math. Soc. (3) **55**, 59–126 (1987)
52. N.J. Hitchin, Lie groups and Teichmüller space. Topology **31**(3), 449–473 (1992)
53. C.D. Hodgson, S.P. Kerckhoff, Universal bounds for hyperbolic Dehn surgery. Ann. of Math. (2) **162**(1), 367–421 (2005)
54. Y. Huang, Z. Sun, McShane identities for higher Teichmüller theory and the Goncharov-Shen potential. arXiv: 1901.02032 (2019)
55. J.H. Hubbard, Teichmüller theory and applications to geometry, topology, and dynamics, vol. 1, Matrix Editions (Ithaca, NY 2006)
56. W. Jaco, U. Oertel, An algorithm to decide if a 3-manifold is a Haken manifold. Topology **23**(2), 195–209 (1984)
57. D. Johnson, J.J. Millson, Deformation spaces associated to compact hyperbolic manifolds, in *Discrete Groups in Geometry and Analysis* (New Haven, CT 1984), Progr. Math., vol. 67 (Birkhäuser, Boston, 1987), pp. 48–106
58. J. Jost, *Compact Riemann Surfaces. An Introduction to Contemporary Mathematics*, 2nd edn. (Springer, Berlin, 2002), xvi+278 pp.
59. R. Kirby, A calculus for framed links. Invent. Math. **45**, 35–56 (1978)
60. H. Konno, Construction of the moduli space of stable parabolic Higgs bundles on a Riemann surface. J. Math. Soc. Japan **45**(2), 253–276 (1993)
61. B. Kostant, The principal three-dimensional subgroup and the Betti numbers of a complex simple Lie group. Amer. J. Math. **81**, 973–1032 (1959)
62. C. Kourouniotis, Deformations of hyperbolic structures. Math. Proc. Cambridge Philos. Soc. **98**(2), 247–261 (1985)
63. C. Kourouniotis, Bending in the space of quasi-Fuchsian structures. Glasgow Math. J. **33**(1), 41–49 (1991)
64. C. Kourouniotis, The geometry of bending quasi-Fuchsian groups, in *Discrete Groups and Geometry*, (Birmingham, 1991), London Math. Soc. Lecture Note Ser. 173 (Cambridge University Press, Cambridge, 1992), pp. 148–164
65. G. Kydonakis, Gluing constructions for Higgs bundles over a complex connected sum. Ph.D. thesis, University of Illinois at Urbana-Champaign, 2018
66. G. Kydonakis, Model Higgs bundles in exceptional components of the Sp(4,$\mathbb{R}$)-character variety. Int. J. Math. **32**(9), Paper No. 2150067, 50 pp. (2021)
67. G. Kydonakis, H. Sun, L. Zhao, Topological invariants of parabolic G-Higgs bundles. Math. Z. **297**(1–2), 585–632 (2021)
68. F. Labourie, Anosov flows, surface groups and curves in projective space. Invent. Math. **165**(1), 51–114 (2006)
69. F. Labourie, Cross ratios, surface groups, PSL $(n, \mathbb{R})$ and diffeomorphisms of the circle. Publ. Math. Inst. Hautes Études Sci. **106**, 139–213 (2007)
70. F. Labourie, G. McShane, Cross ratios and identities for higher Teichmüller-Thurston theory. Duke Math. J. **149**(2), 279–345 (2009)
71. M. Lackenby, R. Meyerhoff, The maximal number of exceptional Dehn surgeries. Invent Math. **191**(2), 341–382 (2013)
72. I. Le, Higher laminations and affine building. Geom. Topol. **20**(3), 1673–1735 (2016)
73. G.-S. Lee, T. Zhang, Collar lemma for Hitchin representations. Geom. Topol. **21**(4), 2243–2280 (2017)
74. W.B.R. Lickorish, A representation of orientable combinatorial 3-manifolds. Ann. Math. **76**, 531–540 (1962)
75. J. Luecke, Dehn surgery on knots in the 3-sphere, in *Proceedings of the International Congress of Mathematicians, Zürich 1994* (Birkhäuser, 1995), pp. 585–594
76. G. Lusztig, Total positivity in reductive groups, in *Lie Theory and Geometry*, Progr. Math., vol. 123 (Birkhäuser, Boston, 1994), pp. 531–568
77. R. Mazzeo, J. Swoboda, H. Weiss, F. Witt, Ends of the moduli space of Higgs bundles. Duke Math. J. **165**(12), 2227–2271 (2016)

78. V.B. Mehta, C.S. Seshadri, Moduli of vector bundles on curves with parabolic structures. Math. Ann. **248**(3), 205–239 (1980)
79. J.J. Millson, On the first Betti number of a constant negatively curved manifold. Ann. Math. **104**, 235–247 (1976)
80. G. Mondello, Topology of representation spaces of surface groups in PSL $(2, \mathbb{R})$ with assigned boundary monodromy and nonzero Euler number. Pure Appl. Math. Q. **12**(3), 399–462 (2016)
81. J.W. Morgan, On Thurston's uniformization theorem for three-dimensional manifolds, in *The Smith Conjecture*, ed. by J.W. Morgan, H. Bass (New York, 1979), Pure Appl. Math., vol. 112 (Academic Press, Orlando, 1984), pp. 37–125
82. J. Morgan, G. Tian, The geometrization conjecture, in *Clay Mathematics Monographs 5*. American Mathematical Society, Providence, RI (Clay Mathematics Institute, Cambridge, MA, 2014), x+291 pp.
83. L. Nicolaescu, On the Cappell-Lee-Miller gluing theorem. Pac. J. Math. **206**(1), 159–185 (2002)
84. G. Perelman, Ricci flow with surgery on three-manifolds. arXiv: math/0303109 (2003)
85. R. Potrie, A. Sambarino, Eigenvalues and entropy of a Hitchin representation. Invent. Math. **209**(3), 885–925 (2017)
86. M.B. Pozzetti, Higher rank Teichmüller theories. Astérisque **422**, 327–354 (2020)
87. M.S. Raghunathan, *Discrete Subgroups of Lie Groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 68 (Springer, New York-Heidelberg, 1972)
88. J.G. Ratcliffe, *Foundations of Hyperbolic Manifolds*, Graduate Texts in Mathematics, 2nd edn., vol. 149 (Springer, New York, 2006)
89. A.W. Reid, A non-Haken hyperbolic 3-manifold covered by a surface bundle. Pacific J. Math. **167**(1), 163–182 (1995)
90. D. Rolfsen, Rational surgery calculus. Extension of Kirby's theorem. Pacific J. Math. **110**, 377–386 (1984)
91. P. Safari, A gluing theorem for Seiberg-Witten moduli spaces. Ph. D. Thesis, Columbia University, 2000
92. A.H.W. Schmitt, Moduli for decorated tuples of sheaves and representation spaces for quivers. Proc. Indian Acad. Sci. Math. Sci. **115**(1), 15–49 (2005)
93. A.H.W. Schmitt, *Geometric Invariant Theory and Decorated Principal Bundles*. Zürich Lectures in Advanced Mathematics (European Mathematical Society, 2008)
94. H. Seifert, Komplexe mit Seitenzuordnung. Nachr. Akad. Wiss. Göttingen Math.Phys. Kl. II **6**, 49–80 (1975)
95. C.S. Seshadri, Moduli of vector bundles on curves with parabolic structures. Bull. AMS **83**(1), 124–126 (1977)
96. C.T. Simpson, Constructing variations of Hodge structures using Yang-Mills theory and applications to uniformization. J. Amer. Math. Soc. **1**, 867–918 (1988)
97. C.T. Simpson, Harmonic bundles on noncompact curves. J. Amer. Math. Soc. **3**(3), 713–770 (1990)
98. C.T. Simpson, Higgs bundles and local systems. Inst. Hautes Études Sci. Publ. Math. **75**, 5–95 (1992)
99. J. Swoboda, Moduli spaces of Higgs bundles on degenerating Riemann surfaces. Adv. Math. **322**, 637–681 (2017)
100. C.H. Taubes, Self-dual Yang-Mills connections over non self-dual 4-manifolds. J. Differential Geom. **17**, 139–170 (1982)
101. W.P. Thurston, *The Geometry and Topology of Three-Manifolds*. Princeton University Mathematics Department Lecture Notes (Princeton University Press, Princeton, 1979)
102. N.G. Vlamis, A. Yarmola, Basmajian's identity in higher Teichmüller-Thurston theory. J. Topol. **10**(3), 744–764 (2017)
103. F. Waldhausen, On irreducible 3-manifolds which are sufficiently large. Ann. Math. (Second Series) **87** (1), 56–88 (1968)
104. C.T.C. Wall, On the work of W. Thurston, in *Proceedings of the International Congress of Mathematicians, Warsaw 1983* (Warszawa PWN, 1984), pp. 11–14

105. A.H. Wallace, Modifications and cobounding manifolds. Can. J. Math. **12**, 503–528 (1960)
106. A. Weil, On discrete subgroups of Lie groups. Ann. Math. (2) **72**, 369–384 (1960)
107. A. Wienhard, An invitation to higher Teichmüller theory, in *Proceedings of the International Congress of Mathematicians, Rio de Janeiro 2018*, vol. II. Invited Lectures (World Scientific Publishing, Hackensack, 2018), pp. 1013–1039
108. M. Wolf, The Teichmüller theory of harmonic maps. J. Differential Geom. **29**(2), 449–479 (1989)
109. M. Wolf, Infinite energy harmonic maps and degeneration of hyperbolic surfaces in moduli space. J. Diff. Geom. **33**, 487–539 (1991)
110. S. Wolpert, The Fenchel-Nielsen deformation. Ann. Math. **115**, 501–528 (1982)
111. Q. Zhou, The moduli space of hyperbolic cone structures. J. Differential Geom. **51**(3), 517–550 (1999)

Chapter 7
Acute Geodesic Triangulations of Manifolds

Sang-hyun Kim

Abstract We give a brief survey on acute geodesic triangulations of certain manifolds such as higher dimensional manifolds, Riemannian surfaces and flat cone surfaces. In the special case of a round two-sphere we review the result of the author with Walsh that gives a complete combinatorial characterization of acute geodesic triangulations. We particularly focus on results that are related with hyperbolic geometry, including Thurston's geometric description for the Deligne–Mostow lattices and the Koebe–Andreev–Thurston theorem on circle packings. We will briefly sketch the proofs of the key results, and list relevant outstanding open problems.

Keywords Acute geodesic triangulation · Koebe–Andreev–Thurston theorem · Polytope · Hyperbolic space · Complex hyperbolic lattice

2010 Primary: 57M60; Secondary: 20F36, 30F45, 52B05, 65M50

7.1 Introduction

Let M^d be a topological manifold. To avoid technicalities, we always assume that M is equipped with a smooth structure and with a compatible PL structure; we allow isolated singularities when $d = 2$. By a *combinatorial triangulation* of M we will mean a simplicial complex that is PL–homeomorphic to M.

If M is equipped with a Riemannian metric, a *geodesic triangulation* means a homeomorphism from a combinatorial triangulation L to M such that the image of each simplex in L is a geodesic simplex. Here, a *geodesic simplex* is inductively defined as the geodesic join of a geodesic simplex of one lower dimension; we will only consider geodesic simplices in uniquely geodesic neighborhoods to avoid

S.-h. Kim (✉)
School of Mathematics, Korea Institute for Advanced Study (KIAS), Seoul, South Korea
https://www.cayley.kr

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_7

any ambiguity of this definition. We declare a point to be a geodesic simplex of dimension zero.

The main object of study in this survey is an *acute (geodesic) triangulation* of a Riemannian manifold (M^d, g), i.e. a geodesic triangulation such that every pair of $(d-1)$ faces in each geodesic simplex makes an angle less than $\pi/2$ at the intersection points. We will mostly focus on the three questions below:

Question 7.1.1 *Let* (M^d, g) *be a Riemannian manifold.*

(1) Does there exist an acute triangulation of M *at all?*
(2) Which combinatorial triangulations are realized as acute triangulations of M*?*
(3) Are there "qualitatively good" acute triangulations of M*?*

The phrase "qualitatively good" may have several meaningful interpretations. One motivation of studying acute triangulations comes from the *finite element method*, a technique in numerical analysis used to find approximate solutions to differential and integral equations. As one form of the method, a Euclidean domain is divided into small geodesic simplices and a given function is interpolated linearly in each subdomain. It is well-known that not just smaller sizes, but also more regular shapes of simplices are important in reducing interpolation errors, gradient interpolation errors, round-off errors, the worst-case running time and so forth [55]. An acute triangulation is suggested as one way of achieving such regularity of shapes, based on empirical evidences; see [63, 67] and references therein. Many other qualitative criteria of triangulations, such as the ratios of the circumscribed balls to the inscribed ones, are considered in the literature [8, 55].

As a more theoretical instance, let us consider a geodesic triangulation T of a compact Riemannian surface S. Given a function f on S, we can consider a piecewise linear interpolation F defined by the values of f on the 0-skeleton $T^{(0)}$. Colin de Verdière points out [16] that the Dirichlet integral $\int |dF|^2$ has a certain desirable form that is a discrete analogue of the classical *Schrödinger operator* if the geodesic triangulation T is acute.

In this survey, we start from a discussion on dimensions $d \geq 5$ where part (1) of Question 7.1.1 has the negative answer. Namely, using the classical Dehn–Sommerville relations we sketch Kalai's proof that manifolds of dimension $d \geq 5$ do not admit acute triangulations. For certain three and four manifolds, we will survey known negative and affirmative answers for the same question. Part (2) of the question on the classification of acute triangulations in these dimensions still seems far from reach.

We will then move to manifolds of dimension two, i.e. surfaces. We consider two classes of surfaces: general Riemannian surfaces and flat cone surfaces. These two types of surfaces are acutely triangulable, admitting the affirmative answer to part (1) of the above question. More strongly, an "almost equilateral" triangulation of a general Riemannian surface is described, in relation to part (3) of Question 7.1.1. We will then illustrate a remarkable discovery of Thurston that completely classifies equilateral triangulations of certain flat cone spheres. The proof of a result by Margulis–Mozes is included, which shows that the hyperbolic plane can be aperiodically tiled by a single acute triangle.

The last object to be discussed is a round sphere S^2. This is one of a few spaces where the first two questions are completely answered. We will include necessary background material including the Koebe–Andreev–Thurston theorem and CAT(κ) geometry, and then describe the complete classification by Walsh and the author [33] of acute triangulations of S^2.

7.2 In Dimension Three and Higher

It is a fundamental question as to whether or not a given space admits an acute triangulation. An affirmative answer can be given by prescribing such a triangulation, while the negative answer is often much trickier to exhibit. One of the main tools in dimension $d \geq 3$ is the theory of convex polytopes.

7.2.1 Polytopes and Dehn–Sommerville Equations

Let us first introduce basic concepts of the general polytope theory. Let P be a *d-polytope*, namely the convex hull of a finite set in $\mathbb{R}^d$. We let $\mathscr{F}_i(P)$ denote the set of its i-dimensional faces, and $f_i(P)$ the cardinality of $\mathscr{F}_i(P)$. We regard the empty set as a (-1)-dimensional face. The *face lattice* of P is a partial order structure on

$$\mathscr{F}(P) := \bigcup_{i=-1}^{d} \mathscr{F}_i(P)$$

defined by inclusion. The *face vector* of P is defined as

$$f(P) := (f_0(P), f_1(P), \ldots, f_{d-1}(P)).$$

Let $\langle x, y\rangle$ denote the usual inner product in $\mathbb{R}^d$. If P contains the origin in its interior, then its *(polar) dual* P^* is defined by

$$P^* := \{y \in \mathbb{R}^d \mid \langle x, y\rangle \leq 1 \text{ for all } x \in P\},$$

There exists an order–reversing bijection between the face lattices $\mathscr{F}(P)$ and $\mathscr{F}(P^*)$. In particular, we have $f_i(P) = f_{d-1-i}(P^*)$.

For each face $F \in \mathscr{F}(P)$ the interval $[F, P]$ in the lattice $\mathscr{F}(P)$ is order–isomorphic to the face lattice of some polytope $\mathrm{Link}_P(F)$, called the *(combinatorial) link* of F in P. In the special case when F is a vertex then $\mathrm{Link}_P(F)$ is combinatorially isomorphic to the intersection between P and a small sphere around F.

Setting $f_d(P) = 1$ and using the Euler characteristic formula for $P \approx B^d$, we obtain

$$\sum_{i=0}^{d}(-1)^i f_i(P) = \chi(P) = \chi(\text{point}) = 1.$$

The above equality is called the *Euler relation* for the polytope P.

A polytope is said to be *simplicial* if it is combinatorially isomorphic to a simplicial complex. While the Euler relation is *the only* affine relation satisfied by the face vectors of all (possibly non-simplicial) d-polytopes [27], a simplicial polytope satisfies extra affine relations called the *Dehn–Sommerville equations*:

$$\sum_{j=k}^{d-1}(-1)^j \binom{j+1}{k+1} f_j(M) = (-1)^{d-1} f_k(P).$$

Here, we let $k = -1, 0, \ldots, d-1$, and $f_{-1}(P) = 1$. These equations are obtained by repeatedly applying the Euler relation to all the links of the faces of the polytope P.

More generally, let M be a closed oriented triangulated n-manifold and let $f_i(M)$ denote the number of i-simplices in M. We let $f_{-1}(M)$ be half of the Euler characteristic of M. Then a similar idea as above yields the *Dehn–Sommerville equations for closed manifolds* [34], which again hold for each $k = -1, \ldots, n$:

$$\sum_{j=k}^{n}(-1)^j \binom{j+1}{k+1} f_j(M) = (-1)^n f_k(M).$$

The equation for $k = -1$ is simply the Euler characteristic formula. Setting $M = S^{d-1}$, we recover the Dehn–Sommerville equations for d-polytopes.

7.2.2 Spherical Complexes

In order to study dihedral angles of a spherical, Euclidean or hyperbolic d-polytope P it is natural to consider the *(metric) link* of a vertex v in P, which is defined as the space $\mathrm{Link}_P(v) = \mathrm{T}^1_v(P)$ of unit tangent vectors at v. The Riemannian angle at v naturally induces a spherical metric on $\mathrm{Link}_P(v)$, making this space isometric to a convex subset of a unit $(d-1)$-sphere.

We say a polytope is *simple* if the link of every vertex is combinatorially a simplicial complex. In the case when P is acute and simple, every link is an acute spherical simplicial complex. A *facet* of a d-dimensional polytope means the top (i.e. $(d-1)$-) dimensional face of P. We have the following simple observation.

Lemma 7.2.1 *If $d \geq 2$ then every facet of an acute spherical d-simplex is acute.*

Remark 7.2.2 When $d = 2$ the lemma shows that the side lengths of an acute spherical triangle are acute, i.e. less than $\pi/2$. The converse of the lemma does not hold, even when $d = 2$. For instance, consider a spherical triangle of dihedral angles $(\pi/3, \pi/3, \pi/2)$. This triangle is obtained by dividing a sphere into four isometric equilateral triangles (tetrahedral subdivision) and then further dividing each triangle into six isometric triangles sharing the center of the gravity as a common vertex. Although each side length is acute, the triangle itself is not acute. By a small perturbation one can even find an obtuse spherical triangle with acute side lengths.

Proof of Lemma 7.2.1 Let P be a d-dimensional acute spherical simplex. We first consider the case $d = 2$. If P has side lengths a, b, c and their opposite angles A, B, C then it satisfies the spherical law of cosines for angles:

$$\cos c = \frac{\cos C + \cos A \cos B}{\sin A \sin B}.$$

This implies that the cosine of each side length is positive, implying that each facet of P is acute.

Consider the case that $P = [v_0, \ldots, v_d]$ is $d \geq 3$ dimensional. We may inductively assume that the conclusion holds for an acute spherical $(d-1)$-simplex. We claim that an arbitrary facet F is acute. We will fix $F := [v_0, v_1, v_2, \ldots, v_{d-1}]$ and estimate the dihedral angle θ between these two faces of F:

$$E_0 := [v_0, \widehat{v_1}, v_2, v_3, \ldots, v_{d-1}],$$

$$E_1 := [v_0, v_1, \widehat{v_2}, v_3, \ldots, v_{d-1}].$$

Let $\bar{v}_i$ be the image of the geodesic ray $v_0 v_i$ in $\mathrm{Link}_P(v_0) = \mathrm{T}^1_{v_0}(P)$. Then the angle θ is equal to the dihedral angle between

$$\bar{E}_0 := [\bar{v}_2, \bar{v}_3, \ldots, \bar{v}_{d-1}],$$

$$\bar{E}_1 := [\bar{v}_1, \bar{v}_3, \ldots, \bar{v}_{d-1}].$$

Since $\mathrm{Link}_P(v_0)$ is an acute spherical $(d-1)$-simplex we inductively see that the $(d-2)$-simplex

$$[\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_{d-1}]$$

is also acute. It follows that θ is acute, as required. □

7.2.3 Dimension Four and Five

Regarding manifolds of dimension $d = 4$, one has the following result; this is implicit in [31] and [35] for $M = S^4$ and $M = (S^1)^4$, respectively.

Theorem 7.2.3 ([31, 35]) *If a closed Riemannian* 4*-manifold admits an acute triangulation then its Euler characteristic is larger than four.*

Proof Assume that a closed Riemannian 4-manifold M is given with a geodesic triangulation, and that $\chi(M) \leq 4$. We will prove that some 2-face is shared by four or fewer 4-faces. Plugging the face vector

$$f := (f_0, f_1, f_2, f_3, f_4)$$

of M into the Dehn–Sommerville equations with $k = -1, 0$ and 3, we obtain the following.

$$\begin{pmatrix} 1 & -1 & 1 & -1 & 1 \\ 0 & -2 & 3 & -4 & 5 \\ 0 & 0 & 0 & -4 & 10 \end{pmatrix} f = \begin{pmatrix} \chi(M) \\ 0 \\ 0 \end{pmatrix}$$

Eliminating f_1 and f_3, we have

$$10 f_4 - 5 f_2 + 10 f_0 = 10\chi(M).$$

For each $0 \leq i < j \leq 4$ we define a *flag number* as

$$f_{i,j} := \#\{(u_i, u_j) \mid u_t \text{ is a } t\text{–face for } t = i \text{ and } t = j \text{ such that } u_i \subseteq u_j\}.$$

It is easy to see that $f_{i,j} = \binom{j+1}{i+1} f_j$. Since each 4-simplex contains five vertices we have

$$f_{2,4} = 10 f_4 = 5 f_2 + 10(\chi(M) - f_0) \leq 5 f_2 + 10(\chi(M) - 5) < 5 f_2.$$

This implies that the average number of 4-simplices that contain a given 2-simplex in the triangulation of M is less than five. In particular, some 2-simplex u_2 must belong to four or fewer 4-simplices. Since the sum of the dihedral angles at u_2 of those 4-simplices is 2π we conclude that the given triangulation is not acute. □

Remark 7.2.4 Kalai [31] actually proved the following result: if $d \geq 5$ then every d-dimensional polytope must contain a $(d-3)$-face shared by at most four $(d-1)$-faces. For a simplicial polytope P this follows from applying the argument of Theorem 7.2.3 to the boundary ($\approx S^4$) of the link of each $(d-5)$-face in P. The case of general polytopes would require a deep result of rigidity theory of frameworks [32].

Let us note another consequence of the above theorem.

Corollary 7.2.5 (cf. [35, Theorem C]) *If $d \geq 5$, then a Riemannian manifold of dimension d does not admit an acute triangulation.*

In particular, the space $\mathbb{R}^d$ does not admit an acute triangulation for $d \geq 5$; see also [37, 38]. One may compare this result with the non-existence of hyperbolic Coxeter d-polytopes for sufficiently large d, which relies on the fact that higher dimensional simple polytopes have many triangular and quadrangular faces; see [65].

Proof of Corollary 7.2.5 The key idea is essentially the same as in the proof of Theorem 7.2.3. Let L be such an acute triangulation, which is possibly infinite. Pick an arbitrary vertex v of L. Topologically, the space $Q := \mathrm{Link}_L(v)$ can be identified with the intersection of L with a very small $(d-1)$-sphere centered at v; recall our standing assumption from the introduction that the link of a vertex in a combinatorial triangulation is a sphere. It follows that Q inherits a simplicial complex structure from that of L. As in Theorem 7.2.3 and Remark 7.2.4, there exists some $(d-3)$-face of Q shared by at most four $(d-1)$-faces of Q. In particular, the given geodesic triangulation of Q is non-acute. This implies that a dihedral angle of some $(d-2)$-face in L, measured at v, is non-acute, either. □

Another consequence of Theorem 7.2.3 is that no acute triangulation of $\mathbb{R}^4$ is *periodic*, i.e. invariant under cocompact isometric actions. For example, a 4-dimensional cube or parallelepiped will never admit an acute triangulation since it tiles the space by translations. In [35], the authors proved a stronger fact that a (hypothetical) acute triangulation of $\mathbb{R}^4$ must contain a sequence of simplices in which the largest angles converge to $\pi/2$. In particular, one cannot hope to tile $\mathbb{R}^4$ with only finitely many isometry types of geodesic 4-simplices, periodically or non-periodically. However, the following question is still open.

Problem 7.1 Does $\mathbb{R}^4$ admit an acute (necessarily non-periodic) triangulation?

7.2.4 $\mathbb{R}^3$, S^3 and More

Aristotle [3] stated the following in his treatise *On the Heavens*:

> It is agreed that there are only three plane figures which can fill a space, the triangle, the square, and the hexagon, and only two solids, the pyramid and the cube.

In a certain interpretation (e.g. [57]), Aristotle's remark is understood to mean that $\mathbb{R}^3$ is filled with regular tetrahedra; this remark is then obviously a mathematical error since a regular tetrahedron has the dihedral angle $\arccos 1/3$, which is not a rational multiple of π. See [54] for more historical remarks on this perspective especially in relation to Plato's theory of atoms.

It was not until the early 21st century that an acute triangulation of $\mathbb{R}^3$ was first constructed [22]. This result was further strengthened later as follows.

Theorem 7.2.6 ([35, 64]) *A 3–dimensional cube (and hence, a 3–torus) admits an acute triangulation.*

The authors of [64] used a computational search to find such a triangulation with 1370 tetrahedra. Let us sketch the proof in [35], which gives a more conceptual argument using 2715 tetrahedra. The authors of the latter paper considered the *600-cell*, which is a four-dimensional (Euclidean) polytope consisting of 600 facets. The 600-cell is polar dual to the 120-cell, a four-dimensional polytope consisting of 120 dodecahedra on the boundary. Since each face-adjacent pair of dodecahedra in the 120-cell share five vertices forming a pentagon, each edge in the 600-cell is shared by five tetrahedra.

The 600-cell is homeomorphic to a four-dimensional ball, and the boundary is a simplicial complex homeomorphic to S^3. In this simplicial complex X_{600}, consider the union $\bar{Q}$ of a simplex Q and all the simplices intersecting Q. Note that each vertex v in X_{600} belongs to twenty tetrahedra so that $\mathrm{Link}_v(X_{600})$ is an icosahedron. By inclusion–exclusion, the number of simplices in $\bar{Q}$ can be counted as

$$20 \cdot \#Q^{(0)} - 5 \cdot \#Q^{(1)} + 2 \cdot \#Q^{(2)} - 1 \cdot \#Q^{(3)} = 80 - 30 + 8 - 1 = 57.$$

The union of the remaining $543 = 600 - 57$ tetrahedra in X_{600} is aptly denoted as

$$X_{543} := X_{600} \setminus \mathrm{Int}\,\bar{Q}.$$

It is obvious that X_{543} is a topological ball since it is the complement of a ball from S^3. The authors of [35] prove the following, which implies Theorem 7.2.6 by dividing a cube into one regular tetrahedron and four copies of a *standard* tetrahedron, i.e. a tetrahedron containing three unit normal vectors on its edge set. The proof uses a stereographic projection of $X_{600}^{(0)} \subseteq S^3 \subseteq \mathbb{R}^4$ to $\mathbb{R}^3$ along with a numerical search by a computer.

Lemma 7.2.7 *A regular tetrahedron and a standard tetrahedron admit acute triangulations which is combinatorially equivalent to X_{543}.*

For S^3, we can projectivize the boundary of the Euclidean 600-cell to the set

$$x_1^2 + x_2^2 + x_4^2 + x_4^2 = R$$

with a sufficiently R, and obtain a metric description of X_{600} as a geodesic triangulation of a three-dimensional sphere. As we noted above the projection of each edge belongs to five tetrahedra in X_{600}. The dihedral angle at an edge of a spherical regular tetrahedron in X_{600} is hence $2\pi/5$, which is acute. From this we have the following.

Proposition 7.2.8 *A three-dimensional sphere admits an acute triangulation.*

On the other hand, the classification problem for acute triangulations of S^3 still does not seem to be within reach. In fact, it is not even known exactly which combinatorial triangulation of S^3 can be realized as the boundary of a 4-polytope; see [48] and the references therein. The following basic question is also open.

Problem 7.2 What is the smallest number of tetrahedra needed to acutely triangulate a three-sphere?

The number 600 was a modestly conjectured answer to the above problem [9]. We will note later from Lemma 7.2.1 and Corollary 7.4.9 that each vertex in an acute triangulation of S^3 must belong to at least twenty tetrahedra.

Another intriguing open question is the following; the 600-cell is an example of such a triangulation for S^3, tiled by regular tetrahedra of dihedral angles $2\pi/5$.

Problem 7.3 Does $\mathbb{R}^3$ admit a triangulation by copies of a single (acute) tetrahedron?

7.3 Dimension Two: General Riemannian and Flat Cone Metrics

In this section, we consider acute triangulations of surfaces with general Riemannian metrics, and also with singular flat metrics.

7.3.1 Riemannian Surfaces

Let S be a surface with a Riemannian metric g. If L is a combinatorial triangulation of S, it is natural to wonder if L can be realized as a (not necessarily acute) geodesic triangulation.

For example, let us consider the case that (S, g) is the round sphere, namely the space S^2 with the usual Fubini–Study metric. We will see in the next section (the Koebe–Andreev–Thurston Theorem) that every combinatorial triangulation L of S^2 is realized as the *nerve of a coin graph*; this means that there exists a collection of interior-disjoint non-degenerate disks of various radii on the sphere such that Γ is isomorphic to the graph obtained by placing a vertex at the center of each disk and by joining two vertices whose corresponding disks are tangent to each other. By joining the centers of tangent disks by spherical geodesic arcs, we obtain a geodesic triangulation of S^2 that is combinatorially isomorphic to L.

In the case of surfaces with nonpositive curvatures, Colin de Verdière obtained the following.

Theorem 7.3.1 ([15]) *If X is a closed orientable Riemannian surface with a nonpositive Gaussian curvature, then every combinatorial triangulation of X can be realized by a geodesic triangulation.*

When X is a hyperbolic manifold, the resulting geodesic triangulation above can be regarded as the two-dimensional case of the *straightening map*

$$C_*(X) \to C_*(X)$$

defined by Thurston [59]. This is a chain map chain homotopic to the identity, which sends a singular simplex to a geodesic simplex having the same vertex images.

In the case of a general Riemannian surface as in Theorem 7.3.1, the main idea is to consider a homeomorphism from a given simplicial complex L to S, and suitably define an energy functional determined by the images of edges. Using variational principles, the minimum of the energy functional can be shown to be attained by a geodesic triangulation realizing L.

Strengthening the existence of acute triangulations, Colin de Verdière and Marin proved that Riemann surfaces admit "almost equilateral" triangulations as follows.

Theorem 7.3.2 ([14]) *If X is a closed orientable Riemannian surface with $\sigma \in \{-1, 0, 1\}$ denoting the sign of the Euler characteristic of X, and if K is the smallest compact interval containing*

$$\left\{\frac{4\pi}{12-2\sigma}, \frac{(4-\sigma)\pi}{12-2\sigma}\right\},$$

then for every open interval J containing K there exists a geodesic triangulation of X such that the angles of all the triangles belong to J; furthermore, one can require that such a geodesic triangulation has arbitrarily small diameters of geodesic triangles.

Let us briefly sketch the proof of the theorem. The starting point is the observation that the statement is invariant under a conformal change of the metric $(X, g) \mapsto (X, e^F g)$, where $F\colon X \to \mathbb{R}$ is an arbitrary smooth map. From this the torus case follows immediately, since one can simply assume that X is flat and approximate X by the union of flat equilateral triangles.

In the case when the genus h of X is at least two, one has a sextic holomorphic form $\Phi = \phi(z)dz^6$ on the Riemann surface X with simple zeroes. The cube root of the modulus of Φ then defines a singular flat metric $\bar{g}$ on X conformal to g, and the singularities of this new metric can be shown to have cone angles $7\pi/3$. Colin de Verdière and Marin then establish that for a dense choice of the Riemannian metric g, the resulting singular flat metric $\bar{g}$ admits a triangulation by flat equilateral triangles which arbitrarily approximates a desired triangulation in the metric g. The case $h = 0$ is similar.

It is much easier to see that the choice of the interval K is optimal. Indeed, suppose X admits a geodesic triangulation L and let v, e and f denote the numbers of vertices, edges and faces. We let $\mathrm{val}(x)$ denote the valence of a vertex $x \in L^{(0)}$. Combining the relation

$$2e = 3f = \sum_{x \in L^{(0)}} \mathrm{val}(x)$$

with the Euler characteristic formula, we see that

$$6\chi(X) = 6(v - e + f) = \sum_x (6 - \mathrm{val}(x)).$$

It follows that if $\chi(X) = 2 = 2\sigma$ then there exists vertices of valence at most five; in this case, there exists a triangle T in L such that some angle θ_1 of T satisfies $\theta_1 \geq 2\pi/5$. Moreover, if all the triangles are sufficiently small, then the sum of the angles of T is $\pi + \delta$ for some small $\delta > 0$. So, another angle θ_2 of T satisfies

$$\theta_2 \leq \frac{1}{2}\left(\pi + \delta - \frac{2\pi}{5}\right) = \frac{3\pi}{10} + \frac{\delta}{2}.$$

This means that one cannot replace the interval K by another compact interval not containing K. If $\sigma = -1$, then some vertex of L has valence at least seven and one sees the optimality in a similar way. The case $\sigma = 0$ is obvious.

One tantalizing question is the parametrization of all combinatorial triangulations that can appear in the conclusion of Theorem 7.3.2. Suppose that X is a Riemannian surface, and that the interval J is chosen to be sufficiently close to the compact interval K given in the theorem. Let L be a geodesic triangulation obtained in the conclusion. If $X \approx S^2$, then the valences of the vertices in L are all five or six; this is because we have

$$K = \left[\frac{3\pi}{10}, \frac{2\pi}{5}\right] \subsetneq \left(\frac{2\pi}{7}, \frac{2\pi}{4}\right)$$

and this inclusion implies that there are no vertices of valences seven (or more) or of valences four (or less). By the Euler characteristic argument given above, one has exactly twelve vertices of valence five.

Similarly, when $X \approx T^2$ then all the valences must be six. When the genus h is at least two, we note from the inclusion

$$K = \left[\frac{2\pi}{7}, \frac{5\pi}{14}\right] \subsetneq \left(\frac{2\pi}{8}, \frac{2\pi}{5}\right)$$

that all the valences in the triangulation are either six or seven; more precisely, we have exactly $12(h - 1)$ vertices of valence seven and all the rest have valence six. This naturally leads us to the following problem.

Problem 7.4 Let X be a closed orientable surface and let $\sigma \in \{-1, 0, 1\}$ be the sign of the Euler characteristic of X. Can we parametrize all the combinatorial triangulations of X such that the valences of all the vertices are either 6 or $6 - \sigma$?

See Sect. 7.3.3 for Thurston's highly influential contribution to this problem when X is homeomorphic to the two-sphere [60].

7.3.2 *Euclidean and Flat Cone Surfaces*

One may wonder how many triangles are needed to acutely triangulate an obtuse triangle; this question can be traced back to a recreational mathematical article of Gardner [23]. It was proved soon after the question was asked that the answer is seven by a clever choice of a triangulation [24]. For a general n-gon the minimal number is estimated as $O(n)$ [39].

There are also numerous results on *optimal* acute triangulations, in which dihedral angles are bounded away from 0 or $\pi/2$ by a definite amount, and in which the number of triangles is controlled. Bishop [5] proved that every Euclidean polygon admits a triangulation such that each triangle not containing a boundary vertex has angles in $[\pi/6, 5\pi/12]$.

More generally, one can consider *flat cone surfaces* defined as follows. A *flat cone surface* is a topological surface X equipped with a metric g and with a finite set of *singularities* on X such that X is locally isometric to a Euclidean disk near points away from the singularities and such that X is locally isometric to the Euclidean cone

$$\{z \in \mathbb{C} \mid |z| < \epsilon, 0 \leq \arg z \leq \theta\}/z \sim e^{i\theta} z$$

at each singularity for some $\theta > 0$ and $\epsilon > 0$. This metric gives a Riemann surface structure on X. The angle $\theta > 0$ is called the *cone angle* of the singularity; it is often more convenient to specify the *apex curvature* [60] (or, *cone deficit*), which is $2\pi - \theta < 2\pi$. A flat torus or the boundary of a platonic solid are simple examples of such surfaces.

A result due to Troyanov [61] states that for all distinct points $p_1, \ldots, p_s$ on a closed Riemann surface (X, g) and for all values $\kappa_1, \ldots, \kappa_s$ in $(-\infty, 2\pi)$ satisfying $\sum_i \kappa_i = 2\pi\chi(X)$ there exists a conformal flat cone metric $\bar{g}$ on X such that each p_i is a singularity of apex curvature κ_i; furthermore, such a metric $\bar{g}$ is unique up to homothety. See also Thurston's explanation in [60, Section 8]. The formal sum

$$\sum_i \frac{-\kappa_i}{2\pi} p_i$$

is called the *divisor* of $\bar{g}$.

Y. D. Burago and V. A. Zalgaller established the following fundamental results regarding acute triangulations of flat cone surfaces.

Theorem 7.3.3 ([11] for (1), [12] for (2))

(1) A flat cone surface admits an acute triangulation.
(2) A flat cone surface admits an isometric piecewise linear (continuous) embedding into $\mathbb{R}^3$.

In fact, the second part of the above theorem was crucially used in the proof of the first, which is a discrete analogue of the Nash–Kuiper embedding theorem. See [51] for an elementary proof of the first part. An extensive survey on works relevant to Nash–Kuiper theorem can be found in [26].

The boundaries of platonic solids are examples of flat cone spheres. As an example, a regular icosahedron can be given with a metric such that each facet is an equilateral triangle with unit side lengths; this metric is flat everywhere but the vertices, at which the apex curvature is $\pi/3$. Similarly, a regular dodecahedraon has a flat metric with 20 cone points, at which apex curvatures are $\pi/5$. The minimal number of triangles to acutely triangulate a regular icosahedral and a regular dodecahedral surface is twelve in both cases. The minimal number of all the other platonic solids is also found; see the survey [67] and references therein. The proofs use clever choices of triangulations (for sufficiency) and Euler number arguments (for necessity). Generalizing these results, we have the following intriguing question.

Problem 7.5 Let X be a compact flat cone surface with a geodesic triangulation, in which all the triangles are isometric to the unit Euclidean equilateral triangle. Does there exist an algorithm to determine the minimum number of triangles that acutely triangulate X?

Note that such a mimimum number is finite by the aforementioned result of Burago and Zalgaller.

7.3.3 Parametrizing Equilateral Triangulations

Let us now concisely overview Thurston's solution to Problem 7.4 for the two-sphere. His solution is in fact a by-product of a much far-reaching theory of moduli spaces of flat cone metrics on surfaces and their realizations as complex hyperbolic orbifolds [60]. His theory gave a geometric interpretation of all of the 94 complex hyperbolic lattices in dimension 3 through 9 discovered by Mostow [44] and Deligne–Mostow [18, 19]. For more details on this interpretation, see [45] and references therein.

A *triangle complex structure* on a topological surface X is a 2-dimensional CW-complex homeomorphic to X such that each 2-cell is the image of a triangle. For instance, if ABC denotes the image of a closed triangular 2-cell in X, then we allow that two vertices B and C coincide. Moreover, it is allowed that two triangular 2-cells share two edges in this subsection, which is forbidden for a simplicial complex. Abusing the terminology (as was done in [60]), by a triangulation we will mean a triangle complex structure.

Let $n, s, k_1, \ldots, k_s$ be positive integers such that $1 \leq k_i \leq 5$. The main combinatorial object of interest for us is the space

$$P(n; k_1, \ldots, k_s)$$

of triangle complex structures (simply, triangulations) of the two-sphere having $2n$ faces with distinguished vertices $v_1, \ldots, v_s$ such that the valence of v_i is $6 - k_i$ and such that all the other vertices have valence six, up to isomorphism of triangulations. Each triangulation in $P(n; k_1, \ldots, k_s)$ is said to have *non-negative* curvature, as the apex curvature is either positive or zero at each vertex of the triangulation after all triangles are metrized as Euclidean equilateral triangles. Note from the Euler characteristic formula that

$$\begin{aligned} 2 = \chi(X^2) = v - e + f &= \sum_{w\text{: vertex}} \left(1 - \frac{1}{2}\operatorname{val}(w) + \frac{1}{3}\operatorname{val}(w)\right) \\ &= \sum_{w\text{: vertex}} \frac{6 - \operatorname{val}(w)}{6} = \sum_{i=1}^{s} \frac{k_i}{6}. \end{aligned}$$

It will be convenient for us to introduce the notation

$$a^{\otimes n} := (a, \ldots, a),$$

where a is repeated n times. The parametrization space considered in Problem 7.4 for $\sigma = 1$ is realized as a subset of the space

$$P(n; 1^{\otimes 12})$$

for various choices of n.

For positive real numbers $\alpha_1, \ldots, \alpha_s$ we let $C(\alpha_1, \ldots, \alpha_s)$ denote the moduli space of flat cone metrics with cone angles

$$2\pi - \alpha_1, \ldots, 2\pi - \alpha_s,$$

modulo orientation-preserving similarities without distinguishing the singular points. We also let $\hat{C}(\alpha_1, \ldots, \alpha_s)$ denote the finite cover of $C(\alpha_1, \ldots, \alpha_s)$ where the s singular points are distinguished by labels. By the result of Troyanov mentioned above, for each choice of distinct s points on S^2 with the usual Fubini-Study metric there exists a conformal flat cone structure in $\hat{C}(\alpha_1, \ldots, \alpha_s)$ up to scaling if and only if $\sum \alpha_i = 4\pi$.

Each triangulation T in $P(n; k_1, \ldots, k_s)$ determines a flat cone metric in $C(\frac{1}{3}\pi k_1, \ldots, \frac{1}{3}\pi k_s)$ by metrizing each of the $2n$ triangles in T as a Euclidean triangle with unit side lengths. The resulting Riemann surface has divisor

$$\sum_i \frac{-k_i}{6} v_i$$

up to permutation of the singular vertices. In general, flat cone metrics on a surface S obtained by gluing Euclidean equilateral triangles are called *equilateral triangulations* of S. It is known that every noncompact Riemann surface admits an equilateral triangulation [6]. A necessary and sufficient condition for a compact Riemann surface to admit an equilateral triangulation is given in [66]. A point in $C(\cdot)$ coming from an equilateral triangulation will be called a *triangulation point* in the (unlabeled) moduli space $C(\cdot)$.

Let $d \geq 1$ be an integer. The space $\mathbb{C}^{d+1}$ is equipped with the standard (also called "first" in the literature) Hermitian form

$$\langle (z_0, \ldots, z_d), (w_0, \ldots, w_d) \rangle := \sum_{i<d} z_i \bar{w}_i - z_d \bar{w}_d.$$

The image of complex vectors $\mathbf{z} \in \mathbb{C}^{d+1}$ satisfying $\langle \mathbf{z}, \mathbf{z} \rangle < 0$ under the projection map $\mathbb{C}^{d+1} \setminus \{0\} \to \mathbb{C}P^d$ is called the *complex hyperbolic d-space* and denoted as $\mathbb{H}^d_{\mathbb{C}}$. This space comes with the natural metric $d(\cdot, \cdot)$ given by

$$\cosh^2 \left(\frac{d(\mathbf{z}, \mathbf{w})}{2} \right) := \frac{\langle \mathbf{z}, \mathbf{w} \rangle \langle \mathbf{w}, \mathbf{z} \rangle}{\langle \mathbf{z}, \mathbf{z} \rangle \langle \mathbf{w}, \mathbf{w} \rangle}.$$

The space $\mathbb{H}^d_{\mathbb{C}}$ is often modeled on the open complex unit ball $B^d \subseteq \mathbb{C}^d$ under the projection

$$\mathrm{Proj} \colon (z_0, \ldots, z_d) \mapsto (z_0/z_d, \ldots, z_{d-1}/z_d).$$

The induced metric is called the *Bergman metric* on B^d. More generally, an arbitrary Hermitian form of signature $(d, 1)$ on $\mathbb{C}^{d+1}$ is equivalent to the first standard Hermitian form, defining a metric space on a suitable subset of $\mathbb{C}P^d$ that is isometric to $\mathbb{H}^d_{\mathbb{C}}$.

Recall that a *lattice* Γ isometrically acting on a symmetric space is a discrete subgroup of the corresponding Lie group with finite co-volume. As a particular case, a lattice in a complex space $\mathbb{C}^n$ will mean an additive subgroup of finite co-volume.

Thurston's result on Problem 7.4 can be summarized as follows.

Theorem 7.3.4 ([60]) *The topological space $\mathscr{M} := C((\pi/3)^{\otimes 12})$ admits a Riemannian metric, whose completion $\bar{\mathscr{M}}$ is isometric to $\mathbb{H}^9_{\mathbb{C}}/\Gamma$ for some lattice $\Gamma \leq \mathrm{Isom}(\mathbb{H}^9_{\mathbb{C}})$.*

See a note of Schwartz [53, Remark 1.2] for an alternative account. For the proof, we consider the finite cover $\hat{\mathscr{M}} := \hat{C}((\pi/3)^{\otimes 12})$ and its 9-dimensional complex-projective parametrization as follows. Each metric g in this latter space corresponds to a divisor

$$\sum_i \frac{-1}{6} v_i$$

for some $v_1, \ldots, v_{12}$. There exists a geodesic triangulation T of (S^2, g) with vertices on the singularities [60, Proposition 3.1]. We let $\tilde{T}'$ be the the universal cover of $T' := T \setminus \{v_1, \ldots, v_{12}\}$. One can show that an assignment of complex numbers to 10 edges on this triangulation (which form a spanning tree of 11 vertices) determines a map Z from the set of edges of $\tilde{T}'$ to $\mathbb{C}$, representing the image of an edge (as a complex vector) in $\tilde{T}'$ under a fixed developing map

$$\mathrm{D} \colon \tilde{T}' \to \mathbb{C}.$$

Denoting the orthogonal part of the Euclidean holonomy $\pi_1(T') \to \mathrm{Isom}(\mathbb{R}^2)$ as H_0, we can see a cocycle relation

$$Z(\gamma.e) = H_0(\gamma)Z(e)$$

for each $\gamma \in \pi_1(T')$ and $e \in (\tilde{T}')^{(1)}$. This cocycle Z completely encodes the necessary data for the developing map up to a scalar multiplication. After projectivising by multiplicative complex numbers, the space $\hat{\mathscr{M}}$ admits local coordinate charts in $\mathbb{C}P^9$, and becomes a 9-dimensional complex-projective manifold.

By elementary complex analysis, the negative of the Euclidean area of the flat cone metric with cocycle Z is easily seen to be

$$\mathrm{Area}^-(Z) := -\frac{i}{4} \sum_{\Delta \in T^{(2)}} (z\bar{w} - \bar{z}w).$$

Here, z and w are the images under Z of the two consistently oriented edges of a triangle Δ in T (or, more precisely their lifts in $\tilde{T}'$). See Fig. 7.1.

By an inductive argument on the number of singular points, one can prove that the map Area^- defines a Hermitian form H of signature (9, 1) in the parameter space

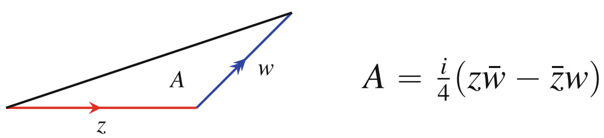


Fig. 7.1 The area of a triangle in the complex plane

$\mathbb{C}^{10}$, which we now denote as $\mathbb{C}^{9,1} := (\mathbb{C}^{10}, H)$. After restricting to the vectors V satisfying $H(V, V) < 0$ and projectivising

$$\mathrm{Proj}\colon V_-^H := \{V \in \mathbb{C}^{9,1} \mid H(V, V) < 0\} \to \mathbb{C}P^9,$$

we obtain complex hyperbolic charts on $\hat{\mathscr{M}}$. Using these charts, Thurston further proves that the completion $\bar{\mathscr{M}}$ of the unlabeled moduli space $\mathscr{M}$ is a complex hyperbolic orbifold, that is, the quotient of $\mathbb{H}^9_{\mathbb{C}}$ by a certain lattice Γ. Furthermore, such an identification of the completed moduli space $\bar{\mathscr{M}}$ of flat cone metrics with a complex hyperbolic orbifold can detect triangulation points.

Theorem 7.3.5 ([60]) *There exists a lattice $L \subseteq \mathbb{C}^{9,1}$ such that H and L are both Γ-invariant, and such that the above parametrization produces a one-to-one correspondence from each vector*

$$V \in (L \cap V_-^H)/\Gamma$$

to a non-negatively curved triangulation T_V of the sphere having total area equal to $-H(V, V)$. Furthermore, the lattice points V which project into $\mathscr{M} \subseteq \bar{\mathscr{M}} = \mathbb{H}^9_{\mathbb{C}}/\Gamma$ correspond to the triangulations T_V in $P(n; 1^{\otimes 12})$ for $n := -H(V, V)/(\sqrt{3}/4)$.

Here, the lattice L can be concretely written as the Eisenstein lattice

$$\mathbb{Z}\left[e^{2\pi i/3}\right]^{9,1} \subseteq \mathbb{C}^{9,1}.$$

One of the key ideas in Thurston's proof is slit-and-patch. Namely, given a flat cone metric

$$g \in \mathscr{M} = C(n; 1^{\otimes 12}),$$

one can find six disjoint Euclidean geodesic arcs joining six pairs of singular points. Slitting along one of these arcs, and patching the two–thirds of a Euclidean equilateral triangle with its side length equal to the length of the slit, one can combine the two singularities at the endpoints of the slit; see Fig. 7.2. Doing this for all six slits, we obtain a metric

$$g_1 \in C(n + m; 2^{\otimes 6}),$$

for some $m \geq 1$. The latter space includes for instance the metric of a regular octahedron. Inductively, we have a complete description of triangulation points with any less number of singularities and can obtain the conclusion of the theorem.

Using the complex hyperbolic volume of the above moduli spaces, Thurston also deduces that the size of the finite set $P(n; 1^{\otimes 12})$ is bounded by $O(n^{10})$; see [21, 36, 43] for precise estimates of these volumes and their asymptotics. This

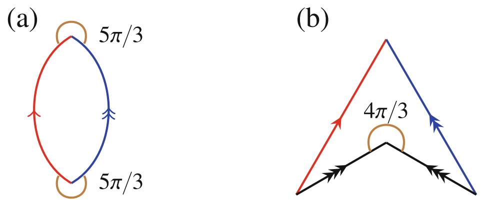


Fig. 7.2 Two cone points of cone angles $5\pi/3$ at a slit are replaced by a single cone point of cone angle $4\pi/3$ in a patch. (**a**) A slit. (**b**) A patch

is a very interesting combinatorial fact in its own, considering that the number of planar triangular tilings grows exponentially [62]. Thurston further notes that $P(n; 1^{\otimes 12})$ is nonempty for all $11 \neq n \geq 10$; one may compare this observation with Corollary 7.4.9.

Problem 7.6 Develop a theory parallel to the above theorems in the case of a higher genus surface. For instance, what is the growth of the number of almost equilateral (that is, having valences six or seven) combinatorial triangulations with n triangles for a closed surface of genus $g \geq 2$?

7.3.4 Aperiodic Tilings

Another interesting question related to the theory of triangulations is to classify *aperiodic tiling sets*, namely sets of topological balls that admit tilings of certain ambient spaces with the additional condition that no such tilings admit invariant cocompact isometric actions [42]. For instance, the famous Penrose tiling of the Euclidean plane provides an aperiodic tiling with two quadrangles. In a different perspective, Thurston solved the "tileability" problem of planar regions by various tiles by converting the problem to a question on finite presentations of groups [58].

It is easy to see that one triangle or one quadrangle can never aperiodically tile the Euclidean plane, since each of them does admit periodic tilings. More strongly, it was recently announced that no single convex Euclidean polygon admits an aperiodic tiling [47]. The analogous question in dimension three is still open; the answer for the dimension four or higher is negative by the results in Sect. 7.2.

Problem 7.7 Does there exist an aperiodic tiling of $\mathbb{R}^3$ or $\mathbb{H}^3$ by an acute simplex?

The hyperbolic three-space $\mathbb{H}^3$ admits a (periodic) tiling by the regular ideal simplex with dihedral angles $\pi/3$. Although unlikely, it is still not determined whether or not $\mathbb{R}^3$ can be tiled, periodically or not, by a single acute simplex (Problem 7.3). On the other hand, Margulis and Mozes discovered an affirmative answer for $\mathbb{H}^2$.

Theorem 7.3.6 ([42]) *The hyperbolic plane admits an aperiodic triangulation by a single acute geodesic triangle.*

The aperiodicity is easy to be guaranteed. If a triangle $T \subseteq \mathbb{H}^2$ tiles the plane periodically, then there exists a cocompact isometry group $\Gamma \leq \mathrm{PSL}(2, \mathbb{R})$ of the hyperbolic plane leaving the tessellation invariant. By considering a finite index torsion-free subgroup of Γ, we obtain a closed hyperbolic surface S tessellated by T. In particular, the area of T divides the value

$$\mathrm{Area}(S) = -2\pi\chi(S).$$

It follows that if the angle sum of T is an irrational multiple of π then every triangulation of $\mathbb{H}^2$ by copies of T is (if exists at all) aperiodic.

So, it suffices for us to find an acute geodesic triangle of $\mathbb{H}^2$, whose angle sum is an irrational multiple of π and which tessellates $\mathbb{H}^2$. Let us consider a hyperbolic isosceles triangle T with angles (A, B, B). We will require the following conditions, which are slight variations of those in [42].

- A and $2B$ are less than $\pi/2$;
- $A + 2B = r\pi$ for some irrational number $r \in (0, 1)$;
- $4A + 6B = 2\pi$.

Solving the above linear equations, we easily obtain a parametrization

$$(A, B) = ((2 - 3r)\pi, (-1 + 2r)\pi)$$

with $r \in (1/2, 5/8)$.

We let Q be a hyperbolic rhombus obtained by gluing two copies of T; in particular, the interior angles of Q are given as $(A, 2B, A, 2B)$. Set $Q_0 := Q$. Suppose we have constructed a topological disk Q_i tiled by isometric copies of Q; in particular, ∂Q_i is a hyperbolic polygon. We assume also that all vertices of ∂Q_i belong to at most two tiles (i.e. isometric copies of Q), and some vertex v_i belongs to a single tile. Let us inductively add copies of Q on the vertices ∂Q_i starting from a corner different from v_i. We eventually obtain a topological disk Q_{i+1} so that the newly glued copies of Q intersect ∂Q_i, and so that the vertices of ∂Q_i are now in the interior of Q_{i+1}. This is possible since whenever at most three copies of Q are glued side-by-side so that the copies share a vertex, one can add more copies of Q at that vertex in such a way that that corner is completed to a full rotation; in other words, whenever one has three (possibly redundant) angles from the set $\{A, 2B\}$ one can choose more angles from the same set so that the total sum is 2π. By continuing this process $i = 0, 1, 2, \ldots$ one obtains a tessellation of $\mathbb{H}^2$ by Q, which gives a tessellation by T. It is also easy to produce infinitely different tilings in this scheme by changing the orders of gluing. This completes the proof of Theorem 7.3.6.

Note that by varying the choice of $r \in (1/2, 5/8)$ one obtains infinitely many examples of acute triangulations tessellating $\mathbb{H}^2$. One can also have different combinatorial types of the tessellation by considering a relation

$$mA + 2nB = 2\pi$$

for various choices of integers $m, n \geq 3$.

7.4 Round Spheres

So far we have mostly focused on the question of whether or not a given metric space admits an acute triangulation, possibly with certain additional desirable conditions. Let us now consider the combinatorial characterization of all possible acute triangulations of the space, in the special case of a sphere equipped with the usual round (Fubini-Study) metric.

7.4.1 Acute Triangulations from Right-Angled Hyperbolic 3-Polytopes

The simplest acute triangulation of a sphere is the regular icosahedral partition, which consists of twenty spherical triangles with dihedral angles $2\pi/5$. Applying Theorem 7.3.2 one can obtain infinitely many distinct acute triangulations with triangles of arbitrarily small diameters.

Another method of producing an acute triangulation of a sphere is as follows. Recall that the hyperbolic 3-space $\mathbb{H}^3$ has a Poincaré ball model, which is the unit open ball B^3 in $\mathbb{R}^3$ centered at the origin. Assigned with the hyperbolic metric, this open ball becomes a Riemannian 3–manifold with sectional curvature constant and equal to -1. The geodesics are either a straight line segment through the origin, or a circular arc perpendicular to $\partial B^3 = S^2$. Similarly, if a sphere S intersects $S^2 = \partial\mathbb{H}^3$ perpendicularly, then the portion of S in the unit open ball is a totally geodesic plane isometric to the hyperbolic plane $\mathbb{H}^2$. This model is conformal, that is, the Euclidean angles in this model coincide with the intrinsic Riemannian angles. At a very small scale the hyperbolic (intrinsic) metric is arbitrarily close to the Euclidean (extrinsic) metric.

A *hyperbolic 3-polytope* is a compact intersection of finitely many half-spaces in $\mathbb{H}^3$, each of which is determined by a totally geodesic plane. There exist infinitely many *right-angled* hyperbolic polytopes, i.e. having dihedral angles precisely $\pi/2$. For instance, take a very small Euclidean regular dodecahedron P_E centered at the origin of the conformal ball model. The hyperbolic convex hull $P_H := \text{conv}(V)$ of the vertices of $V := P_E^{(0)}$ is a hyperbolic 3-polytope and its dihedral angles are very near from those of P_E. In particular, the polytope P_H is obtuse. On the other hand, by radially dilating the polytope tP_H eventually approaches an *ideal* polytope P_H^∞, which is a polytope with vertices on $\partial\mathbb{H}^3 = S^2$. At the ideal vertex three faces then intersect at the angle $\pi/3$, making P_H^∞ a regular ideal acute hyperbolic polytope. It follows that at some moment $t > 1$ during the radial dilation the polytope tP_H is right-angled. We can also glue right-angled polytopes along isometric faces to produce infinitely many non-isometric right-angled ones.

Each right-angled hyperbolic 3-polytope corresponds to an open three-dimensional space of acute spherical triangulations. To describe this space, consider a right-angled hyperbolic 3-polytope P and an arbitrarily chosen point p in the interior of P. After a hyperbolic translation we may assume that $p = O$ in the Poincaré ball model of $\mathbb{H}^3$. The perpendiculars from O to the faces of P are Euclidean radial segments in the model. We now draw the intersections of these perpendiculars with a small sphere S centered at O, and join two intersection points if the corresponding two faces of P are adjacent. The resulting picture on S is a geodesic triangulation $T_{\text{perp}}(P)$ of S, which is combinatorially dual to $P^{(1)}$. Indeed, the three vertices of a triangle Δ in $T_{\text{perp}}(P)$ are on the three perpendiculars α, β, γ from O to the three faces of P that contain some vertex V_0.

The crucial point is that each triangle on S obtained as above is acute. To see this, let Q be the intersection of P with the convex cone K containing the three geodesic rays α, β and γ. Then Q is combinatorially a cube with two distinguished opposite vertices. One is the vertex V_0 of P, and the other is O. At V_0 three faces of Q are mutually orthogonal, and at O the three faces bound the cone K. We will call such a cube Q as a *slanted hyperbolic cube*; see Fig. 7.3. By elementary hyperbolic geometry one sees that the link Δ of Q at O is an acute spherical triangle. This shows that the triangulation $T_{\text{perp}}(P)$ is acute. Moreover, we can freely move P as long as O is contained in P so that the acute triangulations obtained in this way can be parametrized by the interior of P, which is an open 3-ball.

Let us give an alternative description of a slanted hyperbolic cube. We take three mutually orthogonal hyperplanes containing O in the Poincaré ball model, and pick a point V_1 in one of the open octants. By repeatedly reflecting V_1 through the three planes we obtain the points $V_1, V_2, \ldots, V_8$ corresponding to all the octants. We will call the convex hull R of those eight points as a *hyperbolic reflection cube*. The intersection of R with one of the closed octants is a slanted hyperbolic cube. As we have noted above, the link of a vertex of R is an acute spherical triangle. Furthermore, by moving V_1 around in an open octant one sees the following.

Lemma 7.4.1 *For every acute spherical triangle Δ there exists a hyperbolic reflection cube such that all of its links are isometric to Δ.*

Since the angles of the facets in a hyperbolic reflection cubes are acute we deduce that the edge lengths of an acute spherical triangle are acute. By considering polar duals, we also note that if a spherical triangle has obtuse side lengths then its dihedral angles are also obtuse.

There are many acute spherical triangulations that do not have the form of $T_{\text{perp}}(P)$. Indeed, applying small perturbations we see that the space of acute spherical triangulations combinatorially dual to $P^{(1)}$ is locally homeomorphic to

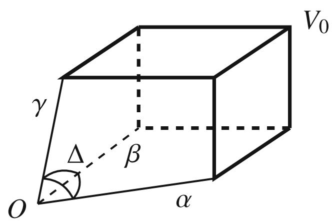


Fig. 7.3 A slanted hyperbolic cube. The dihedral angles at the edges drawn in bold are all $\pi/2$. The link Δ is an acute spherical triangle

$\mathbb{R}^m$, where m is the number of faces in P. A surprising fact is that all the acute spherical triangulations "combinatorially" come from this construction.

Theorem 7.4.2 ([33]) *A combinatorial triangulation L of a two-sphere can be realized as an acute triangulation if and only if there exists a right-angled hyperbolic 3-polytope that is combinatorially dual of L.*

We have so far shown the "if" part of the theorem, and will spend the rest of this section to briefly sketch the idea of the "only if" part.

7.4.2 The Koebe–Andreev–Thurston Theorem and Its Generalizations

Our strategy to prove the forward direction of Theorem 7.4.2 is to cook up a right-angled hyperbolic 3-polytope out of a given acute spherical triangulation. For this we will employ detailed characterizations of hyperbolic 3-polytopes due to Andreev and Thurston, further refined by Rivin–Hodgson.

A *circle packing on a sphere* can be defined as a finite collection of interior-disjoint closed disks (or, *coins*) on a sphere. It is often helpful to regard such a coin as the intersection between the closure of a half space of the Poincaré ball model in the compactification $\mathbb{H}^3 \cup \partial\mathbb{H}^3$ with the boundary $\partial\mathbb{H}^3 = S^2$. The *coin graph* of a circle packing is then defined as the combinatorial graph with vertices on the centers of the coins and with edges joining pairs of centers of mutually tangent coins.

It is trivial that every coin graph is finite and planar. The converse, that every finite planar graph is a coin graph, is a celebrated theorem of Koebe, Andreev and Thurston recalled below.

Theorem 7.4.3 (Koebe–Andreev–Thurston) *Every combinatorial triangulation of a sphere can be represented as a coin graph, in a unique manner up to conformal automorphisms of the sphere.*

Koebe's original proof is based on a form of a uniformization theorem, called the Koebe–Poincaré uniformization theorem for finite planar domains. Andreev gave a combinatorial characterization of finite-volume *non-obtuse* hyperbolic 3-polytohedra, which involves the existence of positive solutions to a certain set of linear inequalities; see [50] for a modern account of Andreev's original proof. Thurston re-interpreted Andreev's Theorem as a theorem on circle packings, generalized the latter theorem to the more general setting of Riemann surfaces [41], and used the theory of circle packings as a crucial foundation for his geometrization theorem of Haken 3–manifolds. As conjectured by Thurston and proved by Rodin–Sullivan [49], circle packings can also be used to approximate a conformal map that maps a given simply connected proper planar domain to an open disk, the existence of which is given by the Riemann mapping theorem.

Thurston's argument for circle packings has several strong consequences. One is the *double circle packing theorem*. This states that for a given combinatorial triangulation L of S^2 there exist two circle packings $\mathscr{A}$ and $\mathscr{B}$ whose coin graphs are L and its combinatorial dual L^*, respectively such that each coin from $\mathscr{A}$ is either disjoint or orthogonal to each coin in $\mathscr{B}$. Indeed, the collection $\mathscr{A}$ is already given in the Koebe–Andreev–Thurston Theorem. The collection $\mathscr{B}$ consists of the circles containing the three tangent points for each triple of mutually tangent circles in $\mathscr{A}$.

Another consequence is the *cage theorem*, which asserts that every Euclidean 3-polytope P is combinatorially isomorphic to another Euclidean polytope P' such that all the edges in P' are tangent to some common 2–sphere; in other words, the polytope P' "cages a sphere". This is straightforward from the double circle packing theorem, since one may simply choose the Euclidean planes containing circles from $\mathscr{B}$ above. These planes form facets of a desired polytope P' caging S^2. The cage theorem generalizes to the boundary of a smooth strictly convex body in $\mathbb{R}^3$, as shown by Schramm [52].

In the double circle packing theorem, the boundary circles of coins in $\mathscr{A} \cup \mathscr{B}$ determine totally geodesic hyperplanes in $\mathbb{H}^3$. These hyperplanes bound a right-angled hyperbolic 3-polyhedron that is *ideal* (i.e. having all the vertices on $\partial\mathbb{H}^3$). So, one may regard the double circle packing theorem as a special case of Andreev's theorem briefly mentioned above.

We will omit the (slightly technical) statement of Andreev's theorem in its full generality, i.e. for all non-obtuse hyperbolic 3-polytopes. For our purpose we only need the special case of right-angled hyperbolic 3-polytopes, where the statement is much simpler. Moreover, we will soon see a generalization by Rivin and Hodgson of Andreev's theorem to hyperbolic polytopes that not necessarily required to be non-obtuse.

Note that the one-skeleton of a non-obtuse (hyperbolic or Euclidean) 3-polytope is *cubic*, i.e. has valence three; for, the link of a vertex is a spherical n-gon with all dihedral angles at most $\pi/2$ and this forces that $n = 3$. Let L be a combinatorial triangulation of a sphere. We say a cycle C in $L^{(1)}$ is *separating* if each of the two open components of $L \setminus C$ contains at least one vertex from $L^{(0)}$. The following result is due to Pogorelov, preceding Andreev's paper [2]. We remark that a compact right-angled hyperbolic d-polytope exists only for $d \leq 4$; this is due to the fact that every simple d-polytope with $d \geq 5$ must have a 2–face that is a triangle or square, which can be proved by the observation in Remark 7.2.4.

Theorem 7.4.4 ([46]) *A combinatorial triangulation L of S^2 having more than four vertices is combinatorially dual to the one-skeleton of a right-angled hyperbolic 3-polytope if and only if L has no separating 3- or 4-cycles.*

For convention, we excluded the case when L has four vertices, i.e. when it is isomorphic to a tetrahedron.

7.4.3 CAT(κ) *Spaces*

For the purpose of finding out the combinatorial type of an acute spherical triangulation T, it will be necessary for us to consider some obtuse-angled hyperbolic 3-polytope P, that will arise as the combinatorial dual of T. We cannot study P by Andreev's Theorem since the latter applies only to non-obtuse polytopes. So, we will take a detour on CAT(κ)-geometry and applyy the Rivin–Hodgson theorem on hyperbolic 3-polytopes.

Consider a real number κ. We let M_κ^n denote a simply connected Riemannian n-manifold of constant curvature κ. We will only consider the case $\kappa = 1, 0$ or -1, where we have that M_κ^n is isometric to a sphere, a Euclidean space or a hyperbolic space of dimension $n \geq 2$.

Fix a value of $\kappa \in \{-1, 0, 1\}$. Let X be a metric space, and $\Delta_0 \subseteq X$ be a geodesic triangle; in the case when $\kappa = 1$ let us further assume that the perimeter ℓ of Δ_0 is less than 2π. We suppose there exists a geodesic triangle $\Delta_1 \subseteq M_\kappa^2$ that has the same side lengths as Δ_0. In particular, we have arc length parameterizations $\gamma_i \colon [0, \ell] \to \Delta_i$ such that the points $\gamma_0(t)$ and $\gamma_1(t)$ are on the vertices at the same parameters $t \in [0, \ell]$. Here, we notice that each side length of Δ_1 is necessarily less than π if $\kappa = 1$. We say Δ_0 satisfies the CAT(κ) *inequality* if

$$d_X(\gamma_0(s), \gamma(t)) \leq d_{M_\kappa^2}(\gamma_1(s), \gamma_1(t))$$

for all $s, t \in [0, \ell]$.

In the case when $\kappa \neq 1$, a geodesic space X is CAT(κ) if every geodesic triangle in X satisfies the CAT(κ) inequality. We also say that a metric space X is CAT(1) if every pair of points with distance less than π can be joined by a geodesic, and if every geodesic triangle of perimeter less than 2π satisfies the CAT(κ) *inequality*.

A space X is *locally* CAT(κ) if every point has a small neighborhood which is CAT(κ). If X is complete and if $\kappa \leq 0$, then this condition is equivalent to saying that the universal cover $\tilde{X}$ of X is CAT(κ); this result is called the *Cartan–Hadamard Theorem*, which was originally proved for Riemannian manifolds of nonpositive curvature. A CAT(κ) space is contractible whenever $\kappa \leq 0$. It is a classical result due to Alexandrov [1] that a smooth (in fact, C^3) Riemannian manifold has sectional curvature at most κ if and only if it is locally CAT(κ). See also [10, Chapter II] for a detailed discussion on CAT(κ) spaces.

A closed hyperbolic n-manifold is a typical example of CAT(-1) spaces. Different examples come from gluing hyperbolic polytopes as follows. An M_κ *cell complex* is a metric space obtained by collecting polytopes in M_κ^n for various dimensions n and gluing some of the faces by isometries. For example, a flat cone surface is an M_0 (or, Euclidean) cell complex as it admits a triangulation by Euclidean triangles with vertices on the singularities. As another example, for each right-angled hyperbolic 3-polytope P there exists a closed hyperbolic 3-manifold M tiled by eight copies of P [29]. The manifold M is a hyperbolic cell complex. Recall that the metric link of a M_κ cell complex at each vertex is a M_1 (i.e. spherical) cell complex.

Theorem 7.4.5 ([10, Chapter II.5]) *Let X be a locally finite M_κ cell complex for some $\kappa \in \{-1, 0, 1\}$.*

(1) The metric space X is locally $\mathrm{CAT}(\kappa)$ *if and only if the link of each vertex in X is* $\mathrm{CAT}(1)$.
(2) In the case when $\kappa = 1$, the metric space X is $\mathrm{CAT}(1)$ *if and only if X is locally* $\mathrm{CAT}(1)$ *and X does not contain a closed geodesic of length less than 2π.*

A two-dimensional spherical cell complex K is *strongly* $\mathrm{CAT}(1)$ if K and all of its links do not contain a closed geodesic of length at most 2π. A strongly $\mathrm{CAT}(1)$ space is $\mathrm{CAT}(1)$ by Theorem 7.4.5. Recall the *Gauss image* $\mathrm{Gauss}(P)$ of a polytope P is the spherical complex obtained by collecting the polar duals of the links of the vertices along the edges of the spherical triangles that come from the same edge of P. The spherical complex $\mathrm{Gauss}(P)$ is homeomorphic to a sphere; it is isometric to S^2 if P is Euclidean.

It is often the case that the "no-short-closed-geodesic" condition in part (2) above is quite tricky to verify. For instance, a link in a spherical complex Y of dimension two is a collection of circular arcs, and it can easily be verified whether or not such a link contains a short (less than 2π) closed geodesic. In other words, the local $\mathrm{CAT}(1)$ condition for Y is easy to check. But verifying that Y is globally $\mathrm{CAT}(1)$ often requires much more extra work; see [20] for related techniques and common difficulties. For us the following characterization of hyperbolic 3-polytopes will come for a rescue:

Theorem 7.4.6 ([28]) *A finite spherical 2-complex homeomorphic to a sphere is the Gauss image of a hyperbolic 3-polytope if and only if it is strongly* $\mathrm{CAT}(1)$.

For instance, for a spherical triangle Δ let us consider the two-dimensional spherical cell complex Y_Δ^{222} obtained from eight copies of Δ glued along the corresponding edges by the octahedral symmetry $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$; see the proof below for an alternative description of Y_Δ^{222}. Then we have the following.

Lemma 7.4.7 *If Δ is a spherical triangle whose edge lengths are all obtuse, then Y_Δ^{222} is strongly* $\mathrm{CAT}(1)$.

Proof The polar dual Δ^* of Δ is an acute spherical triangle. We have seen that there exists a hyperbolic reflection cube R whose links are all isometric to Δ^*. Since the Gauss image of R is isometric to Y_Δ^{222} the "only if" part of Theorem 7.4.6 implies that Y_Δ^{222} is strongly $\mathrm{CAT}(1)$. □

Let L be a finite simplicial complex. The *right-angled Coxeter group* on L is defined as the group presentation

$$C(L) := \langle v \in L^{(0)} \mid v^2 = 1 \text{ for } v \in L^{(0} \text{ and } uv = vu \text{ for } \{u, v\} \in L^{(1)} \rangle.$$

We say that L admits a *separating clique* if L can be written as the union of two proper subcomplexes L_1 and L_2 such that the vertices of $K := L_1 \cap L_2$ form a clique in L and such that $L_1 \setminus K$ is disconnected from $L_2 \setminus K$; in this case, we have

$$C(L) := C(L_1) *_A C(L_2)$$

for the finite group $A = C(K) \cong \mathbb{Z}_2^{\# K^{(0)}}$; in particular, $C(L)$ is not one–ended [56].

Recall that a geodesic metric space X is called δ-*hyperbolic* for some $\delta > 0$ if for each geodesic triangle ABC in X the union of the two sides $AB \cup BC$ is contained in the δ-neighborhood of the remaining side AC; see [25]. It is an elementary exercise to show that the hyperbolic plane is δ-hyperbolic for some $\delta > 0$, which implies that every CAT(-1) space is also δ-hyperbolic.

A finitely generated group is said to be *word-hyperbolic* if it properly and cocompactly acts on a δ-hyperbolic space by isometries. It follows that every finitely generated group acting properly and cocompactly on a CAT(-1) space by isometries is word-hyperbolic.

It is well-known that a word-hyperbolic group does not contain $\mathbb{Z} \times \mathbb{Z}$. As a side remark, we note that it is an outstanding conjecture that a finitely generated group acting properly and cocompactly by isometries on a CAT(0) space is word-hyperbolic if it does not contain $\mathbb{Z} \times \mathbb{Z}$. Note that if L is a square then

$$C(L) \cong (\mathbb{Z}_2 * \mathbb{Z}_2) \times (\mathbb{Z}_2 * \mathbb{Z}_2)$$

contains a copy of $\mathbb{Z}^2$. In particular, if $C(L)$ is word–hyperbolic then L does not contain a chordless square, i.e. a set of four vertices $\{a, b, c, d\}$ that spans a square in L. For right-angled Coxeter groups, the strong converses of the two aforementioned remarks on one-endedness and word-hyperbolicity hold.

Lemma 7.4.8 *For a finite simplicial complex L the following hold.*

(1) The group $C(L)$ is one-ended and if and only if L does not admit a separating clique;
(2) The group $C(L)$ is word-hyperbolic if and only if $L^{(1)}$ does not contain a chordless square.

We are now ready to complete the proof of the main result in this section.

Proof (Proof of Theorem 7.4.2) Let T be a given acute triangulation of S^2. For each $v \in T^{(0)} \subseteq S^2$ we let Π_v denote the tangent plane to S^2 at the point v. Denote by Π_v^+ the closed half space containing 0 and bounding Π_v. We have a Euclidean polytope

$$P_E := \bigcap_v \Pi_v^+.$$

The 3-polytope P_E has the Gauss image T, and is *strongly obtuse*, which means that the face angles and the dihedral angles are all obtuse. For a sufficiently small $t > 0$, the convex hull P_H of the points $tP_E^{(0)}$ in the Poincaré ball model is a hyperbolic 3-polytope that is close to the Euclidean polytope tP_1. In particular, we can require that P_H is also strongly obtuse and combinatorially isomorphic to P_E.

A hyperbolic cell complex $X(P_H)$ is defined as follows. The (reduced) Cayley graph Cayley($C(T)$) is an undirected graph whose vertex set is $C(T)$ and whose edge set consists of the unordered pairs $\{g, gv\}$ for each $g \in C(T)$ and $v \in T^{(0)}$. We place a copy of P_H on the vertices in the Cayley graph Cayley($C(T)$) of $C(T)$ and whenever there is an edge between two such copies we isometrically glue the corresponding faces; more precisely, we let F_v denote the face of P_H corresponding to $v \in T^{(0)}$ and define an equivalence relation $\sim$ on $P_H \times C(L)$ by

$$(x, g) \sim (x, gv)$$

whenever $v \in T^{(0)}$ and $x \in F_v$. We then obtain a hyperbolic cell complex

$$X(P_H) := P_H \times C(T)/\sim .$$

Let us consider the combinatorial dual $\tilde{\Sigma}_T$ of $X(P_H)$, which has a description of 2- and 3-cells as follows. The one-skeleton of $\tilde{\Sigma}_T$ coincides with Cayley($C(T)$). Whenever there is a 4-cycle in Cayley($C(T)$) written as

$$(g, gu, guv, gv)$$

for some element $g \in C(T)$ and for some edge $\{u, v\}$ in T, we glue a 2-cell along the 4–cycle. Similarly, whenever we have eight vertices in Cayley($C(T)$) that can be written as

$$gu^p v^q w^r$$

for $g \in C(T)$, $\{u, v, w\} \in T^{(2)}$ and $p, q, r \in \{0, 1\}$, we glue a 3–cube along those vertices. When T has no separating 3-cycle, then $\tilde{\Sigma}_T$ coincides with the *Davis complex* of $C(T)$, which can be defined for an arbitrary Coxeter group [17].

In our situation where T is a triangulation of a sphere not equal to the boundary of a tetrahedron, the following three conditions are easily verified to be equivalent [17, Proposition 1.2.3]:

- $C(T)$ is one-ended;
- $\tilde{\Sigma}_T \approx X(P_H)$ is contractible;
- the natural cubical metric on $\tilde{\Sigma}_T$ is CAT(0).

Let v be a vertex in P_H, and Δ be its link in P_H. Then the link of a vertex in $X(P_H)$ that comes from v is isometric to the complex $Y^{222}(\Delta)$. By Theorem 7.4.5 and Lemma 7.4.7, we see that $X(P_H)$ is locally CAT(-1). Since $X(P_H) \approx \tilde{\Sigma}_T$ is simply connected, we see from the Cartan–Hadamard Theorem that $X(P_H)$ is CAT(-1). We saw above that the contractibility of $X(P_H)$ implies that $C(T)$ is one-ended. The space $X(P_H)$ admits a natural proper cocompact isometric action of $C(T)$

$$g.(x, w) = (x, gw)$$

for all $x \in P$ and $g, w \in C(T)$. Since $X(P_H)$ is CAT(−1), we conclude that $C(T)$ is one-ended and word-hyperbolic. We conclude from Lemma 7.4.7 that T has no separating 3- or 4-cycle. □

Suppose an acute triangulations L of S^2 has n faces. Using the hypothesis that the valences of the vertices are at least five, one can easily deduce from the Euler characteristic formula that n is an even number not smaller than 20, and that $n \neq 22$. Itoh [30] explicitly constructed examples of acute triangulations for all such even numbers n except for the cases $n = 28$ and $n = 34$.

As a consequence of Theorems 7.4.2 and 7.4.4, one can algorithmically (i.e. in finite time) recognize whether or not a given spherical combinatorial triangulation L is acutely realizable. Using this fact, Walsh and the author constructed examples of spherical acute triangulations with 28 and 34 faces by computer search, complementing Itoh's result above.

Corollary 7.4.9 ([30] for $n \neq 28, 34$; [33] for $n = 28, 34$) *There exists an acute triangulation of S^2 with n faces if and only if n is an even number satisfying that $n \geq 20$ and that $n \neq 22$.*

A similar statement to Theorem 7.4.2 holds for a Euclidean polygon, as proved by Maehara [40] using plane geometry. Namely, a combinatorial triangulation of a disk admits an acute triangulation in $\mathbb{R}^2$ if and only if there are no cycles of length at most four that bounds an open disk containing at least one vertex. This result can alternatively be deduced from the method of this section. Furthermore, Theorem 7.4.2 generalizes to a combinatorial characterization of acutely triangulated planar surfaces in S^2. For instance, one can show [33] that a hemisphere has an acute triangulation, by using the triangulation in Fig. 7.4.

Let L be a combinatorial triangulation of S^2 having no separating 3- or 4-cycles. We know from Theorem 7.4.2 that the space $\mathscr{A}(L)$ of acute triangulations realizing L is nonempty. Moreover, if n denotes the number of vertices in L then $\mathscr{A}(L)$ is an open n-manifold contained in the configuration space $\mathrm{Conf}_n(S^2)$ of n points in

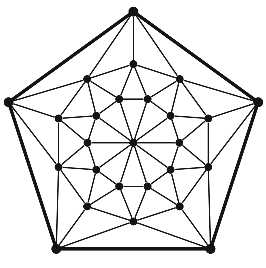


Fig. 7.4 An acute geodesic triangulation of a hemisphere. This graph appeared in [40]

S^2. The configuration space is connected and its fundamental group is the n-strand braid group on S^2. On the other hand, very little is known about the topology of $\mathscr{A}(L)$.

Problem 7.8 Is the space $\mathscr{A}(L)$ connected? What is its homotopy type?

It is a highly nontrivial result due to Cairns that the space $\mathrm{GT}(L)$ of all geodesic triangulations of S^2 isomorphic to L is path-connected [13]. An outstanding conjecture of Cairns asserts that $\mathrm{GT}(L)$ has the homotopy type of SO(3); see [4, 7] for a relevant work.

Acknowledgements The author is supported by Samsung Science and Technology Foundation (SSTF-BA1301–51) and by a KIAS Individual Grant (MG073601) at Korea Institute for Advanced Study. The author is grateful for helpful comments from Thomas Koberda, Jaejeong Lee, and Genevieve Walsh.

References

1. A.D. Aleksandrov, A theorem on triangles in a metric space and some of its applications, in *Trudy Mat. Inst. Steklov., v 38, Trudy Mat. Inst. Steklov., v 38* (Izdat. Akad. Nauk SSSR, Moscow, 1951), pp. 5–23
2. E.M. Andreev, Convex polyhedra in Lobačevskiĭ spaces. Mat. Sb. (N.S.) **81**(123)(123), 445–478 (1970)
3. Aristotle, *On the heavens*. Translated by J. L. Stocks, The Internet Classics Archive (2009)
4. M. Awartani, D.W. Henderson, Spaces of geodesic triangulations of the sphere. Trans. Am. Math. Soc. **304**(2), 721–732 (1987)
5. C.J. Bishop, Uniformly acute triangulations of polygons (2021). preprint. Available at https://www.math.stonybrook.edu/~bishop/papers/acutepoly.pdf
6. C.J. Bishop, L. Rempe, Non-compact riemann surfaces are equilaterally triangulable (2021)
7. E.D. Bloch, R. Connelly, D.W. Henderson, The space of simplexwise linear homeomorphisms of a convex 2-disk. Topology **23**(2), 161–175 (1984)
8. J. Brandts, S. Korotov, M. Křížek, *Simplicial Partitions with applications to the Finite Element Method*. Springer Monographs in Mathematics (Springer, Cham, 2020)
9. J. Brandts, S. Korotov, M. Křížek, J. Šolc, On nonobtuse simplicial partitions. SIAM Rev. **51**(2), 317–335 (2009)
10. M.R. Bridson, A. Haefliger, *Metric spaces of non-positive curvature*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 319 (Springer, Berlin, 1999)
11. Y.D. Burago, V.A. Zalgaller, Polyhedral embedding of a net. Vestnik Leningrad. Univ. **15**(7), 66–80 (1960)
12. Y.D. Burago, V.A. Zalgaller, Isometric piecewise-linear embeddings of two-dimensional manifolds with a polyhedral metric into $\mathbf{R}^3$. Algebra i Analiz **7**(3), 76–95 (1995)
13. S.S. Cairns, Isotopic deformations of geodesic complexes on the 2-sphere and on the plane. Ann. Math. (2) **45**, 207–217 (1944)
14. Y. Colin de Verdière, A. Marin, Triangulations presque équilatérales des surfaces. J. Differ. Geom. **32**(1), 199–207 (1990)
15. Y. Colin de Verdière, Comment rendre géodésique une triangulation d'une surface?. Enseign. Math. (2) **37**(3–4), 201–212 (1991)

16. Y. Colin de Verdière, *Spectres de graphes*. Cours Spécialisés [Specialized Courses], vol. 4 (Société Mathématique de France, Paris, 1998)
17. M.W. Davis, *The Geometry and Topology of Coxeter Groups*. London Mathematical Society Monographs Series, vol. 32 (Princeton University Press, Princeton, NJ, 2008)
18. P. Deligne, G.D. Mostow, Monodromy of Hypergeometric functions and Nonlattice Integral Monodromy. Inst. Hautes Études Sci. Publ. Math. **63**, 5–89 (1986)
19. P. Deligne, G.D. Mostow, *Commensurabilities Among Lattices in* $\mathrm{PU}(1, n)$. Annals of Mathematics Studies, vol. 132 (Princeton University Press, Princeton, NJ, 1993)
20. M. Elder, J. McCammond, CAT(0) is an algorithmic property. Geom. Dedicata **107**, 25–46 (2004)
21. P. Engel, P. Smillie, The number of convex tilings of the sphere by triangles, squares, or hexagons. Geom. Topol. **22**(5), 2839–2864 (2018)
22. D. Eppstein, J.M. Sullivan, A. Üngör, Tiling space and slabs with acute tetrahedra. Comput. Geom. **27**(3), 237–255 (2004)
23. M. Gardner, *New mathematical Diversions*, revised ed. (MAA Spectrum, Mathematical Association of America, Washington, DC, 1995)
24. M. Goldberg, W. Manheimer, Elementary problems and solutions: solutions: E1406. Am. Math. Monthly **67**(9), 923 (1960)
25. M. Gromov, *Hyperbolic groups*, Essays in group theory, Math. Sci. Res. Inst. Publ., vol. 8, Springer, New York, 1987, pp. 75–263.
26. M. Gromov, Geometric, algebraic, and analytic descendants of Nash isometric embedding theorems. Bull. Am. Math. Soc. (N.S.) **54**(2), 173–245 (2017)
27. B. Grünbaum, *Convex Polytopes*, 2nd ed. Graduate Texts in Mathematics, vol. 221 (Springer, New York, 2003). Prepared and with a preface by Volker Kaibel, Victor Klee and Günter M. Ziegler
28. C.D. Hodgson, I. Rivin, A characterization of compact convex polyhedra in hyperbolic 3-space. Invent. Math. **111**(1), 77–111 (1993)
29. T. Inoue, Organizing volumes of right-angled hyperbolic polyhedra. Algebr. Geom. Topol. **8**(3), 1523–1565 (2008)
30. J. Itoh, Acute triangulations of sphere and icosahedron, Differential geometry (Sakado, 2001). Josai Math. Monogr., vol. 3 (Josai Univ., Sakado, 2001), pp. 53–62
31. G. Kalai, On low-dimensional faces that high-dimensional polytopes must have. Combinatorica **10**(3), 271–280 (1990)
32. G. Kalai, Rigidity and the lower bound theorem. I. Invent. Math. **88**(1), 125–151 (1987)
33. S.-H. Kim, G.S. Walsh, Coxeter groups, hyperbolic cubes and acute triangulations. J. Topol. **9**(1), 117–142 (2016)
34. V. Klee, A combinatorial analogue of Poincaré's duality theorem. Canadian J. Math. **16**, 517–531 (1964)
35. E. Kopczyński, I. Pak, P. Przytycki, Acute triangulations of polyhedra and $\mathbb{R}^N$. Combinatorica **32**(1), 85–110 (2012)
36. V. Koziarz, D.-M. Nguyen, Complex hyperbolic volume and intersection of boundary divisors in moduli spaces of pointed genus zero curves. Ann. Sci. Éc. Norm. Supér. (4) **51**(6), 1549–1597 (2018)
37. M. Křížek, Erratum to: There is no face-to-face partition of $\mathbf{R}^5$ into acute simplices [mr2252110]. Discrete Comput. Geom. **44**(2), 484–485 (2010)
38. M. Křížek, There is no face-to-face partition of $\mathbf{R}^5$ into acute simplices. Discrete Comput. Geom. **36**(2), 381–390 (2006)
39. H. Maehara, Acute triangulations of polygons. Eur. J. Combin. **23**(1), 45–55 (2002)
40. H. Maehara, Plane graphs with straight edges whose bounded faces are acute triangles. J. Combin. Theory. Ser. B **88**(2), 237–245 (2003)
41. A. Marden, B. Rodin, On Thurston's formulation and proof of Andreev's theorem, in *Computational Methods and Function Theory* (Valparaíso, 1989). Lecture Notes in Math., vol. 1435 (Springer, Berlin, 1990), pp. 103–115

42. G.A. Margulis, S. Mozes, Aperiodic tilings of the hyperbolic plane by convex polygons. Israel J. Math. **107**, 319–325 (1998)
43. C.T. McMullen, The Gauss-Bonnet theorem for cone manifolds and volumes of moduli spaces. Am. J. Math. **139**(1), 261–291 (2017)
44. G.D. Mostow, Generalized Picard lattices arising from half-integral conditions. Inst. Hautes Études Sci. Publ. Math. **63**, 91–106 (1986)
45. J.R. Parker, Complex hyperbolic lattices, in *Discrete Groups and Geometric Structures*. Contemp. Math., vol. 501 (Amer. Math. Soc., Providence, RI, 2009), pp. 1–42
46. A.V. Pogorelov, Regular decomposition of the Lobačevskiĭ space. Mat. Zametki **1**, 3–8 (1967)
47. M. Rao, Exhaustive search of convex pentagons which tile the plane (2017)
48. J. Richter-Gebert, *Realization Spaces of Polytopes*. Lecture Notes in Mathematics, vol. 1643 (Springer, Berlin, 1996)
49. B. Rodin, D. Sullivan, The convergence of circle packings to the Riemann mapping. J. Differ. Geom. **26**(2), 349–360 (1987)
50. R.K.W. Roeder, J.H. Hubbard, W.D. Dunbar, Andreev's theorem on hyperbolic polyhedra. Ann. Inst. Fourier (Grenoble) **57**(3), 825–882 (2007)
51. S. Saraf, Acute and nonobtuse triangulations of polyhedral surfaces. Eur. J. Combin. **30**(4), 833–840 (2009)
52. O. Schramm, How to cage an egg. Invent. Math. **107**(3), 543–560 (1992)
53. R.E. Schwartz, Notes on shapes of polyhedra (2015)
54. M. Senechal, Which tetrahedra fill space? Math. Mag. **54**(5), 227–243 (1981)
55. J.R. Shewchuk, What is a good linear element? - interpolation, conditioning, and quality measures, in 11th International Meshing Roundtable (2002), pp. 115–126
56. J. Stallings, *Group Theory and Three-Dimensional Manifolds*. Yale Mathematical Monographs, vol. 4 (Yale University Press, New Haven, Conn.-London, 1971), A James K. Whittemore Lecture in Mathematics given at Yale University, 1969
57. D.J. Struik, Het probleem 'de impletione loci' (dutch). Nieuw. Archief voor Wiskunde **15**(3), 121–137 (1925—1928)
58. W.P. Thurston, Conway's tiling groups. Am. Math. Monthly **97**(8), 757–773 (1990)
59. W.P. Thurston, *Three-Dimensional Geometry and Topology.*, vol. 1. Princeton Mathematical Series, vol. 35 (Princeton University Press, Princeton, NJ, 1997). See also its unpublished web version http://library.msri.org/books/gt3m/PDF/Thurston-gt3m.pdf
60. W.P. Thurston, Shapes of polyhedra and triangulations of the sphere, in *The Epstein Birthday Schrift*. Geom. Topol. Monogr., vol. 1 (Geom. Topol. Publ., Coventry, 1998), pp. 511–549
61. M. Troyanov, Les surfaces euclidiennes à singularités coniques. Enseign. Math. (2) **32**(1–2), 79–94 (1986)
62. W.T. Tutte, A census of planar triangulations. Canadian J. Math. **14**, 21–38 (1962)
63. A. Üngör, A. Sheffer, *Pitching Tents in Space-Time: Mesh Generation for Discontinuous Galerkin Method*, vol. 13, 2002, Volume and surface triangulations, pp. 201–221
64. E. Vanderzee, A.N. Hirani, D. Guoy, E.A. Ramos, Well-centered triangulation. SIAM J. Sci. Comput. **31**(6), 4497–4523 (2009/10)
65. È.B. Vinberg, Absence of crystallographic groups of reflections in Lobachevskiĭ spaces of large dimension. Trudy Moskov. Mat. Obshch. **47**, 68–102, 246 (1984)
66. V.A. Voevodskiĭ, G.B. Shabat, Equilateral triangulations of Riemann surfaces, and curves over algebraic number fields. Dokl. Akad. Nauk SSSR **304**(2), 265–268 (1989)
67. C.T. Zamfirescu, Survey of two-dimensional acute triangulations. Discrete Math. **313**(1), 35–49 (2013)

Chapter 8
Signature Calculation of the Area Hermitian Form on Some Spaces of Polygons

İsmail Sağlam

Abstract This chapter is motivated by the paper by Thurston on triangulations of the sphere and singular flat metrics on the sphere. Thurston gave a local parametrization of the moduli space of singular flat metrics on the sphere with prescribed positive curvature data at the singular points by a complex hyperbolic space of an appropriate dimension. This work can be considered as a generalization of the signature calculation of the Hermitian form that he made in his paper.

The moduli space of singular flat metrics on the sphere having unit area and with prescribed curvature data at the singular points can be locally parametrized by certain spaces of polygons. This can be done by cutting singular flat spheres through length minimizing geodesics from a fixed singular point to the others. In that case the space of polygons is a complex vector space of dimension $n-1$ when there are $n+1$ singular points. There is natural area Hermitian form of signature $(1, n-2)$ on this vector space. In this chapter we calculate the signature of the area Hermitian form on some spaces of polygons which locally parametrize the moduli space of singular flat metrics having unit area on the sphere with one singular point of negative curvature. The formula we obtain depends only on the sum of the curvatures of the singular points having positive curvature.

Keywords Singular flat metric · Singular flat surface · Polygon · Hermitian form · Alexandrov unfolding process

Mathematics Subject Classification 51F99, 57M50, 15A69

İ. Sağlam (✉)
Department of Aerospace Engineering, Adana Alparslan Turkes Science and Technology University, Adana, Turkey

Institut de Recherche Mathematique Avancé, CNRS et Université de Strasbourg, Strasbourg, France
e-mail: isaglam@atu.edu.tr

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_8

8.1 Introduction

Let S be a compact singular flat surface perhaps with boundary. By this we mean that

- each interior point of S has a neighborhood isometric to a neighborhood of the apex of a standard cone,
- near each boundary point S possesses the geometry of a surface obtained by cutting a cone through a line passing from the apex.

There is a well-defined notion of angle for each point $p \in S$. Let us denote the angle at p by $\theta(p)$. If p is an interior point we define the curvature at p to be $\kappa_p = 2\pi - \theta(p)$. If p is a boundary point, then $\kappa_p = \pi - \theta(p)$. Note that a point $p \in S$ is called singular if $\kappa_p \neq 0$. Otherwise p is called non-singular.

In [15] Thurston considered the moduli space of singular flat structures on the sphere with prescribed positive curvature data. He showed that if the number of singular points on the sphere is n, then this moduli space is a complex hyperbolic manifold of dimension $n - 3$. To achieve this, he considered singular flat metrics on the sphere and triangulated the sphere so that the vertices of the triangulation are exactly the singular points of the metric. Observing that nearby singular flat metrics admit the same combinatorial triangulation, he obtained local coordinates from the moduli space to the projectifization of the positive part of a certain cocycle space equipped with the Hermitian form induced by the area of a singular flat structure. He showed that the signature of the Hermitian form is $(1, n-3)$, where n is the number of singular points of the sphere. Note that this implies that the local coordinates are from the moduli space to the complex hyperbolic space of dimension $n - 3$. In this chapter we make a similar signature calculation for the case of singular flat spheres with one singular point of negative curvature. Now we return to the theory of triangulation of a singular flat surface S.

It is well known that a singular flat surface S can be triangulated by Euclidean triangles. See [16]. It follows that one can obtain any compact singular flat surface from a finite numbers of Euclidean triangles. However, the result in [16] does not give us any constructive method to obtain a singular flat surface from triangles in the Euclidean plane.

There is a stronger result which says that a triangulation with a minimum number of triangles exists. More precisely, this theorem states that S has a triangulation whose vertex set coincides with the set of singular points of the S. See [6], [14] and [13] for proofs of this fact. However, even by this method, it is not clear how one can construct a singular flat surface from a collection of triangles in the Euclidean plane.

Another way to construct singular flat surfaces is to use flat disks instead of triangles. Here, a flat disk is a singular flat surface which is homeomorphic to a closed disk and has no singular interior points. It is not difficult to see that for any compact singular flat surface S, there exists a flat disk D so that S can be obtained by gluing some of the edges of D appropriately. In [13], it was shown that D can be chosen so that it has a minimum number of edges. Equivalently, there exists a finite number of simple geodesic arcs on S that can intersect only at their endpoints so that when we cut S through these arcs we get a flat disk.

When S has genus 0 and each point in S has angle less than 2π, there is a nice way to construct S from a flat disk. Assume that S has $n+1$ singular points $P_0, P_1, \ldots, P_n$ of positive curvature, that is, assume that the angle at P_i, $\theta(P_i)$, is less than 2π for each i. Let l_i be a length-minimizing geodesic joining P_0 and P_i for each $i > 0$. Then it can be shown that each l_i is simple and l_i intersects l_j only at P_0 when $i \neq j$. It follows that we get a flat disk if we cut S through $l_1, l_2, \ldots, l_n$. The Alexandrov Unfolding Theorem [1] states that this polygon can be embedded into the Euclidean plane. Therefore, any singular flat sphere with angle less than 2π at each singular point can be obtained from a polygon in the Euclidean plane.

As an example, consider the polygon in Fig. 8.1. The edges that are denoted by the same letters have the same length. The polygon has two vertices with angle π: A and B. It has also one vertex having angle $\frac{3\pi}{2}$. If we glue a and a', b and b', c and c', then we get a flat sphere. Note that this flat sphere has four singular points and three of them are obtained from A, B and C. At these points the angles are π,π and $\frac{3\pi}{2}$, respectively. Also, after this gluing operation the vertices X,Y and Z of the polygon come together to form a singular point of the flat sphere. This singular point has angle $\frac{\pi}{2}$. Note that the Alexandrov Unfolding Theorem states that any singular flat sphere with 4 singular points of angle $\frac{\pi}{2}$, $\frac{3\pi}{2}$, π, π can be obtained from such a polygon in the Euclidean plane.

Let us consider the general case again, that is, the case where S has genus 0 and $n+1$ singular points $P_0, \ldots, P_n$ so that each $\theta(P_i)$ is less than 2π. If we cut S through $l_1, \ldots, l_n$, then we get a flat disk which is isometric to a polygon. Note that

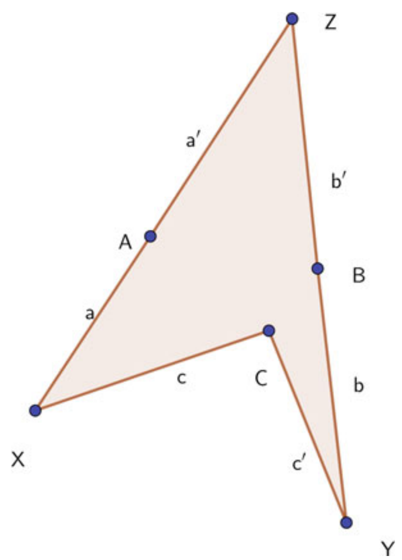


Fig. 8.1 A flat sphere with 4 singular points from a planer polygon

this polygon is not arbitrary since it satisfies certain conditions. First of all, it has $2n$ vertices and $2n$ edges. It has n vertices coming from $P_1, \ldots, P_n$ so that the interior angles of the polygon at these vertices are $\theta(P_1), \ldots \theta(P_n)$. Let

$$\kappa_i = 2\pi - \theta(P_i) \text{ for } i = 0, 1, \ldots, n.$$

The Gauss-Bonnet Formula [18] implies that

$$\sum_{i=0}^{n} \kappa_i = 4\pi$$

and we have

$$0 < \kappa_0 = 4\pi - \sum_{i=1}^{n} \kappa_i < 2\pi.$$

This polygon has n vertices which are induced from P_0 and the sum of the angles at these vertices is equal to P_0. Furthermore, a vertex which comes from the singular point P_i ($1 \leq i \leq n$) is incident to two edges of the same length. Let us move the polygon in the plane so that one of the vertices induced from P_0 is at the origin in the complex plane. If we assume that the vertices of the polygon having angles $\theta(P_1), \ldots, \theta(P_n)$ are in counter-clockwise orientation, then these vertices give us an element $\hat{z} = (z_1, \ldots, z_{2n}) \in \mathbb{C}^{2n}$ so that

1. $z_1 = 0$,
2. $e^{\imath \kappa_k}(z_{2k-1} - z_{2k}) = z_{2k+1} - z_{2k}, \ 1 \leq k \leq n.$

Here, $\imath$ denotes the complex number $\sqrt{-1}$. The elements in $\mathbb{C}^{2n}$ which satisfy the above conditions form an $n-1$ dimensional complex vector space. We denote this vector space by $\mathfrak{P}(\kappa) = \mathfrak{P}(\kappa_1, \ldots, \kappa_n)$. Note that the elements in $\mathfrak{P}(\kappa)$ can be considered as possibly self-intersecting polygons. If such a polygon is positively oriented and does not have any self-intersection, then it represents a flat sphere. In this case the flat sphere and the polygon have the same area. If this polygon has coordinates $\hat{z} = (z_1, \ldots, z_{2n})$, then its area is given by the following formula:

$$\frac{\sqrt{-1}}{4} \sum_{i=1}^{2n-1} (z_i \overline{z}_{i+1} - z_{i+1} \overline{z}_i).$$

Therefore it is natural to consider the following area Hermitian form h_A on $\mathfrak{P}(\kappa)$:

$$h_A(\hat{z}, \hat{w}) = \frac{\sqrt{-1}}{4} \sum_{i=1}^{2n-1} (z_i \overline{w}_{i+1} - z_{i+1} \overline{w}_i).$$

This Hermitian form is called *area Hermitian form*. Fillastre [3] computed the signature of the area Hermitian form and showed that it is $(1, n-2)$. Note that this computation is consistent with the one that was done by Thurston [15].

In this chapter, we calculate the signature of the area Hermition form h_A on $\mathfrak{P}(\kappa) = \mathfrak{P}(\kappa_1, \ldots, \kappa_n)$ by dropping some conditions on the curvature data $\kappa = (\kappa_1, \ldots, \kappa_n)$. As before, we assume that $0 < \kappa_i < 2\pi$ for each $1 \leq i \leq n$. But we do not require that

$$0 < \kappa_0 = 4\pi - \sum_{i=1}^{n} \kappa_i < 2\pi.$$

That is, κ_0 may be any number less than 4π. This has the following geometric significance. Assume that $\kappa_1 + \cdots + \kappa_n > 4\pi$ and $\hat{z} \in \mathfrak{P}(\kappa)$ is a positively oriented polygon. Then we can obtain a flat sphere by identifying equal edges of this polygon appropriately. This flat sphere has $n+1$ singular points and one of them has angle

$$2\pi - \kappa_0 = 2\pi - (4\pi - (\kappa_1 + \cdots + \kappa_n)) > 2\pi.$$

Therefore, at least some part of the moduli space of flat spheres with exactly one singular point having angle greater than 2π can be parametrized by using $\mathfrak{P}(\kappa)$, for some κ. This gives us the hope to endow the moduli space of flat spheres with prescribed curvature data with new geometric structures.

The paper [10] is closely related to the present one. Indeed our work can be considered as a generalization of this paper by Nishi and Ohshika in which the authors calculated the signature of the area Hermitian form on $\mathfrak{P}(\kappa)$, where $\kappa = (\pi, \ldots, \pi)$. This calculation led them to put a pseudo-metric on the moduli space/Teichmüller space of flat metrics on the sphere with $n+1$ singular points, where the cone angles are $(n-2)\pi, \pi, \ldots, \pi$. This new metric structure enabled them to put a pseudo-metric on the moduli space of hyperelliptic curves, since a hyperelliptic curve is a degree 2 branched cover of the complex projective line. We also point out that the papers [7] and [9] are closely related to the present one. Furthermore, in [2], the authors calculated the signature of a symmetric bilinear form on a space of polygons. The formulae and the proofs given in this chapter are similar. Finally, we note that Nishi [8] addressed the question on the signature of a Hermitian form given by the area function on the space of singular flat metrics on the sphere with conical singularities of possibly negative curvatures.

For more information on the geometry of polygons, see [4] and [5]. For more information on the geometry of flat surfaces, see [12, 16–18].

8.2 Basic Facts on Hermitian Forms

In this section we introduce the main facts that we use from the theory of the Hermitian forms. Let V be a complex vector space. A Hermitian form on V is a function

$$h : V \times V \to \mathbb{C}$$

such that

1.
$$h(\alpha u + \beta v, w) = \alpha(u, w) + \beta h(u, w)$$
for all $u, v, w \in V$ and for all $\alpha, \beta \in \mathbb{C}$,
2.
$$h(w, \alpha u + \beta v) = \bar{\alpha} h(w, u) + \bar{\beta} h(w, v)$$
for all $u, v, w \in V$ and for all $\alpha, \beta \in \mathbb{C}$.
3.
$$h(u, v) = \overline{h(v, u)}$$
for all $u, v \in V$.

Note that if $u \in V$, then $h(u, u) \in \mathbb{R}$. $h(u, u)$ is called the square-norm of u.

Assume that V has dimension n and $\mathcal{U} = \{u_1, \ldots u_n\}$ is an ordered basis for V. Then the matrix

$$H = [h(u_i, u_j)]$$

is called the matrix of h in the ordered basis $\mathcal{U}$. This matrix has the property that $H_{ji} = \overline{H}_{ij}$ for all $1 \leq i, j \leq n$. That is, the transpose of the conjugate matrix of H, which is denoted by H^*, is equal to H. Note that such a matrix is called Hermitian matrix.

Definition 8.1 The rank of a Hermitian form h is the rank of the matrix H. It is denoted by $Rank(h)$.

Definition 8.2 A Hermitian form h on an n-dimensional complex vector space V is called non-singular if $Rank(h) = n$.

Note that h is non-singular if and only if for each $u \neq 0 \in V$ there exist $v \in V$ such that $h(u, v) \neq 0$.

Definition 8.3 Let h be a Hermitian form on a complex vector space V.

- h is called positive definite if $h(u, u) > 0$ for all $u \neq 0 \in V$.
- h is called negative definite if $h(u, u) < 0$ for all $u \neq 0 \in V$.

The following theorem asserts that each finite-dimensional complex vector space has a basis $\mathcal{U}$ such that the matrix of h in $\mathcal{U}$ is diagonal.

Theorem 8.1 *Let V be a finite-dimensional complex vector space and let h be a Hermitian form on V. Then there is an ordered basis for V in which h is represented by a diagonal matrix.*

One can sharpen the previous theorem so that the entries of the diagonal matrix are 1, −1 or 0. Here is the precise statement.

Theorem 8.2 *Let V be an n-dimensional complex vector space and h be a Hermitian form on V which has rank r. Then there is an ordered basis $\{u_1, \ldots, u_n\}$ for V in which the matrix of h is diagonal and such that*

$$h(u_j, u_j) = \pm 1 \textit{ for all } j = 1, \ldots r,$$

$$h(u_j, u_j) = 0 \textit{ for all } j = r + 1, \ldots, n.$$

Furthermore, the number of vectors u_j such that $h(u_j, u_j) = 1$, $h(u_j, u_j) = -1$ and $h(u_j, u_j) = 0$ are independent of the choice of basis.

We say that two elements $u, v \in V$ are orthogonal if $h(u, v) = 0$. Let W be a subspace of V. Let us define $W^{\perp}$ as the subspace of V which consists of elements of V that are orthogonal to each element in W. Note that $V^{\perp}$ has dimension $n - r$ and it has a basis $\{u_{r+1} \ldots, u_n\}$.

Let U and W be subspaces of V such that

1. $U \cap W = 0$;
2. any element in $v \in V$ can be written as $v = u + w$, where $u \in U$ and $w \in W$;
3. $h(u, w) = 0$ for all $u \in U$ and $w \in W$.

In this case we write

$$V = U \bigoplus W,$$

and call W an orthogonal complement of U in V.

Let us denote the cardinality of a set A by $|A|$.

Definition 8.4 Let V be an n-dimensional complex vector space and h a Hermitian form on V. Let $\mathcal{U} = \{u_1, \ldots u_n\}$ be a basis as in Theorem 8.2. We introduce the following quantities:

1. $P(h) = |\{u_j : h(u_j, u_j) = 1\}|$;
2. $N(h) = |\{u_j : h(u_j, u_j) = -1\}|$;

3. $Z(h) = |\{u_j : h(u_j, u_j) = 0\}|$.

$(P(h), N(h))$ is called signature of h.

Clearly, $P(h) + N(h) = Rank(h)$ and the dimension of $V^\perp$ is equal to $Z(h)$. Also, $P(h) + N(h) + Z(h)$ is equal to n. We have $Z(h) = 0$ if and only if h is non-singular. Furthermore, h is positive definite if and only if $P(h) = n$, and h is negative definite if and only if $N(h) = n$.

Now we define isomorphisms of complex vector space equipped with Hermitian forms. Let V and V' be complex vector spaces together with Hermitian forms h and h'. We say that V and V' are isomorphic as vector spaces equipped with Hermitian forms if there is a vector space isomorphism $f : V \to V'$ such that

$$h(v, w) = h'(f(v), f(w))$$

for all $v, w \in V$. Note that f is an isomorphism vector spaces having Hermitian forms if it satisfies the following weaker condition:

$$h(u, u) = h'(f(u), f(u))$$

for all $u \in V$. Note that the rank and signature of a Hermitian form is invariant under isomorphisms. Also, two vector spaces having Hermitian forms are isomorphic if and only if they have the same signature and dimension.

8.3 Spaces of Polygons and Signature Calculation

In this section, we introduce the spaces of polygons that we consider. Each of these spaces is a complex vector space and admits a natural *area Hermitian form* on it. We calculate the signature of the area Hermitian form for each of these spaces. Let

$$\kappa = (\kappa_1, \ldots, \kappa_n),\ n > 1,\ 0 < \kappa_i < 2\pi,\ 1 \le i \le n,$$

be an n-tuple of real numbers. We will sometimes call it as curvature data. Let

$$\begin{aligned}\mathfrak{P}(\kappa) =& \{\hat{z} = (z_1, \ldots, z_{2n}) \in \mathbb{C}^{2n} :\\ & z_1 = 0, e^{\imath\kappa_i}(z_{2i-1} - z_{2i}) = z_{2i+1} - z_{2i}, 1 \le i \le n\}.\end{aligned}$$

$\mathfrak{P}(\kappa)$ can be thought as the set of oriented polygons

$$z_1 \to z_2 \to z_3 \ldots z_{2n} \to z_1.$$

Note that each element in $\mathfrak{P}(\kappa)$ has an outer angle κ_i at the vertex $2i$, z_{2i}, where the outer angle is the angle between the vectors $z_{2i-1} - z_{2i}$ and $z_{2i+1} - z_{2i}$ measured counter-clockwise. Also, for all $1 \le i \le n$ and for all $z \in \mathfrak{P}(\kappa)$, $|z_{2i-1} - z_{2i}| = |z_{2i+1} - z_{2i}|$. See Fig. 8.2.

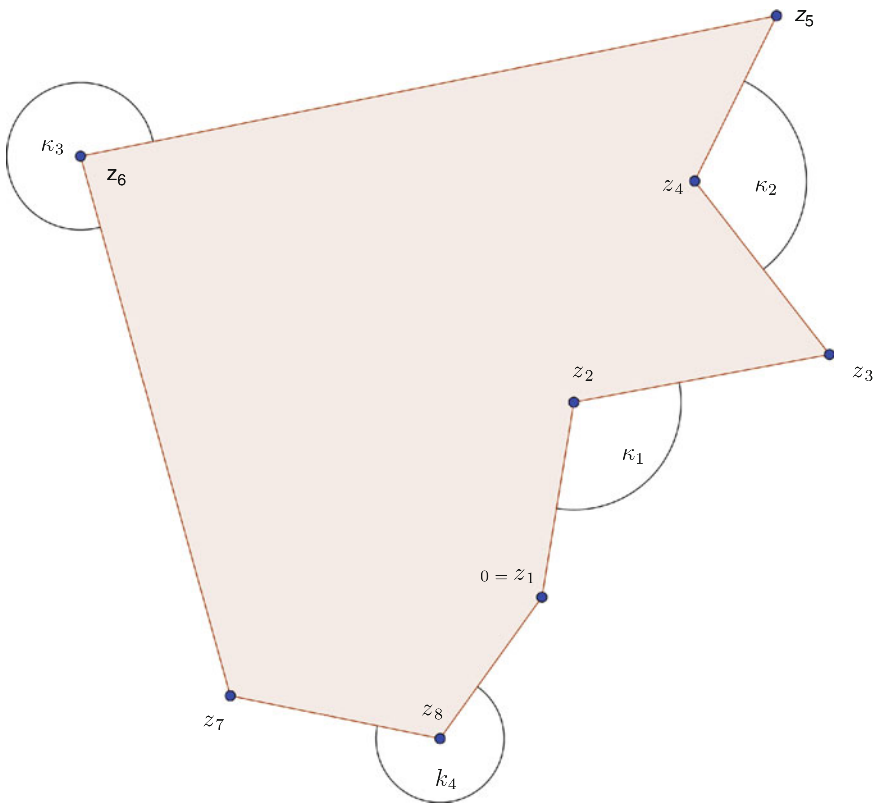


Fig. 8.2 An element of $\mathfrak{P}(\kappa_1, \kappa_2, \kappa_3, \kappa_4)$ as a polygon in the complex plane. Note that an edge of an element of $\mathfrak{P}(\kappa_1, \kappa_2, \kappa_3, \kappa_4)$ considered as a polygon may intersect another edge

Remark 8.1 Dimension of $\mathfrak{P}(\kappa)$ is $n - 1$ since each element $\hat{z} = (0, z_2, \ldots, z_{2n})$ is determined by its coordinates $z_3, \ldots, z_{2n-1}$.

8.3.1 The Area Hermitian Form and the Formula for Its Signature

In this section we introduce the Hermitian form that we are interested in. Also we give the formula for its signature. Consider the Hermitian form

$$h_A(\hat{z}, \hat{w}) = \frac{\sqrt{-1}}{4} \sum_{i=1}^{2n-1} (z_i \bar{w}_{i+1} - z_{i+1} \bar{w}_i)$$

on $\mathfrak{P}(\kappa)$, where $\hat{z} = (z_1, \ldots, z_{2n})$, $\hat{w} = (w_1, \ldots, w_{2n})$. We know that if $\hat{z}$ is a simple polygon, then the area of z is just the square-norm of $\hat{z}$, $h_A(\hat{z}, \hat{z})$. Therefore this form is called area Hermitian form.

We now state the formula that we prove. First, we introduce some notation.

Let

$$\epsilon(\kappa) := \begin{cases} 1 & \text{if } \sum_{i=1}^{n} \kappa_i \in 2\pi\mathbb{Z}, \\ 0 & \text{else.} \end{cases}$$

We denote the cardinality of a set A by $|A|$. Let

$$q(\kappa) = \left|\{i : 1 \leq i < n, \left\lfloor \sum_{k=1}^{i+1} \frac{\kappa_k}{2\pi} \right\rfloor = \left\lfloor \sum_{k=1}^{i} \frac{\kappa_k}{2\pi} \right\rfloor\}\right|, \tag{8.1}$$

and

$$p(\kappa) = n - 1 - q(\kappa) - \epsilon(\kappa), \tag{8.2}$$

where $\lfloor\rfloor$ is the floor function.

Lemma 8.1 *Let* $f(i) = \left\lfloor \sum_{k=1}^{i} \frac{\kappa_k}{2\pi} \right\rfloor$. *Then*

$$n - 1 - q(\kappa) = f(n).$$

In particular, $q(\kappa)$ *only depends on* $\sum_{k=1}^{n} \kappa_k$.

Proof Note that $f(i+1) - f(i) = 0$ or 1. The definition of $q(\kappa)$ implies that

$$n - 1 = q(\kappa) + \left|\{i : 1 \leq i < n, \left\lfloor \sum_{k=1}^{i+1} \frac{\kappa_k}{2\pi} \right\rfloor \neq \left\lfloor \sum_{k=1}^{i} \frac{\kappa_k}{2\pi} \right\rfloor\}\right|.$$

Therefore, it follows that

$$\begin{aligned} n - 1 - q(\kappa) &= \left|\{i : 1 \leq i < n, \left\lfloor \sum_{k=1}^{i+1} \frac{\kappa_k}{2\pi} \right\rfloor \neq \left\lfloor \sum_{k=1}^{i} \frac{\kappa_k}{2\pi} \right\rfloor\}\right| \\ &= \left|\{i : 1 \leq i < n, f(i+1) \neq f(i)\}\right| \\ &= \left|\{i : 1 \leq i < n, f(i+1) - f(i) = 1\}\right| \\ &= f(n) - f(n-1) + f(n-1) - f(n-2) + \ldots f(2) - f(1) = f(n), \end{aligned}$$

since $f(1)$ is equal to 0. The particular case is obvious. □

Lemma 8.2 *If σ is a permutation of $\{1, \ldots, n\}$ and $\kappa(\sigma) = (\kappa_{\sigma(1)}, \ldots, \kappa_{\sigma(n)})$, then*

$$q(\kappa) = q(\kappa(\sigma)), \text{ and } p(\kappa) = p(\kappa(\sigma)) \text{ and } \epsilon(\kappa) = \epsilon(\kappa(\sigma)).$$

Proof It is clear from the definition of ϵ that $\epsilon(\kappa) = \epsilon(\kappa(\sigma))$. Also Lemma 8.1 implies that $q(\kappa) = q(\kappa(\sigma))$. Since

$$p(\kappa) = n - 1 - q(\kappa) - \epsilon(\kappa)$$

it follows that $p(\kappa) = p(\sigma(\kappa))$ for any permutation σ. □

We will prove that the signature of the area Hermitian h_A is

$$(p(\kappa), q(\kappa)),$$

that is, we will prove that

1. $P(h_A) = p(\kappa) = \left\lfloor \sum_{k=1}^{n} \frac{\kappa_k}{2\pi} \right\rfloor - \epsilon(\kappa)$;
2. $N(h_A) = q(\kappa) = n - 1 - \left\lfloor \sum_{k=1}^{n} \frac{\kappa_k}{2\pi} \right\rfloor$;
3. $Z(h_A) = \epsilon(\kappa)$.

Note that this will imply that h_A is non-singular if and only if $\epsilon(\kappa) = 0$.

8.3.2 The Case n=2

In this section we consider the case where $\kappa = (\kappa_1, \kappa_2)$. It follows that $\mathfrak{P}(\kappa)$ has dimension 1. This means that the polygons that we consider have 4 edges and we can easily draw them. See Fig. 8.3.

Lemma 8.3 *If $n = 2$, then the signature of the area Hermitian form is $(p(\kappa), q(\kappa))$.*

Proof *There are 3 cases to consider.*

1. *$\kappa_1 + \kappa_2 > 2\pi$. From the left of Fig. 8.3, it is clear that the polygon corresponding to a non-zero element of $\mathfrak{P}(\kappa)$ has positive area. Since the area of the polygon corresponding to a non-zero element of $\mathfrak{P}(\kappa)$ is the square-norm of that element, we see that each non-zero element of $\mathfrak{P}(\kappa)$ has positive square-norm. Since $\mathfrak{P}(\kappa)$ is one-dimensional, it follows that the signature of the area Hermitian form is* $(1, 0)$. *On the other hand, it follows directly from the definition of $p(\kappa)$ and $q(\kappa)$ that $(p(\kappa), q(\kappa)) = (1, 0)$. The result follows.*
2. *$\kappa_1 + \kappa_2 = 2\pi$. In this case, every element in $\mathfrak{P}(\kappa)$ has area 0. See the middle of Fig. 8.3 . Therefore, the signature is* $(0, 0)$. *Also, it is clear that $p(\kappa) = q(\kappa) = 0$.*
3. *$0 < \kappa_1 + \kappa_2 < 2\pi$. In this case each polygon corresponding to a non-zero element is negatively oriented. This means that the orientation of these polygons is clockwise. Therefore, each non-zero element of this set has negative area. See*

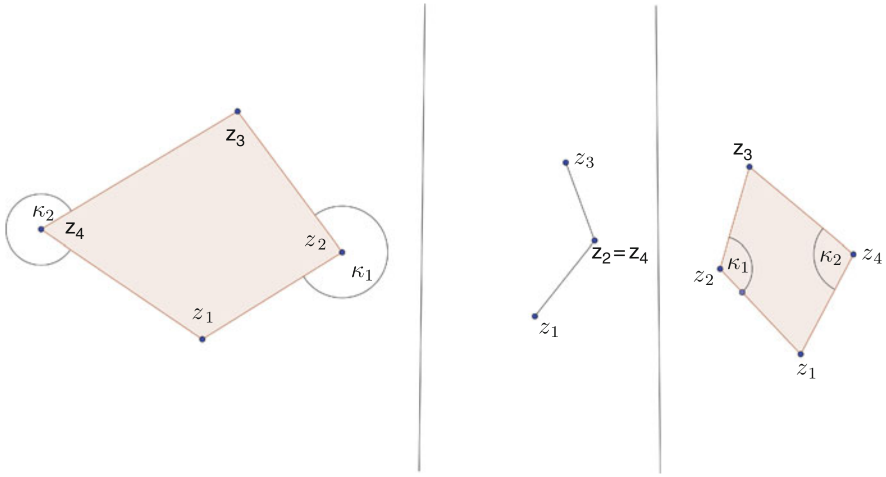


Fig. 8.3 In each part of the figure, a generic element in $\mathfrak{P}(\kappa_1, \kappa_2)$ is given. In the leftmost picture, we have $\kappa_1 + \kappa_2 > 2\pi$. In the picture in the middle, we have $\kappa_1 + \kappa_2 = 2\pi$. In the rightmost picture, we have $\kappa_1 + \kappa_2 < 2\pi$

the right hand side of Fig. 8.3. It follows that the signature of the form is (0, 1). *Also, it is clear from the definition of* $p(\kappa)$ *and* $q(\kappa)$ *that* $q(\kappa) = 1$ *and* $p(\kappa) = 0$. □

8.3.3 A Special Family of Polygons

In this section, we assume that $\kappa = \kappa(n) := (\pi, \pi, \ldots, \pi)$, where the curvature data $(\pi, \ldots, \pi)$ has length n.

Lemma 8.4 *Let* $\kappa = \kappa(n) = (\pi, \ldots, \pi)$.

1. *If* $n = 2k+1$, *then the signature of the area Hermitian form on* $\mathfrak{P}(\kappa(n))$ *is* (k, k).
2. *If* $n = 2k$, *then the signature of the area Hermitian form on* $\mathfrak{P}(\kappa(n))$ *is* $(k - 1, k - 1)$.

Proof The case $n = 2$ was proven in Lemma 8.3. We will prove the lemma by induction on the length of curvature data $(\pi, \ldots, \pi)$, n. Assume that $n \geq 3$. We start with a useful observation. Consider the map

$$\hat{z} = (0, z_2, z_3 \ldots, z_{2n-1}, z_{2n}) \mapsto \hat{z}' = (0, z_{2n}, z_{2n-1}, \ldots, z_3, z_2)$$

which sends $\mathfrak{P}(\kappa(n))$ to itself. It is clear that this map is a vector space isomorphism. It simply gives a polygon the opposite orientation. Note that the following formula holds:

$$h_A(\hat{z}', \hat{z}') = -h_A(\hat{z}, \hat{z}).$$

It follows that $P(h_A) = N(h_A)$. Consider the following vector subspace of $\mathfrak{P}(\kappa(n))$:

$$\mathfrak{P}' = \{\hat{z} \in \mathfrak{P}(\kappa(n)) : z_2 = 0\}.$$

Note that $\mathfrak{P}'$ is an $(n-2)$-dimensional vector subspace. Also, consider the restriction of the area Hermition form on $\mathfrak{P}'$. It is easy to see that $\mathfrak{P}'$ and $\mathfrak{P}(\kappa(n-1))$ are isomorphic as complex vector spaces with Hermitian forms.

Consider a basis $\{u_1, \ldots, u_r, u_{r+1}, \ldots, u_{r+s}, \ldots, u_{n-2}\}$ for $\mathfrak{P}'$ so that

1. $h_A(u_i, u_j) = 0$ if $i \neq j$;
2. $h_A(u_i, u_i) = 1$ if $1 \leq i \leq r$;
3. $h_A(u_i, u_i) = -1$ if $r + 1 \leq i \leq r + s$;
4. and $h_A(u_i, u_i) = 0$ if $i > r + s$.

Let

1. $\mathfrak{P}'^{+}$ be the vector subspace spanned by $\{u_1, \ldots, u_r\}$;
2. $\mathfrak{P}'^{-}$ be the vector subspace spanned by $\{u_{r+1}, \ldots, u_{r+s}\}$;
3. $\mathfrak{P}'^{\perp}$ be the vector subspace spanned by $\{u_{r+s+1}, \ldots, u_{n-2}\}$.

Assume that $n = 2k$, $k \geq 2$. Then $\mathfrak{P}'$ is $(2k-2)$-dimensional and the induction hypothesis implies that the signature of the area Hermitian form on $\mathfrak{P}'$ is $(k-1, k-1)$. Therefore

$$\mathfrak{P}' = \mathfrak{P}'^{+} \bigoplus \mathfrak{P}'^{-},$$

where $\mathfrak{P}'^{+}$ and $\mathfrak{P}'^{-}$ have dimension $k-1$. Consider an element $u \in \mathfrak{P}(\kappa(n)) \setminus \mathfrak{P}'$. Applying the Gram–Schmidt orthogonalization process if necessary, we can choose u so that it is orthogonal to $\mathfrak{P}'$. Therefore the signature of the area Hermitian form on $\mathfrak{P}(\kappa(n))$ is

1. $(k-1, k-1)$,
2. $(k, k-1)$ or
3. $(k-1, k)$.

Since $P(h_A) = N(h_A)$, it follows that this signature is $(k-1, k-1)$.

Assume that $n = 2k+1$. Then $\mathfrak{P}'$ is a $2k-1$ dimensional subspace of $\mathfrak{P}(\kappa(n))$. The induction hypothesis implies that the signature of the area Hermitian form on $\mathfrak{P}'$ is $(k-1, k-1)$. Therefore the dimension of $\mathfrak{P}'^{+}$ is $k-1$, the dimension of $\mathfrak{P}'^{-}$ is $k-1$ and the dimension of $\mathfrak{P}'^{\perp}$ is 1.

Take an element $v \in \mathfrak{P}(\kappa(n)) \setminus \mathfrak{P}'$ so that v is orthogonal to $\mathfrak{P}'^{+}$ and $\mathfrak{P}'^{-}$. We can assure this by the Gram–Schmidt orthogonalization process. Let W be the vector space spanned by v and $\mathfrak{P}'^{\perp}$. It is clear that

$$\mathfrak{P}(\kappa(n)) = \mathfrak{P}'^{+} \bigoplus \mathfrak{P}'^{-} \bigoplus W. \tag{8.3}$$

Assume that W is orthogonal to $\mathfrak{P}(\kappa(n))$, or equivalently, that the area Hermitian form restricted to W is zero. Let

$$w = (0, 0, 0, w_4, \dots, w_{2n})$$

be a generator of $\mathfrak{P}'^{\perp}$. There is an integer l such that $w_k = 0$ for $k < 2l$ and $w_{2l} \neq 0$. Note that $2w_{2l} = w_{2l+1}$. Consider the following element of $\mathfrak{P}(\kappa(n))$:

$$a = (0, \dots, 0, 1, 2, 1, 0, \dots, 0),$$

where the first 1 is in the coordinate $2l - 2$. Then

$$h_A(a, w) = \sqrt{-1}\bar{w}_{2l} \neq 0.$$

This is a contradiction. It follows that the area Hermitian form restricted to W is not trivial. Therefore the signature of the area Hermitian form restricted to W is

1. $(0, 1)$,
2. $(1, 0)$ or
3. $(1, 1)$.

Regarding the decomposition 8.3, it follows that the signature of h_A on $\mathfrak{P}(\kappa(n))$ is

1. $(k - 1, k)$,
2. $(k, k - 1)$ or
3. (k, k).

Since $P(h_A) = N(h_A)$, the signature is (k, k). □

The following corollary is an immediate application of Lemma 8.4.

Corollary 8.1 *If* $\kappa = (\pi, \dots, \pi)$*, then the signature of the area Hermitian form is* $(p(\kappa), q(\kappa))$.

Proof It is not difficult to see that

- $(p(\kappa), q(\kappa)) = (k, k)$ if $n = 2k$, and
- $(p(\kappa), q(\kappa)) = (k - 1, k - 1)$ if $n = 2k$.

Therefore the statement follows from Lemma 8.4. □

8.3.4 Cutting-Gluing Operations

In this section, we explain why for any permutation $\sigma \in S_n$, $\mathfrak{P}(\kappa)$ and $\mathfrak{P}(\kappa(\sigma))$ are isomorphic as complex vector spaces equipped with Hermitian forms.

We prove the claim by using some cutting and gluing operations. Before proceeding to the general case, we first introduce cutting and gluing operations on $\mathfrak{P}(\kappa)$, where $\kappa = (\pi, \pi, \pi)$. We know that $\mathfrak{P}(\kappa)$ is 2-dimensional and that the signature of h_A on it is $(1, 1)$. See Lemma 8.4. Take an element $\hat{z} = (z_1, z_2, z_3, z_4, z_5, z_6) \in \mathfrak{P}(\kappa)$. Then

$$\hat{z} = (z_1, z_2, z_3, z_4, z_5, z_6) = (0, \frac{z_3}{2}, z_3, \frac{z_3 + z_5}{2}, z_5, \frac{z_5}{2})$$

and

$$h_A(\hat{z}, \hat{w}) = \frac{\sqrt{-1}}{4}(z_3\bar{w}_5 - z_5\bar{w}_3).$$

By abusing notation, we denote an element $\hat{z} \in \mathfrak{P}(\kappa)$ as $\hat{z} = [[z_3, z_5]]$. Now consider a positively oriented element $\hat{z} \in \mathfrak{P}(\kappa)$. This element has positive square-norm, that is,

$$h_A(\hat{z}, \hat{z}) = \frac{\sqrt{-1}}{4}(z_3\bar{z}_5 - z_5\bar{z}_3) > 0.$$

Recall that $h_A(\hat{z}, \hat{z})$ is the area of the corresponding polygon which actually is a triangle.

In Fig. 8.4, consider the line segment $[z_2, z_5]$ and cut the triangle $[0, z_3, z_5]$ through this line segment to get two triangles. Glue the edges $[z_2, z_3]$ and $[0, z_2]$ by a rotation of angle π around z_2. In this way, we get another element in $\mathfrak{P}(\kappa)$ having the same area with coordinates $[[z_3 - z_5, z_5]]$. Therefore we have a map

$$\mathfrak{P}(\kappa) \to \mathfrak{P}(\kappa)$$

sending $[[z_3, z_5]] \mapsto [[z_3 - z_5, z_5]]$. Clearly this map is a vector space isomorphism and we realized that it respects the area Hermitian form.

In general, even if the entries of the curvature data κ are not equal, we can use these cutting-gluing operations to find isomorphisms between $\mathfrak{P}(\kappa)$ and $\mathfrak{P}(\kappa(\sigma))$. Note that it is enough to consider the cases for which $\sigma = (i, i+1)$ is a transposition to prove that $\mathfrak{P}(\kappa)$ and $\mathfrak{P}(\kappa(\sigma))$ are isomorphic. Take an element $\hat{z} \in \mathfrak{P}(\kappa)$, and consider it as a polygon in the complex plane. Assume that the line segment joining z_{2i+3} and z_{2i} does not intersect the polygon except at its endpoints. Cut the polygon through the line segment joining z_{2i+3} and z_{2i}. Glue the edge $[z_{2i}, z_{2i+1}]$ in the resulting quadrangle with the edge $[z_{2i}, z_{2i-1}]$ of the polygon to get the element $\hat{z}'$. In this way, we get an area-preserving map from a subset of $\mathfrak{P}(\kappa)$ to $\mathfrak{P}(\kappa(\sigma))$,

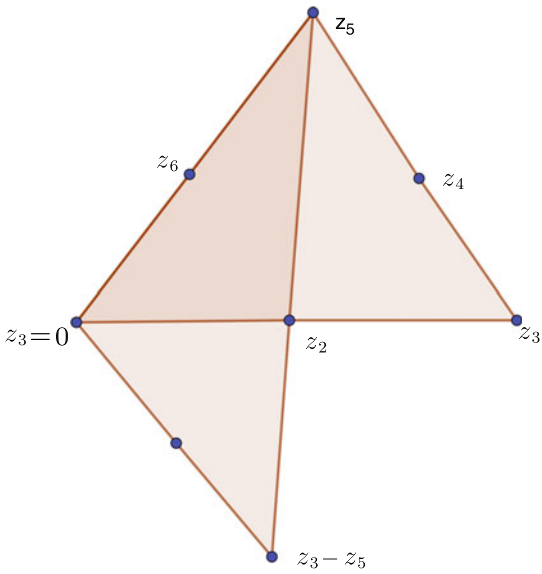


Fig. 8.4 Cutiing-gluing operation on $\mathfrak{P}(\pi, \pi, \pi)$

where $\sigma = (i, i+1)$. Note that this map extends to an area-preserving linear map between $\mathfrak{P}(\kappa)$ and $\mathfrak{P}(\kappa(\sigma))$. Indeed, this linear map is an isomorphism; one can reverse the cutting and gluing operations to get an inverse for the map. See Fig. 8.5.

Therefore, we have proved the following lemma.

Lemma 8.5

1. *For any* $\sigma \in S_n$, $\mathfrak{P}(\kappa)$ *and* $\mathfrak{P}(\kappa(\sigma))$ *are isomorphic as complex vector spaces equipped with Hermitian forms.*
2. *The signature of the area Hermitian form on* $\mathfrak{P}(\kappa)$ *is equal to the signature of the area Hermitian form on* $\mathfrak{P}(\kappa(\sigma))$.

Note that these cutting-gluing operations were introduced in [11].

8.3.5 Signature Calculation

In this section we prove the signature formula. Let $\kappa = (\kappa_1, \ldots, \kappa_n)$. Assume that $n > 2$ and $\kappa_1 + \kappa_2 < 2\pi$. Let

$$\kappa_{12} = \begin{cases} \kappa_1 + \kappa_2 & \text{if } \kappa_1 + \kappa_2 < 2\pi \\ \kappa_1 + \kappa_2 - 2\pi & \text{if } \kappa_1 + \kappa_2 > 2\pi \end{cases}$$

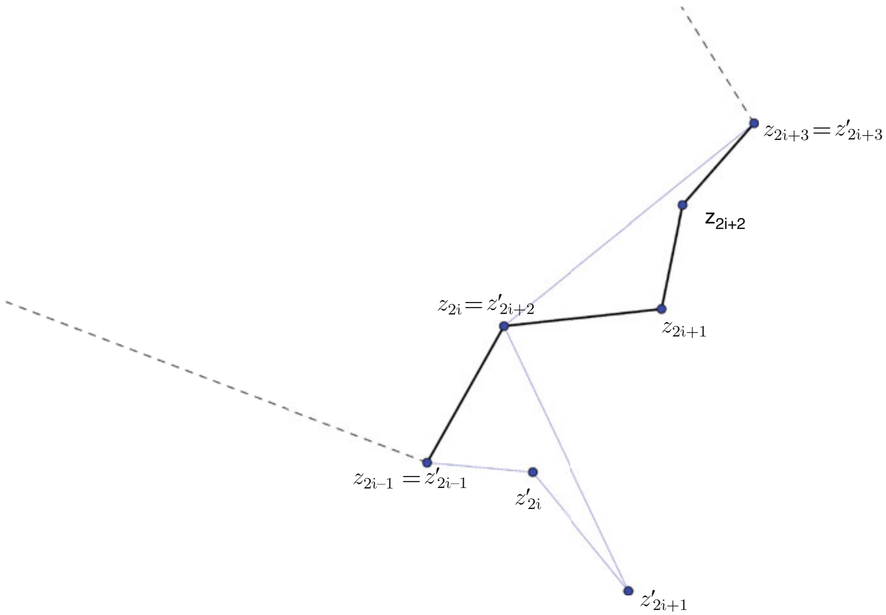


Fig. 8.5 We obtain an element $\mathfrak{P}(\kappa(\sigma))$ from an element of $\mathfrak{P}(\kappa)$

Also let $\kappa' = (\kappa_{12}, \kappa_3, \ldots, \kappa_n)$. Consider the following $n-2$ dimensional subspace of $\mathfrak{P}(\kappa)$:

$$\overline{\mathfrak{P}}(\kappa) = \{\hat{z} \in \mathfrak{P}(\kappa) : z_2 = z_4\}.$$

A generic element of $\overline{\mathfrak{P}}(\kappa)$ is shown in Fig. 8.6. It is not difficult to see that $\overline{\mathfrak{P}}(\kappa)$ together with the induced Hermitian form and $\mathfrak{P}(\kappa')$ are isomorphic. We want to find an orthogonal complement of $\overline{\mathfrak{P}}(\kappa)$ in $\mathfrak{P}(\kappa)$. In $\mathfrak{P}(\kappa)$, there is a unique element of the form

$$X = (0, -1, -1 + e^{\imath\kappa_1}, x, 0, \ldots, 0).$$

Since the angle at the fourth vertex, x, is $\kappa_2 \neq 2\pi - \kappa_1$, it follows that $X \notin \bar{\mathfrak{P}}(\kappa)$ and

$$(-1 + e^{\imath\kappa_1} - x)e^{\imath\kappa_2} = -x,$$

$$e^{\imath(\kappa_1+\kappa_2)} - e^{\imath\kappa_2} = x(e^{\imath\kappa_2} - 1),$$

$$1 - e^{-\imath\kappa_1} + x(-e^{-\imath\kappa_1} + e^{-\imath(\kappa_1+\kappa_2)}) = 0.$$

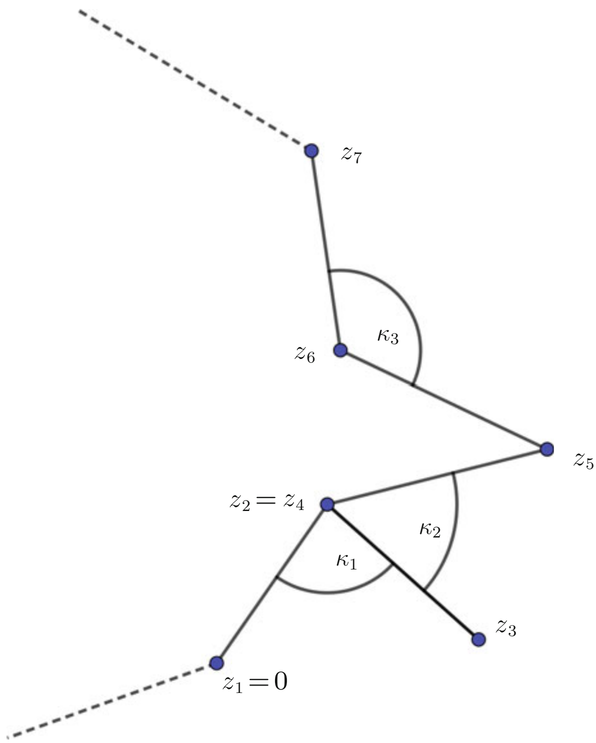


Fig. 8.6 A generic element of $\overline{\mathfrak{P}}(\kappa)$ when $\kappa_1 + \kappa_2 < 2\pi$

On the other hand, any element in $\overline{\mathfrak{P}}(\kappa)$ is a constant multiple of an element of the form

$$Y = (0, -1, -1 + e^{\imath\kappa_1}, -1, -1 + e^{\imath(\kappa_1+\kappa_2)}, \dots).$$

It follows that

$$h_A(X, Y) = 1 - e^{-\imath\kappa_1} + x(-e^{-\imath\kappa_1} + e^{-\imath(\kappa_1+\kappa_2)}) = 0.$$

Let $\mathbb{C}X$ denote the vector space generated by X. Clearly $\mathbb{C}X \equiv \mathfrak{P}(\kappa_1, \kappa_2)$. Note that we have proved the following lemma.

Lemma 8.6 *If $n > 2$ and $\kappa_1 + \kappa_2 \neq 2\pi$, then $\mathbb{C}X \oplus \overline{\mathfrak{P}}(\kappa) = \mathfrak{P}(\kappa)$.*

Theorem 8.3 *The signature of the area Hermitian form for $\mathfrak{P}(\kappa)$ is $(p(\kappa), q(\kappa))$. Also we have the following formulas for $p(\kappa)$ and $q(\kappa)$:*

$$q(\kappa) = n - 1 - \left\lfloor \sum_{k=1}^{n} \frac{\kappa_k}{2\pi} \right\rfloor$$

and

$$p(\kappa) = \left\lfloor \sum_{k=1}^{n} \frac{\kappa_k}{2\pi} \right\rfloor - \epsilon(\kappa),$$

where

$$\epsilon(\kappa) := \begin{cases} 1 & \text{if } \sum_{i=1}^{n} \kappa_i \in 2\pi\mathbb{Z}, \\ 0 & \text{else.} \end{cases}$$

Proof We prove the first part of the theorem by induction on n. If $n = 2$ we know that the statement is true. See Lemma 8.3. Assume that $n > 2$. If all of the κ_i are π, then we know the theorem is true. See Corollary 8.1. Assume that not all of the κ_i are π. It follows that there are $i, j \in \{1, 2, \ldots n\}$, $i \neq j$, such that $\kappa_i + \kappa_j \neq 2\pi$. By Lemma 8.2 and Lemma 8.5, we may assume that $\kappa_1 + \kappa_2 \neq 2\pi$.

Assume that $\kappa_1 + \kappa_2 < 2\pi$. Note that Lemma 8.6 and the induction hypothesis on the vector spaces $\mathbb{C}X \equiv \mathfrak{P}(\kappa_1, \kappa_2)$, $\overline{\mathfrak{P}}(\kappa) \equiv \mathfrak{P}(\kappa_1 + \kappa_2, \kappa_3, \ldots \kappa_n)$ imply that

$$\begin{aligned} N(h_A) &= q(\kappa_1, \kappa_2) + q(\kappa_1 + \kappa_2, \kappa_3, \ldots, \kappa_n) \\ &= \left|\{i : 1 \leq i < 2, \left\lfloor \sum_{k=1}^{i+1} \frac{\kappa_k}{2\pi} \right\rfloor = \left\lfloor \sum_{k=1}^{i} \frac{\kappa_k}{2\pi} \right\rfloor\}\right| \\ &+ \left|\{i : 2 \leq i < n, \left\lfloor \sum_{k=1}^{i+1} \frac{\kappa_k}{2\pi} \right\rfloor = \left\lfloor \sum_{k=1}^{i} \frac{\kappa_k}{2\pi} \right\rfloor\}\right| \\ &= \left|\{i : 1 \leq i < n, \left\lfloor \sum_{k=1}^{i+1} \frac{\kappa_k}{2\pi} \right\rfloor = \left\lfloor \sum_{k=1}^{i} \frac{\kappa_k}{2\pi} \right\rfloor\}\right| \\ &= q(\kappa). \end{aligned}$$

And we also have

$$\begin{aligned} P(\kappa) &= p(\kappa_1, \kappa_2) + p(\kappa_1 + \kappa_2, \kappa_3, \ldots \kappa_n) \\ &= 1 - q(\kappa_1, \kappa_2) - \epsilon(\kappa_1, \kappa_2) + n - 2 \\ &- q(\kappa_1 + \kappa_2, \kappa_3, \ldots, \kappa_n) - \epsilon(\kappa_1 + \kappa_2, \kappa_3, \ldots, \kappa_n) \\ &= n - 1 - q(\kappa) - \epsilon(\kappa_1 + \kappa_2, \kappa_3, \ldots, \kappa_n) \\ &= n - 1 - q(\kappa) - \epsilon(\kappa) \\ &= p(\kappa). \end{aligned}$$

Therefore $(P(h_A), N(h_A)) = (p(\kappa), q(\kappa))$. Note that a similar calculation holds for the case $\kappa_1 + \kappa_2 > 2\pi$.

The formulae for $p(\kappa)$ and $q(\kappa)$ follow easily from Lemma 8.1. □

Corollary 8.2 *Let* $\kappa = (\kappa_1, \ldots, \kappa_n)$.

1. *If* $2\pi < \sum_{i=1}^n \kappa_i < 4\pi$, *then the signature is* $(1, n-2)$.
2. *If* $\sum_{i=1}^n \kappa_i = 2\pi$, *then the signature is* $(0, n-2)$.
3. *If* $\sum_{i=1}^n \kappa_i = 2\pi(n-1)$, *then the signature is* $(n-2, 0)$.
4. *If* $\sum_{i=1}^n \kappa_i < 2\pi$, *then the signature is* $(0, n-1)$.
5. *If* $2\pi(n-1) < \sum_{i=1}^n \kappa_i < 2\pi n$, *then the signature is* $(n-1, 0)$.
6. *If* $2\pi(n-2) < \sum_{i=1}^n \kappa_i < 2\pi(n-1)$, *then the signature is* $(n-2, 1)$.

Acknowledgements The author is grateful to Susumu Tanabe, Muhammed Uludağ, Athanase Papadopoulos, Ke'nishi Ohshika and Haruko Nishi for their suggestions and help. He is financially supproted by TÜBİTAK.

References

1. A. Alexandrov, *Convex Polyhedra*. Springer Monographs in Mathematics (Springer, Berlin, 2005)
2. C. Bavard, É. Ghys, Polygones du plan et polyedres hyperboliques. Geom. Dedicata **43**(2), 207–224 (1992)
3. F. Fillastre, From spaces of polygons to spaces of polyhedra following Bavard, Ghys and Thurston. Enseign. Math. **57**(2), 23–56 (2011)
4. J. Fisher, D. Ruoff, J. Shilleto, Perpendicular polygons. Am. Math. Month. **92**(1), 23–37 (1985)
5. H. Groemer, *Geometric Applications of Fourier Series and Spherical Harmonics*. (Encyclopedia of Mathematics and its Applications) (Cambridge University Press, Cambridge, 1996)
6. I. Izmestiev, A simple proof of an isoperimetric inequality for Euclidean and hyperbolic cone-surfaces. Differ. Geom. Appl. **43**, 95–101 (2015)
7. S. Kojima, Y. Yamashita, Shapes of stars. Proc. Am. Math. Soc. **117**(3), 845–851 (1993)
8. H. Nishi, On the signature of area form on the polygon space. Josai Math. Monographs **6**, 147–150 (2013)
9. H. Nishi, K. Ohshika, A parametrisation of Teichmüller spaces by polygons, in *Proceedings of NACA* (2011)
10. H. Nishi, K. Ohshika, A pseudo-metric on moduli space of hyperelliptic curves. Josai Math. Monographs **5**, 51–59 (2012)
11. İ. Sağlam, Triangulations and quadrangulations of the sphere. Int. J. Math. **26**(8), 1550063 (2015)
12. İ. Sağlam, Complete flat cone metrics on punctured surfaces. Turkish J. Math. **43**, 813–832 (2019)
13. İ. Sağlam, Ideal triangulation and disc unfolding of a singular flat surface. Turkish J. Math. **45**, 1635–1648 (2021)
14. G. Tahar, Geometric triangulations and flips. Comptes Rendus Math. **357**(7), 620–623 (2019)
15. W. Thurston, Shapes of polyhedra and triangulations of the sphere. Geom. Topol. Monographs **1**, 511–549 (1998)
16. M. Troyanov, Les surfaces euclidiennes à singularités coniques. L'Enseignement Math. **32**, 79–94 (1986)

17. M. Troyanov, Prescribing curvature on compact surfaces with conical singularities. Trans. Am. Math. Soc. **324**(2), 793–821 (1991)
18. M. Troyanov, On the moduli space of singular Euclidean surfaces, in *Handbook of Teichmüller Theory*, vol. I (EMS Publishing House, Zürich, 2007), pp. 507–540

Chapter 9
Equilateral Convex Triangulations of $\mathbb{R}P^2$ with Three Conical Points of Equal Defect

Mikhail Chernavskikh, Altan Erdnigor, Nikita Kalinin, and Alexandr Zakharov

Abstract Consider triangulations of $\mathbb{R}P^2$ whose all vertices have valency six except three vertices of valency 4. In this chapter we prove that the number $f(n)$ of such triangulations with no more than n triangles grows as $C \cdot n^2 + O(n^{3/2})$ where $C = \frac{1}{20}\sqrt{3} \cdot л(\frac{\pi}{3})\zeta^{-1}(4)\zeta(Eis, 2) \approx 0.2087432125056015\ldots$, where л is the Lobachevsky function and $\zeta(Eis, 2) = \sum\limits_{(a,b)\in\mathbb{Z}^2\setminus 0} \frac{1}{|a+b\omega^2|^4}$, and $\omega^6 = 1$.

Keywords Flat metric · Equilateral triangulation · Conical singularity · Zeta function · Epstein zeta function · Hyperbolic volume

AMS Codes 51M09, 57N45, 11P21, 11M36, 11E45

9.1 Introduction

Consider a triangulation T of $\mathbb{R}P^2$ such that each vertex of T is contained in at most six triangles. These triangulations are called *convex*. Let each triangle in T be the equilateral triangle with sides of length one. This supplies $\mathbb{R}P^2$ with a flat metric outside of the vertices of T. If at a vertex v of T exactly k triangles come together then we say that the defect at v is equal to $(6-k)\pi/3$. Convex triangulations are exactly those with non-negative defects. By counting edges, vertices, and triangles in T one can see that the sum of all defects of the vertices of T is equal to 2π, because the Euler characteristic of $\mathbb{R}P^2$ is one. Hence this construction gives a flat metric on $\mathbb{R}P^2$ except at most six points (vertices of valency less than six).

M. Chernavskikh
Lomonosov Moscow State University, Moscow, Russia

A. Erdnigor
Russian Federation, Department of Mathematics, HSE University, Moscow, Russia

N. Kalinin (✉) · A. Zakharov
Saint Petersburg State University, Saint Petersburg, Russia

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_9

Consider the covering of $\mathbb{R}P^2$ by S^2. Naturally we obtain a metric on S^2 which is flat everywhere except at most twelve points with so-called conical singularities. A conical singularity with defect θ, $0 \leq \theta < 2\pi$ is locally modeled on the sector $0 \leq \phi \leq 2\pi - \theta$ of the unit disk (r, ϕ) with identified boundaries $(r, 0) \sim (r, 2\pi - \theta)$.

By Alexandrov's theorem [1] each flat metric on S^2 with conical singularities can be realized as the surface of a (possible degenerate) convex polytope in $\mathbb{R}^3$ with intrinsic metric. If we have only two conical points (with defects θ and $2\pi - \theta$) on an everywhere else flat $\mathbb{R}P^2$, then its covering S^2 has four conical points, and they should be identified by the central symmetry. Thus this metric is realized as a two-sided planar parallelogram (a degenerate polytope) with angles $\theta/2, \pi - \theta/2$. If we consider $\mathbb{R}P^2$ with three conical points, then its covering S^2 is isometric to a centrally-symmetric octahedron.

Thurston [4] (see also the lecture notes [3] which contain more detailed proofs) studied convex triangulations of S^2 and the moduli space of flat metrics on S^2 with a finite number of arbitrary conical singularities; the set of convex equilateral triangulations lives as a discrete subset in this moduli space. Following Thurston's ideas, we study equilateral triangulations of $\mathbb{R}P^2$ whose vertices have all valency six except three vertices of valency four.

9.2 Triangulations of $\mathbb{R}P^2$ with Three Marked Points with Defects $2\pi/3$

A graph without loops and multiple edges, drawn on $\mathbb{R}P^2$, is called *a triangulation* of $\mathbb{R}P^2$ if each face of this graph has three edges. Note that two faces of such a triangulation can intersect in zero, one, two, or three vertices.

Consider a triangulation T of $\mathbb{R}P^2$ such that only three vertices A, B, C have valency four, and all the other vertices have valency six.

T gives a flat metric $\mu_{\mathbb{R}P^2}$ on $\mathbb{R}P^2$ except at A, B, C. Passing to the universal covering sphere S^2 one gets a flat metric μ_{S^2} on S^2 except six points. By Alexandrov's theorem, μ_{S^2} is realised as the intrinsic metric of the surface of a certain centrally symmetric octahedron F. The projections of the edges of F give six geodesic paths between A, B, C in $\mathbb{R}P^2$, thus cutting $\mathbb{R}P^2$ into four triangles (all with vertices A, B, C, so we have four triangles ABC). Choose one of these four triangles, call it Δ. Call A, B, C, Δ the *label* of T.

Denote by $T_{\mathbb{R}P^2}$ the set $\{T, A, B, C, \Delta\}$ of labelled triangulations of $\mathbb{R}P^2$. Two such triangulations $(T_1, A_1, B_1, C_1, \Delta_1)$, $(T_2, A_2, B_2, C_2, \Delta_2)$ are said *isometric* if there exists a map between triangulations T_1, T_2, which sends vertices and edges of T_1 to vertices and edges of T_2, A_1 to A_2, B_1 to B_2, C_1 to C_2 and Δ_1 to Δ_2.

Consider the smallest possible triangulation of $\mathbb{R}P^2$ which consists of three vertices, four triangles, and six edges. We can label it in $3 \cdot 2 \cdot 1 \cdot 4$ different ways, but all the obtained labelled triangulations are isometric.

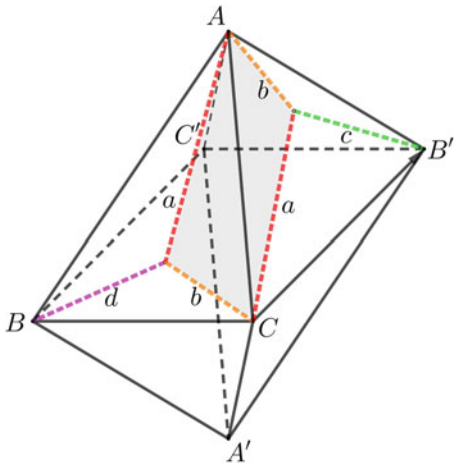


Fig. 9.1 An octahedron with vertices $AA'BB'CC'$, two Fermat–Torricelli points in the faces ABC, ACB', and the corresponding parallelogram with sides a, b

Let $f(n)$ be the cardinality of the set of isometry classes of labelled triangulations in $\mathbb{R}P^2$ with no more than n triangles.

Consider a labelled triangulation $\{T, A, B, C, \Delta\}$ of $\mathbb{R}P^2$. Consider the octahedron F as above. Then Δ lifts to F as two triangles Δ_1, Δ_2. Call A, B, C the vertices of Δ_1 and A', B', C' the vertices of Δ_2, then $A, A' \in F$ are projected to $A \in \mathbb{R}P^2$, $B, B' \in F$ are projected to $B \in \mathbb{R}P^2$, $C, C' \in F$ are projected to $C \in \mathbb{R}P^2$ under the covering map $F \to \mathbb{R}P^2$. Who is Δ_1 and who is Δ_2 is uniquely defined by the condition that the order of vertices A, B, C is counterclockwise (looking from outside of $F \subset \mathbb{R}^3$, see Fig. 9.1).

We can reverse the procedure. Consider a convex triangulations $\tilde{T}$ of S^2 with six points with defects $2\pi/3$. Mark three of these points as A, B, C and suppose that by supplying S^2 with a flat metric as above and realising it as the surface of a polyhedron we obtain a centrally symmetric octahedron F, and ABC is a face of F, and its orientation gives the counterclockwise order of ABC (Fig. 9.1). The central symmetry of F preserves $\tilde{T}$ and provides us with a projection $p : F \to \mathbb{R}P^2$. Projecting $\tilde{T}$ to a triangulation T of $\mathbb{R}P^2$ we mark the images of $A, B, C \in F$ as $A, B, C \in \mathbb{R}P^2$. Label by Δ the image of the face ABC of F under p.

Consider a centrally symmetric octahedron $F \subset \mathbb{R}^3$, such that the sum of angles at each vertex of F is $4\pi/3$. Suppose that $\tilde{T}$ is a convex equilateral triangulation of F. Choose any face of F and call its vertices A, B, C in such a way that the order of A, B, C is counterclockwise (if looking from outside of $F \subset \mathbb{R}^3$, see Fig. 9.1) and call the opposite faces A', B', C'. (A, B, C) is a label of $\tilde{T}$. We consider labeled triangulations $(\tilde{T}, A, B, C)$ up to isometry.

We proved the following lemma

Lemma 9.2.1 *There exist a bijection between labelled triangulations* (T, A, B, C, Δ) *of* $\mathbb{R}P^2$ *with* n *triangles and labelled triangulations* $(\tilde{T}, A, B, C)$ *with* $2n$ *triangles.*

Therefore $f(n) = \#\{(\tilde{T}, A, B, C)$ with no more than $2n$ triangles$\}$.

9.3 Moduli Space of Flat Metrics on S^2 with Six Pair-Wise Centrally Symmetric Conical Points of Equal Defect

Consider the set of all centrally symmetric octahedra F, such that the sum of angles at each vertex of F is $4\pi/3$. There exist natural coordinates on this space as follows [5].

Recall that for a triangle ABC whose angles are all less than $2\pi/3$ the Fermat–Torricelli point is the unique point X inside the triangle such that all the angles AXB, BXC, CXA are equal to $2\pi/3$. If the angle ABC is equal to $2\pi/3$ then we say that B is the Fermat–Torricelli point of the triangle ABC. The Fermat–Torricelli point X is the point minimizing $|XA| + |XB| + |XC|$.

Pick the Fermat–Torricelli point in each face of F and connect it with the vertices of this face. Then, among the lengths of these 24 intervals there are only four different ones [5], let us denote them by a, b, c, d.

Conversely, given four non-negative numbers a, b, c, d (we allow at most one of them to be zero, see the examples below), we can construct 12 parallelograms with acute angle $\pi/3$ and sides (a, b), (a, c), (a, d), (b, c), (b, d), (c, d) (two copies of each parallelogram). Let us bend each of them along its diagonal and glue them in an octahedron F. The diagonals of the parallelograms become edges of F. If $a = 0$ then the angle BAC is $2\pi/3$ and the parallelogram in Fig. 9.1 degenerates to the edge AC of F.

Let us say that the counterclockwise order around A of the intervals from A to the Torricelli points of the adjacent faces gives a, b, c, d, and let us also fix that the interval of length a belongs to the triangle ABC, see Fig. 9.1. Note also that there are six rotational orderings of a, b, c, d and that all of them are realized at exactly six vertices of F.

Given these coordinates on the moduli space of such octahedra (see details in [5] for octahedra with general defects at the vertices), we easily compute the area of the octahedron $F = (a, b, c, d)$ (note that $2\sin\frac{\pi}{3} = \sqrt{3}$): it is

$$Area(a, b, c, d) = \sqrt{3}(ab + ac + ad + bc + bd + cd).$$

Let $Q(a, b, c, d) = ab + ac + ad + bc + bd + cd$; Q is a quadratic form of signature $(1, 3)$ since

$$Q(a, b, c, d) = \frac{1}{8}(3(a+b+c+d)^2 - (c+d-a-b)^2 - 2(c-d)^2 - 2(a-b)^2).$$

If we start with an equilateral triangulation of $\mathbb{R}P^2$ with n triangles, then its covering sphere has $2n$ triangles, and each triangle has area $\sqrt{3}/4$, so the area of the sphere is

$$n\sqrt{3}/2 = \sqrt{3}Q(a, b, c, d))$$

which gives $n = 2Q(a, b, c, d)$.

Define

$$X = \left\{(a, b, c, d) \in \mathbb{R}^4_{\geq 0} \mid Q(a, b, c, d) \leq 1\right\}.$$

Lemma 9.3.1

$$\mathrm{Vol}(X) = \sqrt{3}\text{л}(\frac{\pi}{3}),$$

where $\text{л}(\phi) = -\int_0^\phi \ln|2\sin\theta|d\theta$ *is the Lobachevsky function.*

Proof Recall that each bilinear symmetric form $(\cdot, \cdot)$ yields a volume form on $\mathbb{R}^4$. Namely, $\mathrm{Vol}_{(\cdot,\cdot)}(v_1, v_2, v_3, v_4) = \pm\sqrt{|\det(v_i, v_j)|}$, the square root of the Gramian of $(\cdot, \cdot)$ with respect to this system of vectors. The sign of the (oriented) volume is defined by the orientation of (v_1, v_2, v_3, v_4).

Denote by $\bar{Q}$ the bilinear symmetric form associated with Q. Define a 3-form α on $\mathbb{R}^4$ as follows:

$$\forall x \in \mathbb{R}^4, \alpha : \bigwedge^3 T_x\mathbb{R}^4 \to \mathbb{R}, \alpha(v_1, v_2, v_3) = \mathrm{Vol}_{\bar{Q}}(x, v_1, v_2, v_3).$$

Note that Q induces a hyperbolic structure in the set $Q(v) = 1$ and that α is the corresponding volume form. Next (see [5] for details),

$$\int\limits_{v\in\mathbb{R}^4_{\geq 0}, Q(v)=1} \alpha = 3\text{л}(\frac{\pi}{3}).$$

Let dQ be the differential of Q, namely

$$dQ : T_x\mathbb{R}^4 \to \mathbb{R}, dQ(w) = 2\bar{Q}(x, w).$$

Let $v = (a, b, c, d)$, consider $Q'(v) = Q'(a, b, c, d) = a^2 + b^2 + c^2 + d^2$. Let ω be the standard Euclidian volume form $\omega(v_1, v_2, v_3, v_4) = \mathrm{Vol}_{\bar{Q}'}(v_1, v_2, v_3, v_4)$.

Let us prove that

$$Q^{-1}dQ \wedge \alpha = \frac{\sqrt{3}}{2}\omega. \tag{9.1}$$

Denote the coordinate basis in $\mathbb{R}^4$ by (e_1, e_2, e_3, e_4). Take any $x, v_1, v_2, v_3 \in \mathbb{R}^4$, and denote by $A \in \mathrm{Mat}_{4\times 4}(\mathbb{R})$ the matrix of their coordinates.

On $T_x\mathbb{R}^4$ we have

$$(Q^{-1}dQ \wedge \alpha)(x, v_1, v_2, v_3) = Q(v)^{-1}2\bar{Q}(x, x)\mathrm{Vol}_Q(x, v_1, v_2, v_3) =$$

$$= 2\det A\mathrm{Vol}_Q(e_1, e_2, e_3, e_4) = \frac{\sqrt{3}}{2}\omega(x, v_1, v_2, v_3)$$

which proves (9.1).

Now,

$$\mathrm{Vol}(X) = \int\limits_X \omega = \frac{2}{\sqrt{3}} \int\limits_X Q^{-1} dQ \wedge \alpha =$$

$$= \frac{2}{\sqrt{3}} \int\limits_0^1 q^{-1} dq \int\limits_{a,b,c,d \geq 0,\, Q(a,b,c,d)=q} \alpha = \frac{2}{\sqrt{3}} \int\limits_0^1 q^{-1} (dq)\, q^2 \int\limits_{a,b,c,d \geq 0,\, Q(a,b,c,d)=1} \alpha =$$

$$= \frac{2}{\sqrt{3}} \int\limits_0^1 q dq \cdot 3 л(\frac{\pi}{3}) = \sqrt{3} л(\frac{\pi}{3}).$$

□

Denote

$$g(n) = \# \left\{ (a, b, c, d) \in \mathbb{Z}^4_{>0} \mid Q(a, b, c, d) \leq n \right\}.$$

Theorem 9.3.1

$$g(n) = \sqrt{3} л(\frac{\pi}{3}) n^2 + O(n^{3/2}),$$

where

$$\sqrt{3} л(\frac{\pi}{3}) \approx 0.58597680967236472265039057221806926727385075240896\ldots$$

Proof Define

$$Y_t = \left\{ (a, b, c, d) \in \mathbb{R}^4_{\geq 0} \mid 1 \leq Q(a, b, c, d) \leq t \right\}.$$

Note that $g(n) = |Y_n \cap \mathbb{Z}^4_{>0}|$. It follows from Lemma 9.3.1 that

$$g(n) \approx \mathrm{Vol}(Y_n) \approx \sqrt{3} л(\frac{\pi}{3}) n^2.$$

Note that the error term is proportional to the Euclidean three-dimensional volume of the boundary of Y_n since the three-dimensional volume of the boundary of X is finite (one can use a similar reasoning as in Lemma 9.3.1).

For $t \geq 1$, denote the three-dimensional volume of the boundary of Y_t by $r(t)$. Denote by $2Y_t$ the image of Y_t under the homothety with center at 0 and coefficient 2. Then the three-dimensional volume of the boundary of $2Y_t$ is $8r(t)$. On the other hand $Y_{4t} = Y_4 \cup 2Y_t$ hence $r(4t) \le r(4) + 8r(t)$, thus

$$r(4t) + \frac{1}{7}r(4) \le 8\big[r(t) + \frac{1}{7}r(4)\big].$$

Letting $b(t) = r(t) + \frac{1}{7}r(4)$ we obtain $b(4t) \le 8b(t)$ and this leads to the estimate $b(4^k t) \le 8^k b(t)$. Let $n = 4^k x, 1 \le x < 4$. Note that $8^k \le n^{3/2}$. Then $b(n) \le 8^k b(x)$. Let $c = \max_{1 \le x \le 4} b(x)$. Thus we obtain $b(n) \le cn^{3/2}$. This can be rewritten as $r(n) + \frac{1}{7}r(4) \le cn^{3/2}$ and so the volume of the boundary of Y_n is $O(n^{3/2})$. □

Let us also introduce

$$h(n) = \#\left\{(a, b, c, d) \in \mathbb{Z}^4_{>0} \mid a \equiv b \equiv c \equiv d \pmod 3, Q(a, b, c, d) \le n\right\}.$$

The covolume (in $\mathbb{Z}^4$) of the lattice generated by such quadruples is 27, so, repeating the arguments of our proof of Theorem 9.3.1 we obtain

Theorem 9.3.2

$$h(n) = \frac{\sqrt{3}}{27} Л(\frac{\pi}{3})n^2 + O(n^{3/2}).$$

9.4 A Parametrization of Equilateral Triangulations of S^2 with Six Centrally-Symmetric Points with Defects $2\pi/3$

Let $\omega = e^{\frac{2\pi i}{6}} = \frac{1+\sqrt{-3}}{2}$. Consider the Eisenstein lattice

$$Eis = \mathbb{Z} \oplus \mathbb{Z}\omega^2 \subset \mathbb{C}.$$

Define $\widetilde{Eis} = \frac{1}{1-\omega^2} Eis$. Note that $\widetilde{Eis}$ contains Eis, and $\widetilde{Eis} \setminus Eis$ is the set $z + Eis$ where $z = \frac{1+\omega}{3} = \frac{1}{1-\omega^2}$ is the Torricelli point of the triangle with vertices 0, 1, ω.

Consider a labelled triangulation $(\tilde{T}, A, B, C)$ of a centrally symmetric octahedron with vertices A, A', B, B', C, C'. Take the faces ABC, ACB', $AB'C'$, $AC'B$, make a cut along AC', and develop the obtained polygon onto the plane such that A goes to $0 \in Eis$ under our developing map, and the vertices of T go to Eis. Then the developing map is defined up to the action of $\mathbb{Z}_6$ by rotations, because under the developing map we preserve the local orientation at A.

Let $\vec{a}, \vec{b}, \vec{c}, \vec{d}$ be the vectors in $\mathbb{C}$ connecting the point $0 \in Eis$ and the Torrichelli points of the four faces ABC, ACB', $AB'C'$, $AC'B$ of F (Fig. 9.2).

Lemma 9.4.1 *Under the developing map vectors* $\vec{a}, \vec{b}, \vec{c}, \vec{d}$ *go to* $\widetilde{Eis}$.

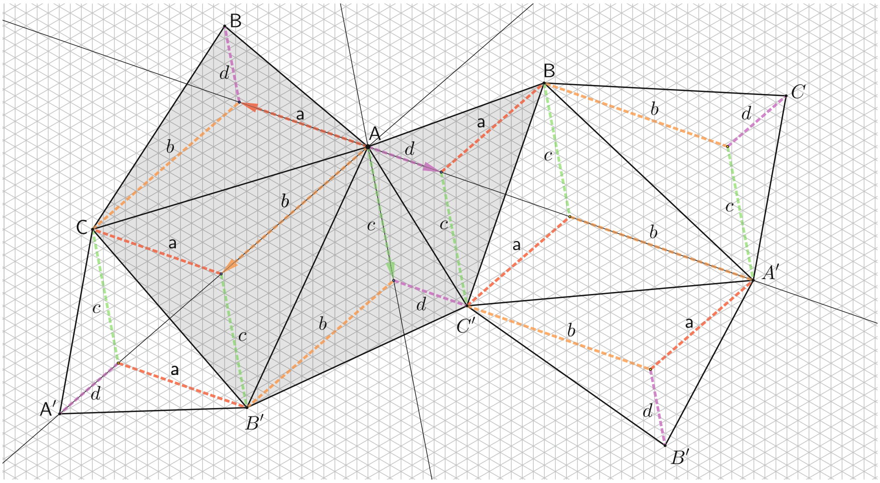


Fig. 9.2 A developing of the octahedron $AA'BB'CC'$ on $\mathbb{R}^2$ is presented, $A = 0$, vertices A, B, C, A', B', C' go to the lattice *Eis*. Note that a triangulation of $\mathbb{R}P^2$ can be obtained from the grey area by gluing AB to AB, then BC to $B'C'$ and then BC' to $B'C$. Note that the Torricelli centers of the faces do not belong to *Eis* but belong to $\widetilde{Eis}$, e.g. see the Torricelli center of ABC

Proof All the vertices of the octahedron are developed into the lattice points. Then the sums $\vec{a}+\vec{b}, \vec{b}+\vec{c}, \vec{c}+\vec{d}, \vec{a}+\vec{d}\omega^2$ also belong to *Eis*. Then, $\vec{a}+\vec{b}-(\vec{b}+\vec{c})+\vec{c}+\vec{d}-(\vec{a}+\vec{d}\omega^2) = \vec{d}(1-\omega^2) \in Eis$, hence $\vec{d} \in \frac{1}{1-\omega^2}Eis = \widetilde{Eis}$. Then, $\vec{c}+\vec{d} \in Eis$ and the latter is a sublattice in $\widetilde{Eis}$, therefore $\vec{c} \in \widetilde{Eis}$. Similarly, $\vec{b}, \vec{a} \in \widetilde{Eis}$. □

Definition 9.4.1 The vectors

$$\vec{e}_x = 1/(1-\omega^2), \vec{e}_y = \omega^2/(1-\omega^2)$$

form a basis in the lattice $\widetilde{Eis}$. Each vector in $\widetilde{Eis}$ can be expressed as $x\vec{e}_x + y\vec{e}_y$, $(x, y) \in \mathbb{Z}^2$. There are three cases for the sum $(x+y) \mod 3$. The lattice $\widetilde{Eis}$ is divided into three subsets:

$$\widetilde{Eis}_k = \{x\vec{e}_x + y\vec{e}_y \; in \; \widetilde{Eis} | (x+y) \equiv k \mod 3\}.$$

Note that $\vec{e}_x - \vec{e}_y = 1/(1-\omega^2) - \omega^2/(1-\omega^2) = 1 \in Eis$ and

$$2\vec{e}_x + \vec{e}_y = 2/(1-\omega^2) + \omega^2/(1-\omega^2) = \frac{2+\omega^2}{1-\omega^2} = \frac{2+\omega^2}{1-\omega^2} = \omega \in Eis.$$

This implies that $\widetilde{Eis}_0 = Eis$.

However the vectors $\vec{a}, \vec{b}, \vec{c}, \vec{d}$ are not arbitrary.

Lemma 9.4.2 *If the vectors* $\vec{a}, \vec{c}$ *lie in* $\widetilde{Eis}_k$ *then* $\vec{b}, \vec{d}$ *lie in* $\widetilde{Eis}_{-k}$. *In other words, there are three cases:*

(1) $\vec{a}, \vec{b}, \vec{c}, \vec{d} \in \widetilde{Eis}_0$;
(2) $\vec{a}, \vec{c} \in \widetilde{Eis}_1$ *and* $\vec{b}, \vec{d} \in \widetilde{Eis}_2$;
(3) $\vec{a}, \vec{c} \in \widetilde{Eis}_2$ *and* $\vec{b}, \vec{d} \in \widetilde{Eis}_1$.

Proof This follows from $\widetilde{Eis}_k + \widetilde{Eis}_m = \widetilde{Eis}_{k+m}$ and fact that $\omega^2 \widetilde{Eis}_k = \widetilde{Eis}_k$. □

Thus we constructed a bijection between the labelled triangulations $(\tilde{T}, A, B, C)$ up to isometry and certain 4-tuples of vectors $\vec{a}, \vec{b}, \vec{c}, \vec{d} \in \widetilde{Eis}$ up to a $\mathbb{Z}_6$ action.

One could consider the sublattice $Eis_0 \subset Eis$,

$$Eis_0 = \left\{ x + y\omega^2 \mid x, y \in \mathbb{Z}, x + y \equiv 0 \pmod 3 \right\}.$$

The cosets of Eis_0 in Eis are Eis_0, Eis_1, Eis_2 where

$$Eis_k = \left\{ x + y\omega^2 \mid x, y \in \mathbb{Z}, x + y \equiv k \pmod 3 \right\}.$$

If L is a lattice, let $\mathrm{Prim}L = \{v \in L \setminus 0 \mid \nexists w \in L, n > 1 : nw = v\}$.

Let $E_0 = \mathrm{Prim}Eis \cap Eis_0$ and $E_{\neq 0} = \mathrm{Prim}Eis \cap (Eis_1 \sqcup Eis_2)$, then

$$\mathrm{Prim}Eis_0 = E_0 \sqcup 3E_{\neq 0}. \tag{9.2}$$

Indeed, it follows from $\mathrm{Prim}Eis = E_0 \sqcup E_{\neq 0}$ that each primitive vector v of Eis_0 is either a primitive vector in Eis (and then it is an element of E_0) or there exists $v' \in Eis, v = kv', k > 1$ and $v' \notin Eis_0, v' \in E_{\neq 0}$. In the latter case, $3v' \in Eis_0$ (this is true for each vector in Eis), therefore k can be equal to three only. Therefore $v \in 3E_{\neq 0}$.

Theorem 9.4.1

$$f(n) = \frac{1}{6}\#\{(z \in \mathrm{Prim}Eis, (a, b, c, d) \in \mathbb{Z}^4_{>0}) | \frac{2}{3}|z|^2 Q(a, b, c, d) \le n\},$$

where (i) $z \in E_0$ *and* a, b, c, d *are arbitrary or (ii)* $z \in E_{\neq 0}, a \equiv b \equiv c \equiv d$ (mod 3).

Proof Each labelled triangulation $(\tilde{T}, A, B, C)$ is determined by the vectors

$$\vec{a}, \vec{b}, \vec{c}, \vec{d} \in \frac{1}{1 - \omega^2} Eis = \widetilde{Eis}$$

with the oriented angles $\angle(\vec{a}, \vec{b}) = \angle(\vec{b}, \vec{c}) = \angle(\vec{c}, \vec{d}) = \frac{\pi}{3}$. One could find $z' \in \mathrm{Prim}\widetilde{Eis}$ — the primitive vector proportional to $\vec{a}$. Then

$$(\vec{a}, \vec{b}, \vec{c}, \vec{d}) = z' \cdot (a, b\omega, c\omega^2, d\omega^3), a, b, c, d \in \mathbb{Z}_{>0}.$$

Let $z = (1 - \omega^2)z' \in \mathrm{Prim}Eis$. The number of triangles in $\tilde{T}$ is equal to the total area of the octahedron divided by the area of one equilateral triangle. The area equals $\sin\frac{\pi}{3} \cdot 2|z'|^2 Q(a, b, c, d) = \frac{1}{\sqrt{3}}|z|^2 Q(a, b, c, d)$ whereas the area of one equilateral triangle is $\frac{\sqrt{3}}{4}$. So the total number of triangles is equal to $\frac{4}{3}|z|^2 Q(a, b, c, d)$. Recall that $f(n)$ is the number of labelled triangulations $(\tilde{T}, A, B, C)$ with at most $2n$ triangles. The last condition is equivalent to $\frac{2}{3}|z|^2 Q(a, b, c, d) \le n$.

Let us study the conditions $\vec{a} + \vec{b}, \vec{b} + \vec{c}, \vec{c} + \vec{d} \in Eis$. In the case $z' \in Eis \iff z \in Eis_0$ the condition is satisfied automatically. Otherwise, $z' \notin Eis \iff z \in Eis_1 \sqcup Eis_2$ the condition on their sums $\vec{a} + \vec{b}$, etc., belonging to Eis is equivalent to $a \equiv b \equiv c \equiv d \pmod 3$ by a direct computation.

Finally, we notice that the triangulations with z and ωz determine isometric triangulations. This adds the factor $\frac{1}{6}$. □

Given a lattice $L \subset \mathbb{R}^2$ and $\mathrm{Re}(s) > 1$ we define the Epstein zeta function

$$\zeta(L, s) = \sum_{\gamma \in L \setminus 0} < \gamma, \gamma >^s .$$

One can prove that

$$\zeta(Eis, s) = \sum_{z \in Eis \setminus 0} |z|^{-2s} = 6\zeta_{\mathbb{Q}[\sqrt{-3}]}(s) = 6\zeta(s)L(\chi_{-3}, s).$$

We refer to [2] for details.

Now we are ready to estimate $f(n)$.

Theorem 9.4.2

$$f(n) = \frac{1}{20}\sqrt{3} \cdot л(\frac{\pi}{3})\zeta^{-1}(4)\zeta(Eis, 2)n^2 + O(n^{3/2}),$$

as $n \to \infty$ *where*

$$\frac{1}{20}\sqrt{3} \cdot л(\frac{\pi}{3})\zeta^{-1}(4)\zeta(Eis, 2) \approx$$

$$\approx 0.2087432125056015707175071603149713862299748799628\,3\ldots$$

Here $\zeta(s)$ *is Riemann's zeta function.*

Proof By the definition,

$$6f(n) = \sum_{z\in E_0} g(\frac{3}{2}|z|^{-2}n) + \sum_{z\in E_{\neq 0}} h(\frac{3}{2}|z|^{-2}n) =$$

$$\sqrt{3}\text{л}(\frac{\pi}{3})\frac{9}{4}[\sum_{z\in E_0}(|z|^{-4}n^2 + O(|z|^{-2}n)^{3/2}) + \sum_{z\in E_{\neq 0}}(\frac{1}{27}|z|^{-4}n^2 + O(|z|^{-2}n)^{3/2})] =$$

$$\sqrt{3}\text{л}(\frac{\pi}{3})\frac{9}{4}n^2(\sum_{z\in E_0}|z|^{-4} + \frac{1}{27}\sum_{z\in E_{\neq 0}}|z|^{-4}) + \text{“error term”}$$

The error term can be estimated as follows:

$$\sum_{z\in E_0\cup E_{\neq 0}} O((|z|^{-2}n)^{3/2}) \leq cn^{3/2}\sum_{z\in Eis}|z|^{-3} = O(n^{3/2}).$$

To compute the summands notice that

$$\sum_{z\in E_0}|z|^{-4} + \sum_{z\in E_{\neq 0}}|z|^{-4} =$$

$$\sum_{z\in \mathrm{Prim}Eis}|z|^{-4} = \zeta^{-1}(4)\sum_{z\in Eis\setminus 0}|z|^{-4} = \zeta^{-1}(4)\zeta(Eis, 2).$$

Indeed,

$$\sum_{z\in Eis\setminus 0}|z|^{-4} = \sum_{k\in\mathbb{Z}_{>0}}[\sum_{z'\in \mathrm{Prim}Eis}|kz'|^{-4}] = \zeta(4)\sum_{z'\in \mathrm{Prim}Eis}|z'|^{-4}$$

since for each vector $z \in Eis$ there exists $k \in \mathbb{Z}_{>0}$ and $z' \in \mathrm{Prim}Eis$ such that $z = kz'$.

Now we use (9.2) which implies

$$\sum_{z\in E_0}|z|^{-4} + \frac{1}{81}\sum_{z\in E_{\neq 0}}|z|^{-4} = \sum_{z\in E_0}|z|^{-4} + \sum_{z\in E_{\neq 0}}|3z|^{-4} =$$

$$= \sum_{z\in E_0}|z|^{-4} + \sum_{z\in 3E_{\neq 0}}|z|^{-4} = \sum_{z\in \mathrm{Prim}Eis_0}|z|^{-4} =$$

$$= \zeta^{-1}(4)\sum_{z\in Eis_0\setminus 0}|z|^{-4} = \zeta^{-1}(4)\sum_{z\in Eis\setminus 0}|(1+\omega)z|^{-4} =$$

$$= \frac{1}{9}\zeta^{-1}(4)\sum_{z\in Eis\setminus 0}|z|^{-4} = \frac{1}{9}\zeta^{-1}(4)\zeta(Eis, 2).$$

From this system of linear equations one finds

$$\sum_{z \in E_0} |z|^{-4} = \frac{1}{10} \zeta^{-1}(4) \zeta(Eis, 2)$$

$$\sum_{z \in E_{\neq 0}} |z|^{-4} = \frac{9}{10} \zeta^{-1}(4) \zeta(Eis, 2)$$

It follows that

$$f(n) = \frac{1}{6}\sqrt{3} \cdot \text{л}(\frac{\pi}{3})\frac{9}{4}n^2(\frac{1}{10} + \frac{1}{27}\frac{9}{10})\zeta^{-1}(4)\zeta(Eis, 2) + O(n^{3/2}) =$$

$$= \frac{1}{20}\sqrt{3} \cdot \text{л}(\frac{\pi}{3})\zeta^{-1}(4)\zeta(Eis, 2)n^2 + O(n^{3/2}).$$

□

9.5 Examples and Computer Computations

It follows from an Euler characteristic computation that no triangulation of $\mathbb{R}P^2$ with an odd number of triangles exists.

Here is the list of $f(2n) - f(2n-1)$, i.e., the number of labelled triangulations (T, A, B, C, Δ) of $\mathbb{R}P^2$ with exactly $2n$ triangles, for $n = 1, \ldots, 74$:

0, 1, 4, 0, 16, 1, 12, 17, 20, 0, 46, 8, 18, 34, 40,

12, 64, 9, 36, 48, 60, 6, 94, 41, 24, 64, 72, 24, 112, 8,

60, 81, 94, 24, 160, 56, 42, 82, 114, 24, 160, 58, 60, 126, 96,

30, 190, 60, 96, 81, 160, 54, 184, 65, 72, 194, 132, 24, 238, 96,

90, 130, 220, 60, 232, 62, 84, 192, 214, 24, 286, 105, 90, 160.

Only one triangulation of $\mathbb{R}P^2$ with four triangles exists, see Fig. 9.3.

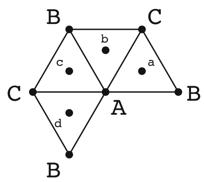


Fig. 9.3 In this case $z = \frac{1+\omega}{3} = \frac{1}{1-\omega^2}$, $(a, b, c, d) = (1, 1, 1, 1)$

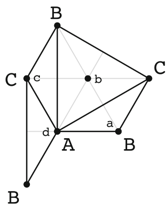


Fig. 9.4 In this case $z = 1$, $(a, b, c, d) = (1, 1, 1, 0)$, and we count this triangulation four times as $(a, b, c, d) = (1, 1, 1, 0), (1, 1, 0, 1), (1, 0, 1, 1), (0, 1, 1, 1)$. Two of them are isometric while another two differ by relabelling $B \to C$

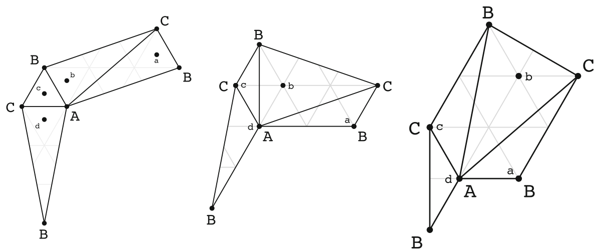


Fig. 9.5 The leftmost picture: $z = \frac{1+\omega}{3}$, $(a, b, c, d) = (4, 1, 1, 1)$ (counted four times). The central pictures and the rightmost picture are representatives for the tuple $(2, 1, 1, 0)$ (counted $12 = 8 + 4$ times). Namely, $z = 1$, $(a, b, c, d) = (2, 1, 1, 0)$ in the central picture (counted eight times). The rightmost picture: $z = 1$, $(a, b, c, d) = (1, 2, 1, 0)$ (counted four times)

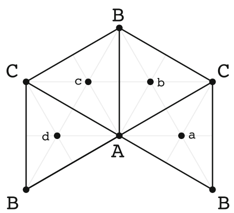


Fig. 9.6 In this case $z = 1$, $(a, b, c, d) = (1, 1, 1, 1)$

Four marked triangulations of $\mathbb{R}P^2$ with 6 triangles exist, see Fig. 9.4.

No triangulation of $\mathbb{R}P^2$ with 8 triangles exists.

Sixteen triangulations with 10 triangles exist, see Fig. 9.5. Only one triangulation with 12 triangles exists, see Fig. 9.6. Plots of the function f(n) and the error term are presented in Figs. 9.7 and 9.8.

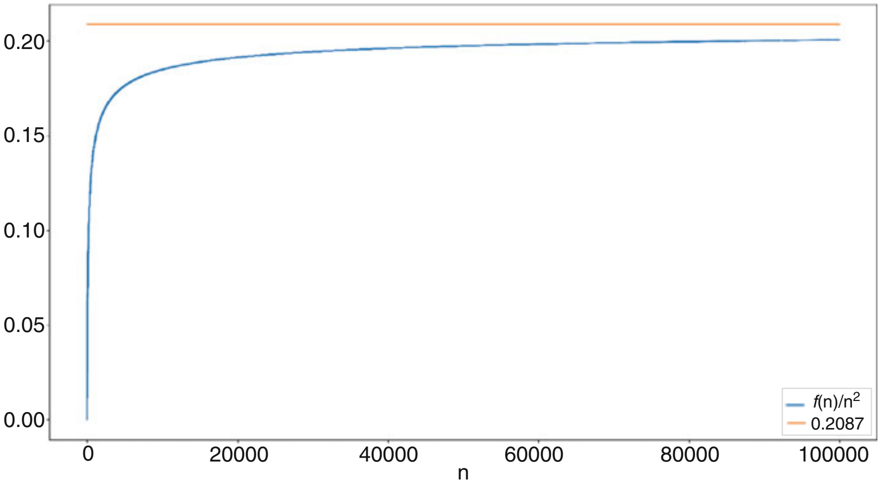


Fig. 9.7 On this plot we see that $f(n)/n^2$ converges to $C = 0.2087\ldots = \frac{1}{20}\sqrt{3}\cdot \text{л}(\frac{\pi}{3})\zeta^{-1}(4)\zeta(Eis, 2)$

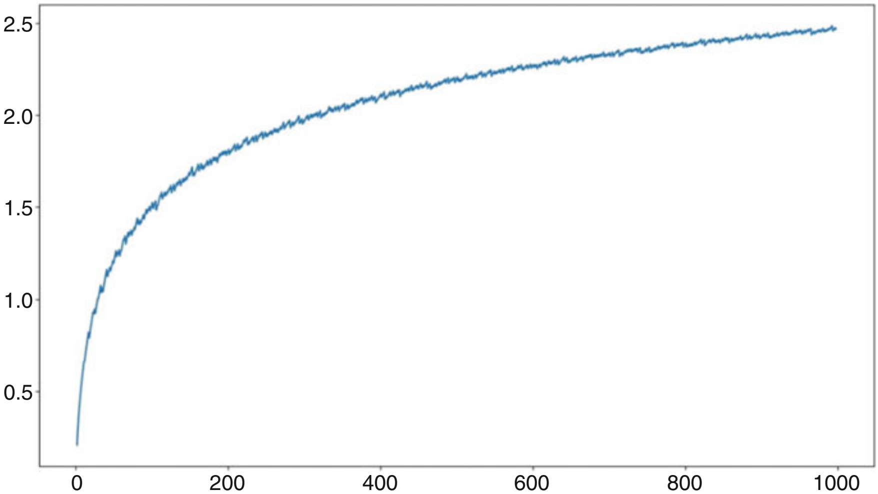


Fig. 9.8 The plot for the error term $\frac{1}{20}\sqrt{3}\text{л}(\frac{\pi}{3})\zeta^{-1}(4)\zeta(Eis, 2)n^2 - f(n)$ divided by $n^{3/2}$ is presented. Thus we see that $f(n) \approx \frac{1}{20}\sqrt{3}\cdot \text{л}(\frac{\pi}{3})\zeta^{-1}(4)\zeta(Eis, 2)n^2 - 2.5n^{3/2}$

Acknowledgements We all are grateful to the Euler Institute for its hospitality and to The Great Mathematical Workshop for the opportunity to start this project. Mikhail Chernavskikh is supported by the Basis Foundation scholarship. Research of Alexander Zakharov is supported by Ministry of Science and Higher Education of the Russian Federation, agreement №075-15-2019-1619.

The research of Nikita Kalinin is supported by the Russian Science Foundation grant №20-71-00007. Theorem 3.1 and Theorem 4.2 have been obtained under support of the RSF grant №20-71-00007. We thank the referee for the careful reading of this chapter.

References

1. A.D. Alexandrov, Existence of a convex polyhedron and of a convex surface with a given metric. Mat. Sbornik **11**(1–2), 15–65 (1942)
2. A. Henn, The hexagonal lattice and the epstein zeta function, in *Dynamical Systems, Number Theory and Applications: A Festschrift in Honor of Armin Leutbecher's 80th Birthday* (World Scientific, Singapore, 2016), pp. 127–140
3. R.E. Schwartz, Notes on shapes of polyhedra (2015). Preprint arXiv:1506.07252
4. W.P. Thurston, Shapes of polyhedra and triangulations of the sphere. *Geometry and Topology monographs, Volume 1: The Epstein birthday schrift, arXiv preprint math/9801088* (1998), pp. 511–549
5. Z. Wang, Shapes of centrally symmetric octahedra with prescribed cone-deficits. Adv. Geom. **21**(2), 179–185 (2021)

Chapter 10
Combination Theorems in Groups, Geometry and Dynamics

Mahan Mj and Sabyasachi Mukherjee

Abstract The aim of this chapter is to give a survey of combination theorems occurring in hyperbolic geometry, geometric group theory and complex dynamics, with a particular focus on Thurston's contribution and influence in the field.

Keywords Kleinian group · Hyperbolic group · Flaring · Holomorphic mating · Simultaneous uniformization · Double limit theorem

AMS 2010 Subject Classification 20F65, 20F67, 37F10, 37F32, 30F60, 30F40, 57M50

10.1 Introduction

The aim of this survey is to give an eclectic account of combination theorems in hyperbolic geometry, geometric group theory and complex dynamics. Thurston's contribution and influence in the theme is pervasive, and we will only be able to touch upon some of these aspects. The hope in writing this survey is therefore only to whet the appetite of the reader and provide some references to more detailed articles and surveys.

Combination theorems have a long history, going back to Klein's paper from 1883 [73]. A major subsequent development in terms of combination theorems for Kleinian groups is due to Maskit [96–100]. See Sect. 10.2 for further details.

Both authors were supported by the Department of Atomic Energy, Government of India, under project no.12-R&D-TFR-5.01-0500 as also by an endowment of the Infosys Foundation. MM was also supported in part by a DST JC Bose Fellowship, Matrics research project grant MTR/2017/000005, CEFIPRA project No. 5801-1. SM was supported in part by SERB research project grant SRG/2020/000018.

M. Mj (✉) · S. Mukherjee
School of Mathematics, Tata Institute of Fundamental Research, Mumbai, India
e-mail: mahan@math.tifr.res.in; sabya@math.tifr.res.in

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_10

A combination theorem of a different, more complex analytic, flavor was introduced by Ahlfors and Bers (see Sect. 10.3). The Bers Simultaneous Uniformization Theorem provided a way of combining two abstractly isomorphic Fuchsian surface groups into a single Kleinian surface group. Equivalently, two discrete faithful representations $\rho_i : \pi_1(S) \to \mathrm{PSL}(2, \mathbb{R})$ are combined in a dynamically natural way into a single discrete faithful representation $\rho : \pi_1(S) \to \mathrm{PSL}(2, \mathbb{C})$. The study of Kleinian groups around this time thus took on a rather complex analytic orientation.

In the 1970s and 1980s, a phase transition occurred in the theory with the advent of Thurston, who combined the above two strands into one unified theme, and vastly generalized both. He introduced a 3-dimensional hyperbolic geometry point of view, leading to his proof of hyperbolization of atoroidal Haken manifolds [70, 121, 122, 143–145]. Particular mention must be made of his Double Limit Theorem that may be thought of as a limiting case of the Simultaneous Uniformization Theorem. (See Sect. 10.4.)

Thurston's work has had a deep and profound influence on hyperbolic geometry ever since, and has provided a template for related developments in geometric group theory and complex dynamics. In geometric group theory, Bestvina and Feighn [12] isolated the coarse geometric features of Thurston's combination theorem and proved a highly influential combination theorem for Gromov-hyperbolic groups [49], spawning considerable activity and several generalizations [4, 28, 45, 115, 116]. In particular, the main theorem of [12] was extended to a relatively hyperbolic setup [4, 28, 45, 115] and also to the setup of a coarse-geometric analog of bundles [116]. (See Sect. 10.5 for further details.)

In a relatively recent major development leading to a resolution of Thurston's virtual Haken conjecture by Agol and Wise [1, 149], Haglund and Wise [54] proved a combination theorem for virtually special cubulable hyperbolic groups [53]. For these and related developments, see Sect. 10.6.

In a closely related theme, Fatou and Julia laid the foundation of the theory of dynamics of rational maps on the Riemann sphere in the first quarter of the twentieth century [40–43, 65, 66]. These early developments in the field drove Fatou to observe similarities between limit sets of Kleinian groups and Julia sets of rational maps: 'L'analogie remarquée entre les ensembles de points limites des groupes Kleineens et ceux qui sont constitués par les frontières des régions de convergence des itérées d'une fonction rationnelle ne paraît d'ailleurs pas fortuite et il serait probablement possible d'en faire la synthèse dans une théorie générale des groupes discontinus des substitutions algébriques.' After several decades, this analogy was set on a firm footing by Sullivan with the introduction of quasiconformal techniques in the study of rational dynamics [140]. Shortly afterwards, Sullivan put forward a dictionary between the aforementioned classes of conformal dynamical systems, Thurston proved a topological characterization for an important class of rational maps [33] as a philosophical analog of the hyperbolization of atoroidal Haken 3-manifolds. The theory of polynomial mating, designed by Douady and Hubbard [35], extends the notion of a combination theorem from the world of Kleinian groups to that of complex dynamics. This theme too bears the tell-tale stamp of Thurston.

In fact, Thurston's topological characterization of rational maps is an invaluable tool in constructing such matings [33, 136] (see Sect. 10.7).

The idea of combining Kleinian groups with rational maps was first conceived by Bullett and Penrose in [21], where they used iterated algebraic correspondences to 'mate' the modular group $\mathrm{PSL}(2, \mathbb{Z})$ with certain quadratic polynomials. More recently, a one complex variable approach was adopted to bind together the actions of Kleinian groups and rational maps in the dynamics of a single map. This perspective can be thought of as a unification of Bers simultaneous uniformization theorem (and in certain cases, Thurston's double limit theorem) with the Douady–Hubbard theory of polynomial mating. The crucial difference between this mating framework and that of Bullett-Penrose is that here one extracts a non-invertible map from a Kleinian group that is *orbit equivalent* to the group on its limit set (i.e., one extracts a semi-group dynamics from the dynamics of a non-commutative group), and 'mates' this map with the dynamics of a polynomial. In the anti-holomorphic setting, this is achieved by associating a piecewise circular reflection map, called the *Nielsen map*, to a Kleinian reflection group. The simplest example of this mating phenomenon is given by the *Schwarz reflection map* associated with a simply connected *quadrature domain*; namely, the exterior of a deltoid curve (which is the conformal mating of the anti-polynomial $\overline{z}^2$ and the ideal triangle reflection group). A series of papers [75, 78, 79, 91] culminated in a comprehensive framework for conformally mating Kleinian reflection groups with anti-holomorphic polynomials (see Sect. 10.8.1).

On the holomorphic side, a framework for combining Kleinian groups with polynomial maps was devised in [114]. The key player in this setting of combination theorems is a class of *piecewise Möbius* maps, termed *mateable* maps. Such maps are dynamically and combinatorially compatible with Kleinian groups on the one hand and polynomials on the other. In particular, a mateable map associated to a Kleinian group is orbit equivalent to the group on the limit set. While the simplest examples of mateable maps are given by the classical *Bowen–Series maps* associated with Fuchsian punctured sphere groups, a new class of examples called *higher Bowen–Series maps* was also described in [114]. These maps enjoy various close connections with Bowen–Series maps, and are interesting in their own right (for instance, they are responsible for failure of *topological orbit equivalence rigidity* of Fuchsian groups). It turns out that any mateable map can be conformally mated with suitable complex polynomials giving rise to disconnected moduli spaces of matings of punctured spheres with complex polynomials (see Sect. 10.8.2).

10.2 Klein–Maskit Combination for Kleinian Groups

A discrete subgroup Γ of $PSL_2(\mathbb{C})$ is called a *Kleinian group*. The *limit set* of the Kleinian group Γ, denoted by Λ_Γ, is the collection of accumulation points of a Γ-orbit $\Gamma \cdot z$ for some $z \in \hat{\mathbb{C}}$. Λ_Γ is independent of z. It may be thought of as the locus of chaotic dynamics of Γ on $\hat{\mathbb{C}}$, i.e. for Γ non-elementary and any $z \in \Lambda_\Gamma$,

$\Gamma \cdot z$ is dense in Λ_Γ. We shall identify the Riemann sphere $\hat{\mathbb{C}}$ with the sphere at infinity $\mathbb{S}^2$ of $\mathbb{H}^3$. The complement of the limit set $\hat{\mathbb{C}} \setminus \Lambda_\Gamma$ is called the domain of discontinuity $\Omega(\Gamma)$ of Γ. If the Kleinian group Γ is torsion-free, it acts freely and properly discontinuously on $\Omega(\Gamma)$ with a Riemann surface quotient.

Definition 10.1 A set D is called a for Γ, if

1. $D \neq \emptyset$,
2. $D \subset \Omega(\Gamma)$, and
3. $g(D) \cap D = \emptyset$, for all $g \in \Gamma$, $g \neq 1$.

If, further, $\bigcup_{g\in\Gamma} g.D = \Omega(\Gamma)$, then D is called a *fundamental domain* for Γ.

The story of combination theorems starts with the following theorem of Klein:

Theorem 10.1 (Klein Combination Theorem [73]) *Let Γ_1, Γ_2 be Kleinian groups with fundamental domains D_1, D_2 respectively. Assume that the interior of D_1 (resp. D_2) contains the boundary and exterior of D_2 (resp. D_1).Then the group Γ generated by Γ_1, Γ_2 is Kleinian, and $D = D_1 \cap D_2$ is a fundamental domain for Γ.*

In the 1960s, Maskit started working on extending the Klein combination Theorem 10.1 to a more general setup. Maskit's work on combination theorems for Kleinian groups started with the following.

Theorem 10.2 (Klein-Maskit Combination Theorem for Free Product with Amalgamation [95]) *Let Γ_1, Γ_2 be Kleinian groups with domains of discontinuity Ω_1, Ω_2 respectively. Let $H = \Gamma_1 \cap \Gamma_2$. Let D_1, D_2, Δ be partial fundamental domains for Γ_1, Γ_2, H respectively. For i = 1,2, set $E_i = H.D_i$. Denote the interior of $D = E_1 \cap E_2 \cap \Delta$ by D'. If*

1. $D' \neq \emptyset$,
2. $E_1 \cup E_2 = \Omega_1 \cup \Omega_2$.

Then the group Γ generated by Γ_1, Γ_2 is Kleinian, D' is a partial fundamental domain for Γ, and $\Gamma = \Gamma_1 *_H \Gamma_2$ is the free product with amalgamation of Γ_1, Γ_2 along H. Further, $gD \cap D = \emptyset$, for all $g \in \Gamma$, $g \neq 1$.

In [96], Maskit strengthened the above theorem by determining precisely a fundamental domain for the group.

Theorem 10.3 (Klein-Maskit Combination Theorem for Free Product with Amalgamation [96]) *Let Γ_1, Γ_2 be Kleinian groups with domains of discontinuity Ω_1, Ω_2 respectively. Let $H \subset \Gamma_1 \cap \Gamma_2$. such that H is either cyclic or consists only of the identity. Let D_1, D_2, Δ be fundamental domains for Γ_1, Γ_2, H respectively. For $i = 1, 2$, set $E_i = H.D_i$. Denote the interior of $D = E_1 \cap E_2 \cap \Delta$ by D'. Suppose $E_1 \cup E_2 = \Omega(H)$ and that $D' \neq \emptyset$. Assume further that there is a simple closed curve γ contained in $int(E_1 \cup E_2) \cup \Lambda_H$ such that γ is invariant under H, the closure of $\gamma \cap \Delta$ is contained in $int(E_1 \cap E_2)$ and γ separates both $E_1 \setminus E_2$ and*

$E_2 \setminus E_1$. *Then the group* Γ *generated by* Γ_1, Γ_2 *is Kleinian,* $\Gamma = \Gamma_1 *_H \Gamma_2$*, and* D *is a fundamental domain for* G.

Subsequently, in [97, 98], Maskit upgraded Theorem 10.3 to the following. We start with two Kleinian groups Γ_1, Γ_2 with $H \subset \Gamma_1 \cap \Gamma_2$, where $H \neq \Gamma_1, \Gamma_2$. We are also given a simple closed curve γ dividing the Riemann sphere $\hat{\mathbb{C}}$ into two closed topological discs, B_1 and B_2, where B_i is precisely invariant under H in Γ_i. More precisely, B_i is $H-$invariant, and if $g \in \Gamma_i \setminus H$, then $g(B_i) \cap B_i = \emptyset$. Then $\Gamma = \langle \Gamma_1, \Gamma_2 \rangle$, the group generated by Γ_1 and Γ_2, is also a Kleinian group. What really needs to be proved in all these cases is the discreteness of Γ.

In all these cases, Maskit shows that $\Gamma = \Gamma_1 *_H \Gamma_2$, i.e. $\Gamma = \langle \Gamma_1, \Gamma_2 \rangle$ is equal to the free product with amalgamation of Γ_1, Γ_2 along H. Further, by carefully choosing fundamental domains for Γ_1, Γ_2 one can ensure that their intersection will be a fundamental domain for Γ. Thus, the basic hypothesis guaranteeing discreteness of Γ can be summarized as follows:

1. The disks B_1 and B_2 are both invariant under H.
2. The $(\Gamma_1 \setminus H)-$translates of B_1 are disjoint disks in B_2.
3. The $(\Gamma_2 \setminus H)-$translates of B_2 are disjoint disks in B_1.

There is also a version of the Klein-Maskit combination theorem for HNN extensions. We are given a single group Γ_0, with two subgroups H_1 and H_2, two closed disks B_1 and B_2 which have disjoint projections to $\Omega(\Gamma_0)/\Gamma_0$, where

1. H_i preserves B_i,
2. there exists a Möbius transformation h mapping the outside of B_1 onto the inside of B_2 and conjugating H_1 to H_2.

Maskit then shows that $\Gamma = \langle \Gamma_0, h \rangle$ is a Kleinian group. Also $\Gamma = \Gamma_0 *_H$ is the HNN-extension of G_0 along H, where the two inclusions of H map to H_1, H_2 and h is the stable letter conjugating one to the other. Further, by carefully choosing a fundamental domain D for Γ_0, one can ensure that $D \setminus (B_1 \cup B_2)$ is a fundamental domain for Γ.

Maskit weakens the hypotheses further in [100], allowing translates of the closed disks B_1, B_2 to have common boundary points. However, in [100] he requires that such points of intersection also be ordinary points of our original group. In [100], it is also shown that Γ is geometrically finite if and only if the original groups are so. The basic topological tool used in the proof is a Jordan curve γ in $\hat{\mathbb{C}}$ and its translates under a Kleinian group. The standard hypothesis in these papers is the 'almost disjointness' of γ from all its translates. More precisely, if $g(\gamma) \cap \gamma \neq \emptyset$, then it is required that $g(\gamma)$ is entirely contained in the closure of one of the open disks bounded by γ. Thus, a substantial amount of the technical difficulty in [98, 100] comes from controlling the points of intersection $g(\gamma) \cap \gamma$.

To conclude this section, we refer the reader to

1. Work of Li et al. [83, 84] for generalizations of the Klein-Maskit combination theorems to higher dimensions.

2. Work of Dey et al. [34] for a combination theorem for Anosov subgroups, a natural class of discrete subgroups of higher rank Lie groups that generalizes convex cocompact subgroups of $PSL_2(\mathbb{C})$.

10.3 Simultaneous Uniformization and Quasi-Fuchsian Groups

The aim of this section is to give a brief account of Bers simultaneous uniformization theorem. The reason is twofold. First, it provides the context for Thurston's double limit theorem in Sect. 10.4.2. Second, it is the Kleinian group analog for the Douady–Hubbard mating construction [35], and more generally the original motivation for the mating constructions in Sect. 10.8.

Fix a surface S. The collection of all representations $\rho : \pi_1(S) \to PSL_2(\mathbb{C})$ up to conjugacy (in $PSL_2(\mathbb{C})$) is called the *character variety* and is represented as $\mathcal{R}(S)$. We note in passing that the appropriate quotient by $PSL_2(\mathbb{C})$ of the space of all representations $\rho : \pi_1(S) \to PSL_2(\mathbb{C})$ is the GIT quotient. This is needed in order to obtain the structure of a variety on $\mathcal{R}(S)$.

10.3.1 Topologies on Space of Representations

For future reference, we summarize here a natural collection of topologies on the space of *discrete faithful* $\rho : \pi_1(S) \to PSL_2(\mathbb{C})$. The *algebraic topology* is the topology of pointwise convergence on elements of $\pi_1(S)$:

Definition 10.2 We shall say that a sequence of representations $\rho_n : \pi_1(S) \to PSL_2(\mathbb{C})$ converges *algebraically* to $\rho_\infty : \pi_1(S) \to PSL_2(\mathbb{C})$ if for all $g \in \pi_1(S)$, $\rho_n(g) \to \rho_\infty(g)$ in $PSL_2(\mathbb{C})$.

The collection of conjugacy classes of discrete faithful representations of $\pi_1(S)$ into $PSL_2(\mathbb{C})$ equipped with the algebraic topology is denoted as $AH(S)$. Thus, $AH(S) \subset \mathcal{R}(S)$ comes naturally equipped with a complex analytic structure. The space of discrete faithful representations of $\pi_1(S)$ into $PSL_2(\mathbb{R})$ equipped with the algebraic topology is precisely the Teichmüller space. Thus, the Teichmüller space sits 'diagonally' in $AH(S)$.

For analyzing convergence from a geometric point of view, the natural topology is the *geometric topology*, or equivalently, the Gromov–Hausdorff topology.

Definition 10.3 Let $\rho_n : \Gamma \to PSL_2(\mathbb{C})$ be a sequence of discrete, faithful representations of a finitely generated, torsion-free, nonabelian group Γ. Thus, $\rho_n(\Gamma)$ is a sequence of closed subsets of $PSL_2(\mathbb{C})$. If $G \subset PSL_2(\mathbb{C})$ is a closed subgroup such that $\rho_n(\Gamma)$ converges to G in the Gromov–Hausdorff topology, then $\rho_n(\Gamma)$ is said to *converge geometrically* to G, and G is called the *geometric limit*.

Definition 10.4 G_n converges *strongly* to G if G_n converges to G both algebraically and geometrically.

10.3.2 Simultaneous Uniformization

Definition 10.5 Let $\rho : \pi_1(S) \to PSL_2(\mathbb{C})$ be a discrete faithful representation such that the limit set of $G = \rho(\pi_1(S))$ is a topological circle in $\mathbb{S}^2$. Then G is said to be *quasi-Fuchsian*. The collection of conjugacy classes of quasi-Fuchsian representations is denoted as $QF(S)$.

Note that $QF(S)$ is contained in $AH(S)$ and hence inherits a complex analytic structure. The domain of discontinuity Ω of a quasi-Fuchsian G consists of two open invariant disks Ω_1, Ω_2. Hence the quotient Ω/G is the disjoint union $\Omega_1/G \sqcup \Omega_2/G$. Hence we have a map $\tau : QF(S) \to Teich(S) \times Teich(S)$, where $Teich(S)$ denotes the Teichmüller space of S. The **Bers simultaneous Uniformization Theorem** asserts:

Theorem 10.4 ([9, 10]) *$\tau : QF(S) \to Teich(S) \times Teich(S)$ is a homeomorphism.*

Hence, given any two conformal structures T_1, T_2 on a surface, there is a unique discrete quasi-Fuchsian G whose limit set Λ_G is topologically a circle, and the quotient of whose domain of discontinuity is $T_1 \sqcup T_2$. See Fig. 10.1 [67], where the inside and the outside of the Jordan curve correspond to Ω_1, Ω_2.

We refer to [61] for a proof of Theorem 10.4 and summarize the main ideas here. Theorem 10.4 is essentially complex analytic in nature and goes back to an understanding of the Beltrami partial differential equation due to Morrey. Let K_S denote the canonical bundle of the Riemann surface S (if S has punctures as a hyperbolic surface, we regard them as marked points in the complex analytic category).

Definition 10.6 A *Beltrami differential* on S is an L^∞ section of $K_S^{-1} \otimes \overline{K_S}$, where $\overline{K_S}$ denotes the complex conjugate of K_S. The space of Beltrami differentials on S will be denoted as $\mathcal{D}_b(S)$

The local expression for an element of $\mathcal{D}_b(S)$ in a complex analytic chart $U \subset S$ is thus given by $\mu \frac{d\bar{z}}{dz}$, where $\mu \in L^\infty(U)$ is called a *Beltrami coefficient*.

Definition 10.7 A *quasiconformal map* between two Riemann surfaces S_1 and S_2 is a homeomorphism $\phi : S_1 \to S_2$ having locally square-integrable weak partial derivatives such that

$$\mu = \frac{\phi_{\bar{z}}}{\phi_z}$$

satisfies $||\mu||_\infty < 1$. Here, μ is called the *Beltrami coefficient* of ϕ.

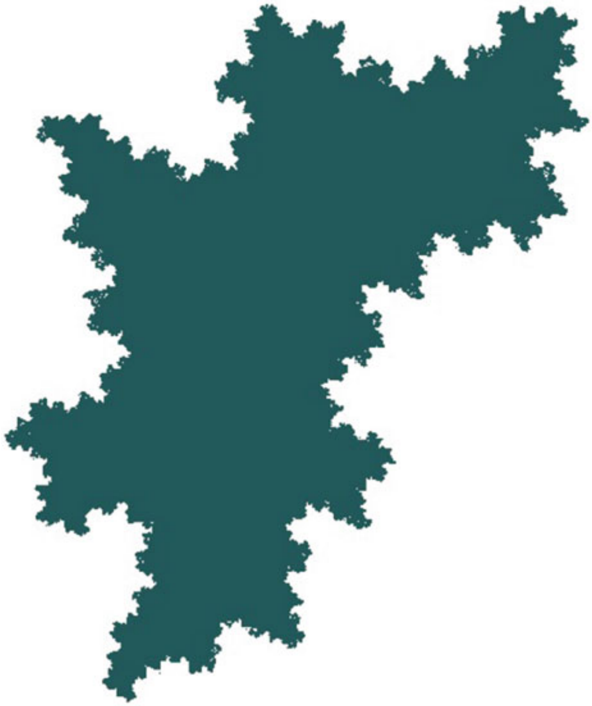


Fig. 10.1 Quasi-Fuchsian group limit set

The first major ingredient in the proof of Theorem 10.4 is the Measurable Riemann mapping theorem. As pointed out by Hubbard in [61, p. 149], the Beltrami coefficient μ really represents an almost-complex structure on U and the Measurable Riemann mapping theorem (due to Ahlfors–Bers–Morrey) below ensures its integrability to a complex structure.

Theorem 10.5 (Measurable Riemann Mapping Theorem [61, Theorem 4.6.1])

Existence of Quasiconformal Maps *Let $U \subset \mathbb{C}$ be open. Let $\mu \in L^\infty(U)$ satisfying $||\mu||_\infty < 1$. Then there exists a quasiconformal mapping $f : U \to \mathbb{C}$ solving the Beltrami equation*

$$\frac{\partial f}{\partial \overline{z}} = \mu \frac{\partial f}{\partial z}.$$

Uniqueness of Quasiconformal Maps *If g is another quasiconformal solution to the Beltrami equation above, then there exists a univalent analytic function ϕ : $f(U) \to \mathbb{C}$ such that $g = \phi \circ f$.*

The rest of this brief account of Theorem 10.4 follows [50] which captures the relevant conformal geometry. Recall that we have fixed a base Riemann surface S.

Let $\Gamma < PSL_2(\mathbb{R})$ be a (base) Fuchsian group uniformizing S. Let S' be an arbitrary point in the Teichmüller space of the underlying topological surface. Theorem 10.4 then associates to S' a quasiconformal map $\Phi : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ fixing the three points $0, 1, \infty$, and conjugating the action of Γ to that of a Kleinian group $\Gamma(S, S')$, such that

1. Φ is conformal on the lower half-plane.
2. $\Gamma(S, S')$ leaves invariant the images of the lower and upper half-planes.
3. The quotient of the lower half-plane by $\Gamma(S, S')$ is S.
4. The quotient of the upper half-plane by $\Gamma(S, S')$ is S'.

To prove the existence of a Φ as above, we first note that the Teichmüller space can be identified (via Theorem 10.5) with Beltrami differentials on S with norm bounded by one. Let μ be the Beltrami differential on S corresponding to S'. Next. lift μ to the upper half plane. Extend to a Beltrami coefficient μ_0 on $\hat{\mathbb{C}}$ by defining it to be identically zero on the lower half plane. The map Φ above is then given by the normalized solution to the Beltrami equation with the Beltrami coefficient μ_0. Invariance of μ_0 under Γ ensures the existence of an isomorphism $\rho : \Gamma \to \Gamma(S, S')$ to the desired Kleinian group $\Gamma(S, S')$ such that

$$\Phi \circ g = \rho(g) \circ \Phi,$$

for all $g \in \Gamma$.

10.3.3 Geodesic Laminations

We turn now to the hyperbolic geometry of quasi-Fuchsian groups.

Definition 10.8 A geodesic lamination on a hyperbolic surface is a foliation of a closed subset with geodesics.

A geodesic lamination on a surface may further be equipped with a transverse (positive) measure to obtain a *measured lamination*. The space $\mathcal{ML}(S)$ of measured (geodesic) laminations on S then has the structure of a positive cone in a vector space, i.e. for every $\lambda \in \mathcal{ML}(S)$ and $t \in \mathbb{R}_+$, $t\lambda \in \mathcal{ML}(S)$. It can be projectivized to obtain the space of projectivized measured laminations $\mathcal{PML}(S)$. It was shown by Thurston [39] that

Theorem 10.6 *$\mathcal{PML}(S)$ is homeomorphic to a sphere and can be adjoined to $Teich(S)$ compactifying the latter to a closed ball.*

Definition 10.9 ([142, Definition 8.8.1]) A *pleated surface* in a hyperbolic three-manifold N is a complete hyperbolic surface S of finite area, together with an isometric map $f : S \to N$ such that every $x \in S$ is in the interior of some geodesic segment which is mapped by f to a straight line segment. Also, f must take every cusp of S to a cusp of N

The pleating locus of the pleated surface $f : S \to M$ is the set $\gamma \subset S$ consisting of those points in the pleated surface which are in the interior of unique line segments mapped to line segments.

Proposition 10.7 ([142, Proposition 8.8.2]) *The pleating locus γ is a geodesic lamination on S. The map f is totally geodesic in the complement of γ.*

The geometry of quasi-Fuchsian groups and their relationship with geodesic laminations arises out of the geometry of the convex core that we now describe.

Definition 10.10 Let Γ be an infinite Kleinian group and let $\Lambda \subset \hat{\mathbb{C}}$ denote its limit set. The *convex hull* of Λ is the smallest non-empty closed convex subset of $\mathbb{H}^3$ whose set of accumulation points in $\hat{\mathbb{C}}$ equals Λ. We denote the convex hull of Λ by $CH(\Lambda)$.

The convex hull $CH(\Lambda)$ of a Kleinian group Γ is invariant under Γ. The quotient $CH(\Lambda)/\Gamma \subset \mathbb{H}^3/\Gamma$ is called the *convex core* of the hyperbolic 3-manifold $M = \mathbb{H}^3/\Gamma$.

For a quasi-Fuchsian group $\Gamma = \rho(\pi_1(S))$, the convex core is homeomorphic to a product $S \times [a, b]$ (the Fuchsian case corresponds to $a = b$). The hyperbolic distance between $S \times \{a\}$ and $S \times \{b\}$ is a measure of the geometric complexity of Γ. In [142][Ch. 8], Thurston further shows:

Proposition 10.8 *Let M be a complete hyperbolic 3-manifold corresponding to a quasi-Fuchsian group Γ, and let $CC(M)$ denote its convex core. Then each component of the convex core boundary $\partial CC(M)$ is a pleated surface.*

10.4 Thurston's Combination Theorem for Haken Manifolds

The material in this section provides the core inspiration for most combination theorems that came subsequently.

Definition 10.11 ([57]) A properly embedded surface $(F, \partial F) \subset (M, \partial M)$ in a 3-manifold M with boundary ∂M (possibly empty) is said to be *incompressible* if the inclusion map $i : (F, \partial F) \subset (M, \partial M)$ induces an injective homomorphism of fundamental groups $i_* : \pi_1(F) \to \pi_1(M)$. Further, we require that for every boundary component γ of ∂F, $i_* : \pi_1(\gamma) \to \pi_1(\partial M)$ is injective. (The second condition is automatic when F is not a disk.)

An embedded incompressible surface $(F, \partial F) \subset (M, \partial M)$ is said to be *boundary parallel* if F can be isotoped into ∂M keeping $\partial F \subset \partial M$ fixed.

A compact 3-manifold M (possibly with boundary ∂M) is said to be *Haken* if

1. $\pi_2(M) = 0$.
2. There exists an embedded incompressible surface $(F, \partial F)$ that is not boundary parallel.

M is said to be atoroidal if $\pi_1(M)$ contains no $\mathbb{Z} \oplus \mathbb{Z}$ subgroups. M is said to be acylindrical if any embedded incompressible annulus in $(M, \partial M)$ is boundary parallel.

We summarize Thurston's celebrated hyperbolization theorem now and then give a brief account of the ingredients that go into the proof.

Theorem 10.9 ([143–145]) *Let M be a compact atoroidal Haken 3-manifold. Then M is hyperbolic.*

There is a version of Theorem 10.9 for 3-manifolds with torus boundary components also. But, in the interests of exposition, we shall largely focus on the compact atoroidal case. The proof of Theorem 10.9 breaks into two principal pieces:

1. M is compact atoroidal Haken and does not fiber over $\mathbb{S}^1$. This case will be described in Sect. 10.4.1.
2. M fibers over the circle with fiber F. This case will be described in Sect. 10.4.2.

10.4.1 Non-fibered Haken 3-Manifolds

There are a number of detailed expositions for the compact atoroidal Haken non-fibered case and we point out [70, 101, 121] in particular.

It is a fundamental fact of 3-manifold topology [57, Chapter 13] that any Haken manifold admits a Haken hierarchy. Cutting $(M, \partial M)$ open along $(F, \partial F)$ gives us a new (possibly disconnected) atoroidal 3-manifold with boundary. The cut open manifold is automatically Haken, and we can proceed inductively. At the last stage, we are left with a finite collection of balls, and these are clearly hyperbolic.

Thus, in order to prove Theorem 10.9 in the non-fibered case, an essential step is the following:

Theorem 10.10 ([121]) *Let M_1 be an acylindrical atoroidal 3-manifold with non-empty incompressible boundary ∂M_1 whose interior admits a complete hyperbolic metric. Let $\tau : \partial M_1 \to \partial M_1$ be an orientation-reversing involution. Then the interior of $M = M_1/\tau$ admits a complete hyperbolic metric.*

To prove Theorem 10.10, a first tool is the following generalization of Theorem 10.4:

Theorem 10.11 *Let M_1 be a compact 3-manifold with boundary such that*

1. *The interior of M_1 admits a complete hyperbolic metric.*
2. *No component of ∂M_1 is homeomorphic to a torus or a sphere.*

Then the space of complete hyperbolic metrics on M_1 is given by $Teich(\partial M_1)$.

Let $\partial M_1 = \sqcup_i \Sigma_i$, where each Σ_i is a surface of genus greater than one. Then $Teich(\partial M_1) = \prod_i Teich(\Sigma_i)$. Fix a complete hyperbolic structure on M_1 (the existence of such a structure is guaranteed by the hypothesis of Theorem 10.11). This is equivalent to a discrete faithful representation $\rho : \pi_1(M_1) \to PSL_2(\mathbb{C})$. Let $\Gamma = \rho(\pi_1(M_1))$. Then each $\Sigma_i \subset \partial M_1$ gives (via inclusion) a conjugacy class of quasi-Fuchsian subgroups of Γ. Thus the involution τ of Theorem 10.10 induces a map

$$\sigma : Teich(\partial M_1) \to Teich(\partial M_1).$$

The map σ is called the *skinning map*. The existence of a complete hyperbolic structure on $M = M_1/\tau$ is equivalent to the existence of a fixed point of the skinning map σ, as such a fixed point ensures an isometric gluing. Thurston's fixed point theorem can now be stated as the following: If M is atoroidal then σ has a fixed point. It now follows from the Klein-Maskit combination theorem (Sect. 10.2) that M admits a complete hyperbolic structure. The acylindricity hypothesis of Theorem 10.10 guarantees that M is atoroidal, completing an outline of the proof of Theorem 10.10.

An effective method of proving the existence of a fixed point of the skinning map σ was taken by McMullen in [101]. As mentioned in [121] Hubbard had observed that the analytical formula for the coderivative of the skinning map relates it to the Theta operator in Teichmüller theory. McMullen studies the fixed point problem via $||D\sigma||$, the norm of the derivative of the skinning map. He reproves Theorem 10.10 by showing that if M_1 is acylindrical, then there exists $c < 1$ such that $||D\sigma|| < c$ guaranteeing a solution to the gluing problem.

Both Thurston's fixed-point theorem and McMullen's estimates in [101] are easiest to state when M_1 is acylindrical. However, both approaches can be refined to conclude hyperbolicity of M as long as M_1 is not of the form $S \times I$ and τ glues $S \times \{0\}$ to $S \times \{1\}$. The excluded case is that of 3-manifolds fibering over the circle and involves a completely different approach that we describe now.

10.4.2 The Double Limit Theorem

We shall follow [122] to give an outline of the steps involved in the hyperbolization of 3-manifolds fibering over the circle. Recall (Theorem 10.6) that the space of projectivized measured laminations $\mathcal{PML}(S)$ compactifies $Teich(S)$. Thurston's double limit theorem may be thought of as an extension of the simultaneous uniformization Theorem 10.4 to the case where the pair (τ_1, τ_2) of Riemann surfaces in $Teich(S) \times Teich(S)$ is replaced by a pair $(\ell_1, \ell_2) \in \overline{Teich(S)} \times \overline{Teich(S)}$, where $\overline{Teich(S)} = Teich(S) \cup \mathcal{PML}(S)$ denotes the Thurston compactification of $Teich(S)$ as in Theorem 10.6.

Dual to any measured lamination $\ell \in \mathcal{ML}(S)$ there is an action of $\pi_1(S)$ on an $\mathbb{R}-$tree. An $\mathbb{R}-$tree is a geodesic metric space such that any two distinct points are joined by a unique arc isometric to an interval in $\mathbb{R}$. We refer to [11] for an expository account of group actions on $\mathbb{R}-$trees and convergence of $\Gamma-$spaces, and mention only the following theorem. Fix a group Γ. A triple (X, o, ρ) is called a based $\Gamma-$space if $o \in X$ is a base-point, and Γ acts on X via a homomorphism $\rho : \Gamma \to Isom(X)$ from Γ to the isometry group $Isom(X)$ of X.

Theorem 10.12 ([11, Theorem 3.3]) *Let* (X_i, o_i, ρ_i) *be a convergent sequence of based* Γ*-spaces such that*

1. *Each* X_i *is* δ *hyperbolic, for some* $\delta \geq 0$.
2. *there exists* $g \in \Gamma$ *such that the sequence* $d_i = d_{X_i}(o_i, \rho_i(g)(o_i))$ *is unbounded.*

Then there is a based $\mathbb{R}$*-tree* (T, o) *and an isometric action* $\rho : \Gamma \to Isom(T)$ *such that* $(X_i, o_i, \rho_i) \to (T, o, \rho)$.

Further, the (pseudo)metric on the $\mathbb{R}-$*tree* T *is obtained as the limit of pseudo-metrics* $\frac{d_{(X_i, o_i, \rho_i)}}{d_i}$.

Finally, we shall need the following theorem of Skora [139] on the structure of groups admitting small actions on $\mathbb{R}-$trees.

Theorem 10.13 ([139]) *Let* S *be a finite area hyperbolic surface. Suppose* $\pi_1(S)$ *acts non-trivially on an* $\mathbb{R}-$*tree* T*, such that for every cusp* $\mathbb{P}$ *of* S*,* $\pi_1(\mathbb{P})$ *fixes a point in* T*. Then the stabilizer of each non-degenerate arc of* T *contains no free subgroup of rank 2 if and only if the action is dual to an element of* $\mathcal{ML}(S)$.

An action of $\pi_1(S)$ on an $\mathbb{R}-$tree T such that the stabilizer of each non-degenerate arc of T contains no free subgroup of rank 2 is called a *small action*. Morgan and Shalen [117–119] constructed a compactification of the variety $\mathcal{R}(S)$ by the space of projectivized length functions arising from small actions of $\pi_1(S)$ on $\mathbb{R}-$trees. Skora's theorem 10.13 allows us to replace $\mathcal{PML}(S)$ in Theorem 10.6 by such length functions.

With this background in place we return to an outline of Thurston's double limit theorem [144] following Otal [122]. Let $(\tau_i^+, \tau_i^-) \in Teich(S) \times Teich(S)$ be a sequence of points converging to $(\ell_+, \ell_-) \in \overline{Teich(S)} \times \overline{Teich(S)}$. By Theorem 10.4, we can identify $Teich(S) \times Teich(S)$ with $QF(S)$ and hence assume that $(\tau_i^+, \tau_i^-) \in QF(S)$. For convenience of exposition, we assume that ℓ_+, ℓ_- are both in $\mathcal{PML}(S)$ (a similar statement holds if only one of ℓ_+, ℓ_- lies in $\mathcal{PML}(S)$). Assume further that ℓ_+, ℓ_- *fill* S, i.e. each component of $S \setminus (\ell_+ \cup \ell_-)$ is either simply connected or else is topologically a punctured disk. Let $\rho_i : \pi(S) \to PSL_2(\mathbb{C})$ be the quasi-Fuchsian representation corresponding to $(\tau_i^+, \tau_i^-) \in QF(S)$ and let $\Gamma_i = \rho_i(\pi(S))$. Thurston's double limit theorem now says:

Theorem 10.14 ([144]) *Under the above assumptions, there exists a Kleinian group* Γ *such that* Γ_i *converges to* Γ *in* $AH(S)$.

We sketch Otal's proof following [122] and argue by contradiction. If Γ_i diverges, then Theorem 10.12 shows that there is a limiting small action of $\pi_1(S)$ on an $\mathbb{R}-$tree T. By Theorem 10.13 such a small action is dual to a measured lamination ℓ on S.

It is then shown in [122] that any measured lamination that intersects ℓ essentially is realizable in T. Hence at least one of ℓ_+ and ℓ_- must be realizable in T, since the two together fill S. Without loss of generality, suppose ℓ_+ is realizable in T. This allows us to approximate ℓ_+ by simple closed curves σ on S and estimate the translation length $l_i(\sigma)$ of σ in $\mathbb{H}^3/\Gamma_i$. The estimate thus obtained contradicts a classical estimate of $l_i(\sigma)$ due to Ahlfors obtained in terms of the length of the geodesic realization of σ in τ_i^+ and τ_i^-. This final contradiction proves Theorem 10.14.

Finally, to hyperbolize an atoroidal 3-manifold fibering over the circle with monodromy ϕ, one picks a base Riemann surface τ, and sets $\tau_i^+ = \phi^i(\tau)$ and $\tau_i^- = \phi^{-i}(\tau)$. Then ℓ_+, ℓ_- turn out to be the stable and unstable laminations of ϕ. The 3-manifold M obtained from the double limit theorem is easily seen to be invariant under ϕ, and hence M admits a quotient which is the required hyperbolic 3-manifold.

10.5 Combination Theorems in Geometric Group Theory: Hyperbolic Groups

The fundamental combination theorem in the context of hyperbolic groups in the sense of Gromov [49] is due to Bestvina and Feighn [12]. The theorem was motivated by Thurston's combination Theorem 10.9. In the context of geometric group theory, free products with amalgamation and HNN extensions can be treated on a common footing by passing to the universal cover and looking at the resulting Bass–Serre tree of spaces [133]. Thus, while the main combination theorem of [12] provides only a weaker conclusion than Theorem 10.9 inasmuch as it establishes Gromov-hyperbolicity, the context is considerably more general and works for trees of spaces. It turns out that the sufficient condition in [12] is also necessary and this converse direction was established by Gersten [48], Bowditch [14] and others. The paper [12] spawned considerable activity in geometric group theory and have been giving rise to a number of combination theorems [4, 28, 45–47, 93, 115, 116] right up to the time of writing this article. A forthcoming book of Kapovich and Sardar [71] furnishes a definitive account and rather general versions of the material in Sects. 10.5.1 and 10.5.2.

10.5.1 Trees of Spaces

The framework of [12] is that of a tree of spaces. We follow the exposition in [113] to define the relevant notions.

Definition 10.12 ([12]) Let (X, d) be a geodesic metric space. Let T be a simplicial tree. Let $\mathcal{V}(T)$ and $\mathcal{E}(T)$ denote the vertex set and edge set of T respectively. Then $P : X \to T$ is said to be a tree of geodesic metric spaces satisfying the *quasi-isometrically embedded condition* (or simply, the *qi condition*) if there exists a map $P : X \to T$, and constants $K \geq 1, \epsilon \geq 0$ satisfying the following:

1. $\forall v \in \mathcal{V}(T)$, $X_v = P^{-1}(v) \subset X$ equipped with the induced path metric d_v is a geodesic metric space X_v. Also, the inclusion maps $i_v : X_v \to X$ are uniformly

proper, i.e. $\forall M > 0$, $v \in T$ and $x, y \in X_v$, there exists $N > 0$ such that $d(i_v(x), i_v(y)) \leq M$ implies $d_v(x, y) \leq N$.

2. Let $e = [v_1, v_2] \in \mathcal{E}(T)$ with initial and final vertices v_1 and v_2 respectively (we assume that all edges have length 1). Let X_e be the pre-image under P of the mid-point of e. There exist continuous maps $f_e : X_e \times [v_1, v_2] \to X$, such that $f_e|_{X_e \times (v_1, v_2)}$ is an isometry onto the pre-image of the interior of e equipped with the path metric d_e.
   Further, we demand that f_e is fiber-preserving, i.e. projection to the second co-ordinate in $X_e \times [v_1, v_2]$ corresponds via f_e to projection to the tree $P : X \to T$.
3. $f_e|_{X_e \times \{v_1\}}$ and $f_e|_{X_e \times \{v_2\}}$ are (K, ϵ)-quasi-isometric embeddings into X_{v_1} and X_{v_2} respectively. We shall often use the shorthand f_{e,v_1} and f_{e,v_2} for $f_e|_{X_e \times \{v_1\}}$ and $f_e|_{X_e \times \{v_2\}}$ respectively.

We shall refer to K, ϵ as the constants or parameters of the qi-embedding condition.

If there exists $\delta > 0$ such that the vertex and edge spaces X_v, X_e above are all δ-hyperbolic metric spaces for all vertices v and edges e of T, then $P : X \to T$ will be called *a tree of hyperbolic metric spaces*.

Definition 10.13 ([12]) A continuous map $f : [-k, k] \times I \to X$ is called a *hallway* of length $2k$ if it satisfies the following:

1. $f^{-1}(\cup X_v : v \in T) = \{-k, \cdots, k\} \times I$
2. f is transverse, relative to condition (1) to $\cup_e X_e$.
3. for all $i \in \{-k, \cdots, k\}$, f maps $i \times I$ to a geodesic in X_v for some vertex space X_v.

Definition 10.14 ([12]) A hallway $f : [-k, k] \times I \to X$ is said to be *ρ-thin* if

$$d(f(i, t), f(i + 1, t)) \leq \rho$$

for all i, t.

A hallway $f : [-m, m] \times I \to X$ is called *λ-hyperbolic* if

$$\lambda l(f(\{0\} \times I)) \leq \max \{l(f(\{-m\} \times I)), l(f(\{m\} \times I))\}.$$

The *girth* of the hallway is defined to be the quantity

$$\min_i \{l(f(\{i\} \times I))\}.$$

A hallway is *essential* if the edge path in T resulting from projecting the hallway under $P \circ f$ onto T does not backtrack (and is therefore a geodesic segment in the tree T).

Definition 10.15 (Hallways Flare Condition [12]) The tree of spaces, X, is said to satisfy the *hallways flare* condition if there exist $\lambda > 1$ and $m \geq 1$ such that the following holds:

$\forall \rho > 0$ there exists $H(= H(\rho))$ such that any ρ-thin essential hallway of length $2m$ and girth at least H is λ-hyperbolic.

The constants λ, m are referred to as the constants or parameters of the hallways flare condition. If, further, the constant ρ is fixed, then H will also be called a constant or parameter of the hallways flare condition.

With these notions in place, we can state the main geometric combination theorem of [12]:

Theorem 10.15 *Let $P : X \to T$ be a tree of hyperbolic spaces satisfying the qi-embedded condition (as in Definition 10.12). Further, suppose that the hallways flare condition (as in Definition 10.15) is satisfied. Then X is hyperbolic.*

The proof of Theorem 10.15 in [12] proceeds by establishing a linear isoperimetric inequality ensuring hyperbolicity. We shall indicate a different proof scheme below in the special case that the edge-to-vertex inclusion maps are uniform quasi-isometries rather than qi-embeddings. The forthcoming book [71] provides a new proof as well.

10.5.2 Metric Bundles

The notion of a metric bundle [116] adapts the idea of a fiber bundle to a coarse geometric context. We shall describe below the main combination theorem of [116] which is an analog of Theorem 10.15 in this context.

Definition 10.16 Let $(\mathcal{X}, d_X)$ and $(\mathcal{B}, d_B)$ be geodesic metric spaces. Let $c, K \geq 1$ be constants and $h : \mathbb{R}^+ \to \mathbb{R}^+$ a function. $P : \mathcal{X} \to \mathcal{B}$ is called an $(h, c, K)-$ *metric bundle* if

1. P is 1-Lipschitz.
2. For each $z \in \mathcal{B}$, $F_z = P^{-1}(z)$ is a geodesic metric space with respect to the path metric d_z induced from $(\mathcal{X}, d_X)$. We refer to F_z as the *fiber* over z.

   We further demand that the inclusion maps $i_z : (F_z, d_z) \to \mathcal{X}$ are uniformly metrically proper as measured with respect to h, i.e. for all $z \in \mathcal{B}$ and $u, v \in F_z$, $d_X(i_z(u), i_z(v)) \leq N$ implies that $d_z(u, v) \leq f(N)$.
3. For $z_1, z_2 \in \mathcal{B}$ with $d_B(z_1, z_2) \leq 1$, let γ be a geodesic in $\mathcal{B}$ joining them. Then for any $z \in \gamma$ and $x \in F_z$, there is a path in $p^{-1}(\gamma)$ of length at most c joining x to both F_{z_1} and F_{z_2}.
4. For $z_1, z_2 \in \mathcal{B}$ with $d_B(z_1, z_2) \leq 1$ and $\gamma \subset \mathcal{B}$ a geodesic joining them, let $\phi : X_{z_1} \to X_{z_2}$, be any map such that for all $x_1 \in X_{z_1}$ there is a path of length at most c in $P^{-1}(\gamma)$ joining x_1 to $\phi(x_1)$. Then ϕ is a $K-$quasi-isometry.

If in addition, there exists δ' such that each X_z is $\delta'-$hyperbolic, then $P : \mathcal{X} \to \mathcal{B}$ is called an $(h, c, K)-$ metric bundle of $\delta'-$hyperbolic spaces (or simply a metric bundle of hyperbolic spaces if the constants are implicit).

It is pointed out in [116] that condition (4) follows from the previous three (with suitable K); but it is more convenient to have it as part of our definition.

A closely related notion of a *metric graph bundle* often turns out to be more useful:

Definition 10.17 ([116, Definition 1.2]) Suppose $\mathcal{X}$ and $\mathcal{B}$ are metric graphs and $f : \mathbb{N} \to \mathbb{N}$ is a proper function. We say that $\mathcal{X}$ is an *f-metric graph bundle* over $\mathcal{B}$ if there exists a surjective simplicial map $\pi : \mathcal{X} \to \mathcal{B}$ such that the following hold.

1. For all $b \in V(\mathcal{B})$, $\mathcal{F}_b := \pi^{-1}(b)$ is a connected subgraph of $\mathcal{X}$. Moreover, the inclusion maps $\mathcal{F}_b \to \mathcal{X}$, $b \in V(\mathcal{B})$ are uniformly metrically proper as measured by f.
2. For all adjacent vertices $b_1, b_2 \in V(\mathcal{B})$, any $x_1 \in V(\mathcal{F}_{b_1})$ is connected by an edge to some $x_2 \in V(\mathcal{F}_{b_2})$.

For all $b \in V(B)$, F_b is called the *fiber* over b and its path metric is denoted by d_b. It is pointed out in [116] that any metric bundle is quasi-isometric to a metric graph bundle, where the quasi-isometry coarsely preserves fibers and restricts to a quasi-isometry of fibers. Condition (2) of Definition 10.17 immediately shows that if $\pi : \mathcal{X} \to \mathcal{B}$ is a metric graph bundle then for any points $v, w \in V(\mathcal{B})$ we have that $Hd(F_v, F_w) < \infty$, where Hd denotes the Hausdorff distance.

Example 10.16 *Let*

$$1 \to N \to G \to Q \to 1$$

be an exact sequence of finitely generated groups. Choose a finite generating set of N and extend it to a finite generating set of G. The image of the finite generating set of G in Q under the quotient map is then a generating set of Q. This gives a natural simplicial map $P : \Gamma_G \to \Gamma_Q$ between the respective Cayley graphs. This is the prototypical example of a metric graph bundle. The fibers are all copies of Γ_N.

Definition 10.18 Suppose $\mathcal{X}$ is an *f-metric graph bundle* over $\mathcal{B}$. Given $k \geq 1$ and a connected subgraph $\mathcal{A} \subset \mathcal{B}$, a *$k$-qi section* over $\mathcal{A}$ is a map $s : \mathcal{A} \to \mathcal{X}$ such that s is a k-qi embedding and $\pi \circ s$ is the identity map on $\mathcal{A}$.

For any hyperbolic metric space F with more than two points in its Gromov boundary ∂F, there is a coarsely well-defined *barycenter map*

$$\phi : \partial^3 F \to F$$

mapping an unordered triple (a, b, c) of distinct points in ∂F to a centroid of the ideal triangle spanned by (a, b, c). We shall say that the barycenter map $\phi : \partial^3 F \to F$ is $N-$coarsely surjective if F is contained in the N-neighborhood of the image of ϕ. A *$K-$qi-section* $\sigma : \mathcal{B} \to \mathcal{X}$ is a $K-$qi-embedding from $\mathcal{B}$ to $\mathcal{X}$ such that $P \circ \sigma$ is the identity map. The following guarantees the existence of qi-sections for metric bundles:

Proposition 10.17 ([116, Section 2.1]) *Given $\delta, N, c, K \geq 0$ and proper $f : \mathbb{N} \to \mathbb{N}$, there exists K_0 such that the following holds.*
Let $P : \mathcal{X} \to \mathcal{B}$ be an (f, c, K)-metric bundle of $\delta-$hyperbolic spaces such that all barycenter maps $\phi_b : \partial^3 F_b \to F_b$ are $N-$coarsely surjective, Then through each point of X, there exists a K_0-qi section.

A similar statement holds for metric graph bundles.

The following gives the analog of Definition 10.15 in the context of metric bundles and metric graph bundles:

Definition 10.19 Let $P : \mathcal{X} \to \mathcal{B}$ be a metric bundle or a metric graph bundle. $P : \mathcal{X} \to \mathcal{B}$ is said to satisfy a *flaring condition* if $\forall k \geq 1$, there exist $\lambda_k > 1$ and $n_k, M_k \in \mathbb{N}$ such that the following holds:
Let $\gamma : [-n_k, n_k] \to \mathcal{B}$ be a geodesic and let $\tilde{\gamma_1}$ and $\tilde{\gamma_2}$ be two k-qi sections of γ in X. If $d_{\gamma(0)}(\tilde{\gamma_1}(0), \tilde{\gamma_2}(0)) \geq M_k$, then

$$\lambda_k.d_{\gamma(0)}(\tilde{\gamma_1}(0), \tilde{\gamma_2}(0)) \leq \max\{d_{\gamma(n_k)}(\tilde{\gamma_1}(n_k), \tilde{\gamma_2}(n_k)), d_{\gamma(-n_k)}(\tilde{\gamma_1}(-n_k), \tilde{\gamma_2}(-n_k))\}.$$

The following Theorem is the analog of Theorem 10.15 in the context of metric (graph) bundles.

Theorem 10.18 *Suppose that $P : \mathcal{X} \to \mathcal{B}$ is a metric bundle or a metric graph bundle such that all fibers F_z are uniformly hyperbolic, and the barycenter maps are uniformly coarsely surjective. Equivalently, by Proposition 10.17, there exists $\rho \geq 1$ such that for every $x \in \mathcal{X}$, there exists a $\rho-$qi section $s : \mathcal{B} \to \mathcal{X}$ passing through x, i.e. $s \circ P(x) = x$.*

Then if $\mathcal{X}$ satisfies the qi-embedded condition and the flaring condition (as in Definition 10.19) corresponding to $\rho-$qi sections, then $\mathcal{X}$ is hyperbolic.

Conversely, if X is hyperbolic, then as a metric bundle or metric graph bundle, $\mathcal{X}$ satisfies the flaring condition.

10.5.2.1 Ladders

A tool that has turned out to be considerably useful in the context of both trees of spaces and metric bundles is the notion of a ladder. In particular, for our proof of Theorem 10.18 (sketched in Sect. 10.5.2.2), we shall use it. The notion is related to, but different from that of a hallway. Ladders were introduced in [109] in the context of trees of spaces and in [108] in the context of groups. Instead of going through the construction in detail, we extract the relevant features from the ladder construction of [108, 109]. The following is a restatement of [109, Theorem 3.6] reformulated to emphasize the connection with hallways.

Theorem 10.19 *Given $\delta \geq 0$, $K \geq 1$, $\epsilon \geq 0$ there exists D such that the following holds.*

We consider one of the two following situations:

1. $P : \mathcal{X} \to \mathcal{T}$ *is a tree of* $\delta-$*hyperbolic spaces as in Definition 10.12 with parameters* K, ϵ. *Let* F_v *be a vertex space,*
2. $P : \mathcal{X} \to \mathcal{B}$ *is a metric bundle or metric graph bundle, and* F_v *is a fiber.*

In both cases, the intrinsic metric on F_v *is denoted by* d_v. *Then for every geodesic segment* $\mu \subset (F_v, d_v)$ *there exists a* $D-$*qi-embedded subset* $\mathcal{L}_\mu$ *of* X *such that the following holds.*

1. $F_v \cap \mathcal{L}_\mu = \mu$,
2. *(a) For* $P : \mathcal{X} \to \mathcal{T}$ *a tree of hyperbolic metric spaces and every* $w \in \mathcal{T}$, $F_w \cap \mathcal{L}_\mu$ *is either empty or a geodesic* μ_w *in* (F_w, d_w). *Further, there exists a subtree* $\mathcal{T}_1 \subset \mathcal{T}$ *such that the collection of vertices* $w \in \mathcal{T}$ *satisfying* $F_w \cap \mathcal{L}_\mu \neq \emptyset$ *equals the vertex set of* T_1.
   (b) For $P : \mathcal{X} \to \mathcal{B}$ *a metric bundle or metric graph bundle,* $F_w \cap \mathcal{L}_\mu$ *is a geodesic* μ_w *in* (F_w, d_w).
3. *There exists* $\rho_0 \geq 1$ *such that through every* $z \in \mathcal{L}_\mu$, *there exists a* ρ_0-*qi-section* σ_z *of* $[v, P(z)]$ *contained in* $\mathcal{L}_\mu$ *satisfying*

$$\sigma_z(P(z)) = z, \quad \sigma_z(v) \in \mu.$$

Further, there exists a $D-$*coarse Lipschitz retraction* $\Pi_\mu : \mathcal{X} \to \mathcal{L}_\mu$, *i.e.*

1. $d(\Pi_\mu(x), \Pi_\mu(y)) \leq Dd(x, y) + D, \ \forall x, y \in X$,
2. $\Pi_\mu(x) = x, \ \forall x \in \mathcal{L}_\mu$.

The qi-embedded set $\mathcal{L}_\mu$ is called a *ladder* in [108, 109]. Theorem 10.19 shows in particular that there is a $(2D, 2D)-$ quasigeodesic of (X, d_X) joining the end-points of μ and lying on $\mathcal{L}_\mu$.

Remark 10.20 *Note that in Theorem 10.19, we have* not *assumed that* X *is hyperbolic: no assumptions on the global geometry of* X *are necessary here.*

10.5.2.2 Idea Behind the Proof of Theorem 10.18

We focus on the metric graph bundle case for convenience. Theorem 10.19 guarantees that for any pair of points x, y in a metric graph bundle $\mathcal{X}$, there exist

1. Qi-sections Σ_x, Σ_y through x, y.
2. A ladder $\mathcal{L}(x, y)$ bounded by Σ_x, Σ_y. In fact, in this case (as shown in [108, 116]), $\mathcal{L}(x, y) \cap F_b$ is equal to a geodesic in F_b joining $\Sigma_x(b), \Sigma_y(b)$ (here we are abusing notation slightly by identifying the qi-sections Σ_x, Σ_y with their images).

Thus, for every $x, y \in \mathcal{X}$ there are preferred quasigeodesics in $\mathcal{X}$ contained in $\mathcal{L}(x, y)$. We have not used the flaring condition so far. The flaring condition guarantees hyperbolicity of $\mathcal{L}(x, y)$. We shall return to this shortly. Hyperbolicity of $\mathcal{L}(x, y)$ ensures (by the Morse Lemma) that all quasigeodesics in $\mathcal{L}(x, y)$ joining x, y are in a bounded neighborhood of each other. This gives a family of paths in $\mathcal{X}$, one for every pair x, y. We then use a path-families argument following Hamenstädt [56] and a criterion due to Bowditch to conclude that $\mathcal{X}$ is hyperbolic.

We return to the proof of hyperbolicity of $\mathcal{L}(x, y)$. We note that $\mathcal{L}(x, y)$ is a bundle over $\mathcal{B}$ where the fibers are intervals. The flaring condition is inherited by $\mathcal{L}(x, y)$ with slightly worse constants. Thus, we are reduced to proving Theorem 10.18 in the special case that fibers are intervals. To do this, we decompose the ladder $\mathcal{L}(x, y)$ using qi-sections contained in $\mathcal{L}(x, y)$ into a finite number of ladders 'stacked one on top of another'. Thus, there exist disjoint sections $\Sigma_x = \Sigma_0, \Sigma_1, \cdots, \Sigma_n = \Sigma_y$ and ladders $\mathcal{L}_i$ bounded by Σ_{i-1}, Σ_i such that distinct $\mathcal{L}_i$'s have disjoint interiors. The ubiquity of qi-sections allows us to ensure that each of these smaller ladders has bounded girth (in the spirit of Definition 10.14), i.e. Σ_{i-1}, Σ_i are at a bounded distance from each other along some fiber F_b and flare away from each other as one goes to infinity in B. A further path families argument following [56] allows us to prove that $\mathcal{L}(x, y)$ is hyperbolic.

A word about the proof sketch above. Note that we use only the 1-dimensional property of quasigeodesics flaring and path families to prove the combination theorem in this case, as opposed to the more '2-dimensional' area argument of [12]. This has been considerably refined in [71] to give a new path-families proof of Theorem 10.15.

10.5.3 Relatively Hyperbolic Combination Theorems

We refer the reader to [15, 38, 49] for the basics of relative hyperbolicity. Theorem 10.15 was generalized to the context of trees of relatively hyperbolic spaces in two different ways:

1. Using an acylindricity hypothesis in [28] and [4]. This is in the spirit of Theorem 10.10.
2. Using the flaring condition in [45, 115]. This in the spirit of Theorem 10.14.

10.5.3.1 Relatively Hyperbolic Combination Theorem Using Acylindricity

Let G be hyperbolic relative to a finite collection $\mathcal{P} = \{P_1, \cdots, P_k\}$ of parabolic subgroups. Let $\partial_h G$ denote the Bowditch boundary of G. Let $H \subset G$ be a relatively quasiconvex subgroup [58]. We shall give Dahmani's version of the combination theorem [28] below. Let $\Lambda_H \subset \partial_h G$ denote the limit set of H. A relatively quasiconvex subgroups H is *full relatively quasi-convex* if it is quasi-convex and if, for any infinite sequence $g_n \in G$ in distinct left cosets of H, the intersection $\cap_n g_n(\Lambda_H)$ is empty.

Lemma 10.1 ([28, Lemma 1.7]) *Let G be hyperbolic relative to a finite collection $\mathcal{P} = \{P_1, \cdots, P_k\}$ of parabolic subgroups. Let H be a full relatively quasi-convex subgroup. Let P be a conjugate of one of the P_i's. Then $P \cap H$ is either finite, or of finite index in P.*

Definition 10.20 ([134]) The action of a group G on a tree T is $k-$acylindrical for some $k \in \mathbb{N}$ if the stabilizer of any geodesic of length k in T is finite. The action of a group G on a tree T is acylindrical if it is $k-$acylindrical for some $k \in \mathbb{N}$.

A finite graph of groups is said to be acylindrical, if the action on the associated Bass–Serre tree is acylindrical.

Then Dahmani's combination theorem states:

Theorem 10.21 ([28]) *Let G be the fundamental group of an acylindrical finite graph of relatively hyperbolic groups, whose edge groups are full quasi-convex subgroups of the adjacent vertex groups. Let $\mathcal{G}$ be the family of images of the maximal parabolic subgroups of the vertex groups, and their conjugates in G. Then G is strongly hyperbolic relative to $\mathcal{G}$.*

The approach in [28] is quite different from that of [12]. From the Bowditch boundaries of the vertex and edge groups, a metrizable compact space Z is constructed in such a way that G naturally acts on Z. It is then shown that this action is a convergence action. Finally, it is shown that the action is geometrically finite, forcing G to have a relatively hyperbolic structure.

The Bass–Serre tree of G has vertex groups G_v and edge groups G_e. Hence, associated to the Bass–Serre tree T there is a natural tree (T) of compact spaces given by $\partial_h G_v$ and $\partial_h G_e$. The set Z is built [28, Section 2] from these copies of $\partial_h G_v$ and $\partial_h G_e$. Suppose $e = [v_1, v_2]$ is an edge of T. For all such edges e, glue together $\partial_h G_{v_1}$ and $\partial_h G_{v_2}$ along the limit set $\partial_h G_e$. The relevant identification space is thus obtained from the set $\sqcup_{v \in V(T)} \partial_h G_v$ by identifying pairs of points according to the images of $\partial_h G_e$. Finally, the base tree T encodes (infinite) directions that are 'transverse' to all the vertex spaces. The set Z is then obtained from topologizing $\partial T \cup \sqcup_{v \in V(T)} \partial_h G_v / \sim$, where $\sim$ is the equivalence relation given by edge spaces.

Alibegovic [4] proves a similar combination theorem for relatively hyperbolic groups following the original strategy of Bestvina and Feighn in Theorem 10.15 using the linear isoperimetric inequality characterization of hyperbolicity.

10.5.3.2 Relatively Hyperbolic Combination Theorem Using Flaring

We next define a tree of relatively hyperbolic spaces in general.

Definition 10.21 ([115]) A tree $P : \mathcal{X} \to \mathcal{T}$ of geodesic metric spaces is said to be a tree of relatively hyperbolic metric spaces if in addition to the conditions of Definition 10.12

1. each vertex space X_v is strongly hyperbolic relative to a collection of subsets $\mathcal{H}_v$ and each edge space X_e is strongly hyperbolic relative to a collection of subsets $\mathcal{H}_e$. The sets $H_{v,\alpha} \in \mathcal{H}_v$ or $H_{e,\alpha} \in \mathcal{H}_e$ are referred to as *horosphere-like sets*.
2. the maps f_{e,v_i} above, for $i = 1, 2$, are *strictly type-preserving*. That is, for $i = 1, 2$ and for any $H_{v_i,\alpha} \in \mathcal{H}_{v_i}$, $f_{e,v_i}^{-1}(H_{v_i,\alpha})$, is either empty or is equal to some $H_{e,\beta} \in \mathcal{H}_e$. Further, for all $H_{e,\beta} \in \mathcal{H}_e$, there exists v and $H_{v,\alpha}$, such that $f_{e,v}(H_{e,\beta}) \subset H_{v,\alpha}$.
3. There exists $\delta > 0$ such that each $\mathcal{E}(X_v, \mathcal{H}_v)$ is δ-hyperbolic (here, $\mathcal{E}(X_v, \mathcal{H}_v)$ denotes the electric space obtained from X_v by electrifying all the horosphere-like sets in $\mathcal{H}_v$).
4. The induced maps of the coned-off edge spaces into the coned-off vertex spaces $\widehat{f_{e,v_i}} : \mathcal{E}(X_e, \mathcal{H}_e) \to \mathcal{E}(X_{v_i}, \mathcal{H}_{v_i})$ $(i = 1, 2)$ are uniform quasi-isometries. This is called the *qi-preserving electrification condition*

We state conditions (4) and (6) in conjunction by saying that X_v is *strongly δ–hyperbolic* relative to $\mathcal{H}_v$.

We explain condition (7) briefly. Given the tree of spaces $P : \mathcal{X} \to \mathcal{T}$ with vertex spaces X_v and edge spaces X_e there exists a naturally associated tree whose vertex spaces are the electrified spaces $\mathcal{E}(X_v, \mathcal{H}_v)$ and edge spaces are the electrified spaces $\mathcal{E}(X_e, \mathcal{H}_e)$ obtained by electrifying the respective horosphere like sets. Condition (4) of the above definition ensures that we have natural inclusion maps of edge spaces $\mathcal{E}(X_e, \mathcal{H}_e)$ into adjacent vertex spaces $\mathcal{E}(X_v, \mathcal{H}_v)$. The resulting tree of coned-off spaces $P : \mathcal{TC}(X) \to \mathcal{T}$ is referred to simply as the *induced tree of coned-off spaces*. The *cone locus* of $\mathcal{TC}(X)$ is the forest given by the following:

1. the vertex set $\mathcal{V}(\mathcal{TC}(X))$ consists of the cone-points $c_{v,\alpha}$ in the vertex spaces X_v resulting from the electrification operation of the horosphere-like sets $H_{v,\alpha} \in \mathcal{H}_v$.
2. the edge set $\mathcal{E}(\mathcal{TC}(X))$ consists of the cone-points $c_{e,\alpha}$ in the edge set X_e resulting from the electrification operation of the horosphere-like sets $H_{e,\alpha} \in \mathcal{H}_e$.

Each connected component of the cone-locus is a *maximal cone-subtree*. The collection of maximal cone-subtrees is denoted by $C\mathcal{T}$ and elements of $C\mathcal{T}$ are denoted as CT_α. Note that each maximal cone-subtree CT_α naturally gives rise to a tree CT_α of horosphere-like subsets depending on which cone-points arise as vertices and edges of CT_α. The metric space that CT_α gives rise to is denoted as C_α. We refer to any such C_α as a *maximal cone-subtree of horosphere-like spaces*. The induced tree of horosphere-like sets is denoted by

$$g_\alpha : C_\alpha \to CT_\alpha.$$

The collection of these maps will be denoted as $\mathcal{G}$. The collection of the maximal cone-subtree of horosphere-like spaces C_α is denoted as C. Note that each CT_α thus appears both as a subset of $\mathcal{TC}(X)$ as well as the underlying tree of C_α.

Definition 10.22 (Cone-Bounded Hallways Strictly Flare Condition)
An essential hallway of length $2k$ is *cone-bounded* if $f(i \times \partial I)$ lies in the cone-locus for $i = \{-k, \cdots, k\}$.

The tree of spaces, X, is said to satisfy the *cone-bounded hallways flare* condition if there are numbers $\lambda > 1$ and $k \geq 1$ such that any cone-bounded hallway of length $2k$ is λ-hyperbolic, where λ, k are called the constants or parameters of the strict flare condition.

We now state the combination theorem for relative hyperbolicity using the flaring condition.

Theorem 10.22 ([45, 115]) *Let $P : \mathcal{X} \to \mathcal{T}$ be a tree of uniformly relatively hyperbolic spaces in the sense of Definition 10.21 satisfying the qi-embedded condition, such that the resulting tree of coned-off spaces satisfies*

1. *the hallways flare condition,*
2. *the cone-bounded hallways flare condition.*

Then X is hyperbolic relative to the maximal cone-subtrees of horosphere-like spaces.

10.5.4 Effective Quasiconvexity and Flaring

We now state a couple of theorems along the lines of Theorems 10.15 and 10.18 ensuring quasiconvexity of a subspace of a vertex space. The first, due to Ilya Kapovich [69] is in the setup of an acylindrical graph of groups:

Theorem 10.23 *Let $\mathcal{G}$ be a finite acylindrical graph of groups where all vertex and edge groups are hyperbolic and edge-to-vertex inclusions are quasi-isometric embeddings. Let $\mathcal{T} \subset \mathcal{G}$ be a maximal subtree. Let G denote the group corresponding to the tree $\mathcal{T}$. (By Theorem 10.15, G is hyperbolic.) Then each vertex group G_v of G is quasiconvex in G.*

Next, we shall consider in a unified way the two following situations:

1. $P : \mathcal{X} \to \mathcal{T}$ is a tree of hyperbolic metric spaces satisfying the qi-embedded condition with constants K, ϵ and the hallways flare condition with parameters λ_0, m_0. Further, if ρ_0 is given we shall assume an additional constant H_0 as a lower bound for girths of ρ_0-thin hallways.
2. $P : \mathcal{X} \to \mathcal{B}$ is a metric bundle or a metric graph bundle satisfying the flaring condition with constants as in Definitions 10.16, 10.17, and 10.19.

Also (X_v, d_v) will be a vertex space of $\mathcal{X}$ (in the tree of spaces case) or $P^{-1}(v)$ equipped with the induced metric in the metric (graph) bundle case.

Definition 10.23 Let $P : \mathcal{X} \to \mathcal{T}$ be a tree of hyperbolic spaces. Let $Y \subset (X_v, d_v)$ be a $C-$quasiconvex subset of (X_v, d_v). We say that *Y flares in all directions with* *parameter* K if for any geodesic segment $[a, b] \subset (X_v, d_v)$ with $a, b \in Y$ and any $\rho-$thin hallway $f : [0, k] \times I \to \mathcal{X}$ satisfying

1. $\rho \leq \rho_0$,
2. $f(\{0\} \times I) = [a, b]$,
3. $l([a, b]) \geq K$,
4. $k \geq K$,

the length of $f(\{k\} \times I)$ satisfies

$$l(f(\{k\} \times I)) \geq \lambda l([a, b]).$$

Similarly, let $P : \mathcal{X} \to \mathcal{B}$ be a metric bundle or metric graph bundle with hyperbolic fiber. Let $Y \subset X_v$ be quasiconvex. Further, assume that there is a $\rho-$qi section through every $x \in \mathcal{X}$ (cf. the second hypothesis of Theorem 10.18).

We say that *Y flares in all directions with parameter* $K \geq 0$, $D \geq 1$, $\lambda > 1$ if the following holds:

Let $\gamma : [0, D] \to \mathcal{B}$ be a geodesic such that $\gamma(0) = v$ and let $\tilde{\gamma_1}$ and $\tilde{\gamma_2}$ be two ρ-qi lifts (sections) of γ in $\mathcal{X}$. If $d_v(\tilde{\gamma_1}(0), \tilde{\gamma_2}(0)) \geq K$, then we have

$$\lambda . d_v(\tilde{\gamma_1}(0), \tilde{\gamma_2}(0)) \leq d_{\gamma(D)}(\tilde{\gamma_1}(D), \tilde{\gamma_2}(D)).$$

We can now state a proposition guaranteeing quasiconvexity of subsets of vertex spaces.

Proposition 10.24 ([113]) *Given* K, C, *there exists* C_0 *such that the following holds.*
Let $P : \mathcal{X} \to \mathcal{T}$ *and* X_v *be as in Theorem 10.19 above. If* Y *is a* $C-$*quasiconvex subset of* (X_v, d_v) *and flares in all directions with parameter* K, *then* Y *is* C_0-*quasiconvex in* (X, d_X).

Conversely, given C_0, *there exist* K, C *such that the following holds.*
For $P : \mathcal{X} \to \mathcal{T}$ *and* X_v *as above, if* $Y \subset X_v$ *is* C_0-*quasiconvex in* (X, d_X), *then it is a* $C-$*quasiconvex subset in* (X_v, d_v) *and flares in all directions with parameter* K.

A similar statement holds for metric (graph) bundles.

Proposition 10.25 *Given* K, D, λ, C, *there exists* C_0 *such that the following holds.*
Let $P : \mathcal{X} \to \mathcal{B}$ *be a metric (graph) bundle and* X_v *be as in Definition 10.23. If* Y *is a* $C-$*quasiconvex subset of* (X_v, d_v) *and flares in all directions with parameters* K, D, λ, *then* Y *is* C_0-*quasiconvex in* $(\mathcal{X}, d_X)$.

Conversely, given C_0, *there exist* K, D, λ, C *such that the following holds.*
For $P : \mathcal{X} \to \mathcal{B}$ *a metric (graph) bundle and* X_v *as above, if* $Y \subset X_v$ *is* C_0-*quasiconvex in* (X, d_X), *then it is a* $C-$*quasiconvex subset in* (X_v, d_v) *and flares in all directions with parameters* K, D, λ.

10.6 Combination Theorems in Geometric Group Theory: Cubulations

We turn now to the remarkable work during the last decade on special cube complexes. We refer to [53] for the basics of special cube complexes. Let $\mathcal{G}$ denote a finite graph, $RAAG(\mathcal{G})$ the right-angled Artin group associated to $\mathcal{G}$, and $\mathcal{S}(\mathcal{G})$ its Salvetti complex. A cube complex C is said to be *special* if there exists a combinatorial local isometry from C to $\mathcal{S}(\mathcal{G})$ for some finite graph $\mathcal{G}$. By Agol's resolution of Wise's conjecture in [1] (see Theorem 10.30 below), hyperbolic groups that are virtually special are precisely those that act geometrically on a CAT(0) cube complex. We give a brief account of some of the combination theorems that have been proved around this theme.

In [59], Hsu and Wise proved the precursor of all virtually special combination theorems by showing that if a hyperbolic group G splits as a finite graph of finitely generated free groups with cyclic edge groups, then G is virtually special. In [60], they later generalized this to amalgamated products of free groups over a finitely generated malnormal subgroup. A landmark combination theorem due to Haglund and Wise concerns the combination of hyperbolic virtually special cubulable groups along malnormal quasiconvex subgroups:

Theorem 10.26 ([54]) *Let A, B, M be compact virtually special cube complexes. Suppose that $G_A = \pi_1(A)$, $G_B = \pi_1(B)$, and $G_M = \pi_1(M)$ are hyperbolic. Let $M \xrightarrow{i_A} A$, and $M \xrightarrow{i_B} B$ be local isometries of cube complexes such that $i_{A*}(G_M)$ and $i_{B*}(G_M)$ are quasiconvex and malnormal in G_A and G_B respectively. Let $X = A \cup_M B$ be the cube complex obtained by gluing A and B together along M using $M \times [0, 1]$. Then X is virtually special.*

Theorem 10.26 generalizes earlier work of Wise [147] where he showed that any 2-complex built by amalgamating (in terms of fundamental group) two finite graphs along a malnormal immersed graph is virtually special. Theorem 10.26 is also a crucial ingredient in Wise's proof of the virtual specialness of hyperbolic groups admitting a quasiconvex hierarchy. This is very much in the spirit of the Haken hierarchy for Haken 3-manifolds and Thurston's hyperbolization of such manifolds (cf. Sect. 10.4.1).

Theorem 10.27 ([149]) *Let G be a hyperbolic group admitting a quasiconvex hierarchy. Then G is the fundamental group of a compact non-positively curved cube complex that is virtually special.*

We list below some of the important consequences of Theorem 10.27. The following resolved a conjecture of Baumslag:

Theorem 10.28 ([149]) *Every one-relator group with torsion is virtually special.*

In the context of hyperbolic 3-manifolds, Wise showed the following.

Theorem 10.29 ([149]) *Compact hyperbolic Haken manifolds are virtually special.*

Using work of Kahn and Markovic [68], Bergeron and Wise [8] proved that all hyperbolic 3-manifolds can be cubulated, i.e. they act geometrically on CAT(0) cube complexes. This led Wise to conjecture that hyperbolic groups that act geometrically on CAT(0) cube complexes are virtually special. The following celebrated theorem of Agol resolved this conjecture affirmatively:

Theorem 10.30 ([1]) *Hyperbolic groups acting geometrically on CAT(0) cube complexes are virtually special.*

A flurry of activity ensued in trying to show that several naturally defined hyperbolic groups are, in fact, cubulable. In [51, 52], Hagen and Wise proved that hyperbolic groups G admitting an exact sequence of the form

$$1 \to F_n \to G \to \mathbb{Z} \to 1$$

are cubulable. (Here F_n denotes the free group on n generators.) Hence, by Agol's Theorem 10.30, such groups G are virtually special. In a different direction, Manning, the first author and Sageev [92] showed that there exist cubulable hyperbolic groups G admitting an exact sequence of the form

$$1 \to \pi_1(S) \to G \to F_n \to 1,$$

where S is a closed surface of genus greater than one. Again, by Agol's Theorem 10.30, such groups G are virtually special.

Finally, we mention work of Przytycki and Wise [126], who proved the virtual specialness of fundamental groups of 3-manifolds whose JSJ decomposition has both a hyperbolic as well as a Seifert-fibered piece. 3-manifolds admitting such a JSJ decomposition are called *mixed*. As a consequence of their result, the authors show that mixed manifolds virtually fiber.

The proof in [126] proceeds by first showing that there are enough codimension one surface subgroups to ensure cubulability. This is established by combining surfaces coming from the graph manifold pieces with those coming from hyperbolic pieces. Once cubulability has been established, the malnormal special quotient theorem [149] (see also [2]) is used to establish specialness of the cube complex thus built.

In this section, we have given only a cursory treatment of a topic that, starting with [131] has undergone tremendous development over the last two decades. We refer the reader to the books [5] and [148] for a more comprehensive treatment.

10.7 Holomorphic Dynamics and Polynomial Mating

Before entering the theme of combination theorems in holomorphic dynamics, we say a few words on the history of the subject, and sketch briefly some of the philosophical parallels between holomorphic dynamics and Kleinian groups and some of the developments inspired by this synergy. These will also serve as a motivation for combination theorems involving complex polynomials and Kleinian groups that we will discuss later in the section.

10.7.1 Historical Comments

The study of dynamics of rational maps on the Riemann sphere started with groundbreaking work of Fatou and Julia [40–44, 65, 66] in the 1920s. The subject remained dormant for several decades barring a handful of important contributions, most notably by Siegel [138] and Brolin [18]. Around the 1970s, the availability of computers allowed Feigenbaum and Mandelbrot to perform numerical experiments on finer structures of dynamical and parameter spaces of real/complex-analytic maps. Their pioneering discoveries infused fresh blood into the field, and gave rise to problems and conjectures that played pivotal roles in the development of the modern theory of holomorphic dynamics.

A revolutionary contribution came from Sullivan, who introduced quasiconformal methods into the study of rational dynamics to prove nonexistence of wandering domains in the Fatou set for rational maps [140]. The seminal work of Douady and Hubbard on the dynamics of quadratic polynomials and the structure of the Mandelbrot set [31, 32] turned out to be equally fundamental in that the techniques devised by them were robust enough to be applied to the study of a wide variety of holomorphic dynamical systems.

Sullivan proposed a dictionary between Kleinian groups and rational dynamics that was motivated by various common features shared by them [140, p. 405]. In addition to the apparent similarities between the topological structures of the *limit set* (respectively, the *domain of discontinuity*) of a Kleinian group and the *Julia set* (respectively, the *Fatou set*) of a rational map, there are deeper similarities between the techniques employed in proving various statements in these two parallel worlds. In fact, in the same paper, Sullivan gave a new proof of Ahlfors' finiteness theorem which closely parallels the proof of the 'no wandering Fatou component' theorem for rational maps.

Around the same time, Thurston proved a topological characterization for an important class of rational maps [33]. This result, which is a philosophical analog of the hyperbolization of atoroidal Haken 3-manifolds, has given rise to a wealth of rich and beautiful results that we will not be touching upon in this survey (see [129, §9] and the references therein).

We should emphasize that the aforementioned dictionary is not an automatic method for translating results in one setting to the other, but rather an inspiration for results and proof techniques. We now list a few prominent pieces of work motivated by this dictionary. Sullivan and McMullen introduced Teichmüller spaces of conformal dynamical systems in the spirit of Teichmüller spaces of Riemann surfaces in [141]. Bullett and Penrose constructed matings of holomorphic quadratic polynomials and the modular group as *holomorphic correspondences* [21]. In [90], Lyubich and Minsky constructed "an explicit object that plays for a rational map the role played by the hyperbolic 3-orbifold quotient of a Kleinian group".McMullen established conceptual connections between renormalization ideas used in holomorphic dynamics and the study of 3-manifolds fibering over the circle [102]. Pilgrim proved a canonical decomposition theorem for Thurston maps as an analog of the torus decomposition theorem for 3-manifolds [124]. Another noteworthy development in the framework of the above dictionary is the recent work of Luo [86–88], where results in rational dynamics were proved using techniques that are closely related to Thurston's work on 3-manifolds.

We refer the reader to [23, 107] for a basic introduction to rational dynamics, to [89] for a comprehensive account on the dynamics of quadratic polynomials and the Mandelbrot set, to the recent survey article by Rees on major advances in the field [129], and a survey by DeZotti on connections between holomorphic dynamics and other branches of mathematics [30].

10.7.2 Mating of Polynomials

The operation of polynomial mating, which was introduced by Douady and Hubbard in [35], constructs a rational map on $\widehat{\mathbb{C}}$ by combining the actions of two complex polynomials. Since the first appearance of the notion, several closely related definitions and perspectives have been put forward. In this survey, we will follow the route adopted in [123] (see [130] for the original formulation and some historical comments, and [106] for a lucid account of the mating construction along with a detailed worked out example).

To define the operation of polynomial mating formally, we need to introduce some terminology. The *Fatou set* of a rational map R, denoted by $\mathcal{F}(R)$, is the largest open subset of $\widehat{\mathbb{C}}$ on which the sequence of iterates $\{R^{\circ n}\}_{n\geq 0}$ forms a normal family. Its complement is called the *Julia set*, and is denoted by $\mathcal{J}(R)$. For a complex polynomial P, the *filled Julia set* (i.e., the set of points with bounded forward orbits) and the *basin of attraction of infinity* (i.e., the complement of the filled Julia set) are denoted by $\mathcal{K}(P)$ and $\mathcal{B}_\infty(P)$, respectively. We refer the reader to [107] for basic topological and dynamical properties of these sets.

A rational map R is called *postcritically finite* if each of its critical points has a finite forward orbit. R is called *hyperbolic* if each of its critical points converges to an attracting cycle under forward iteration.

If P is a monic, centered polynomial of degree d with a connected Julia set, then there exists a conformal map $\phi_P : \widehat{\mathbb{C}} \setminus \overline{\mathbb{D}} \to \mathcal{B}_\infty(P)$ that conjugates z^d to P, and satisfies $\phi_P'(\infty) = 1$ [107, Theorem 9.1, Theorem 9.5]. We will call ϕ_P the *Böttcher coordinate* for P. Furthermore, if $\partial\mathcal{K}(P) = \mathcal{J}(P)$ is locally connected, then ϕ_P extends to a semiconjugacy between $z^d|_{\mathbb{S}^1}$ and $P|_{\mathcal{J}(P)}$. In this case, the map $\phi_P : \mathbb{S}^1 \to \mathcal{J}(P)$ is called the *Carathéodory loop/semi-conjugacy* for $\mathcal{J}(P)$.

Now let P_1, P_2 be two monic polynomials of the same degree $d \geq 2$ with connected and locally connected filled Julia sets. We consider the disjoint union $\mathcal{K}(P_1) \sqcup \mathcal{K}(P_2)$ and the map

$$P_1 \sqcup P_2 : \mathcal{K}(P_1) \sqcup \mathcal{K}(P_2) \to \mathcal{K}(P_1) \sqcup \mathcal{K}(P_2),$$
$$P_1 \sqcup P_2|_{\mathcal{K}(P_1)} = P_1, \quad P_1 \sqcup P_2|_{\mathcal{K}(P_2)} = P_2.$$

Let $\sim$ be the equivalence relation on $\mathcal{K}(P_1) \sqcup \mathcal{K}(P_2)$ generated by $\phi_{P_1}(z) \sim \phi_{P_2}(\overline{z})$, for all $z \in \mathbb{S}^1$. It is easy to check that $\sim$ is $P_1 \sqcup P_2-$invariant, and hence it descends to a continuous map $P_1 \perp\!\!\!\perp P_2$ to the quotient $\mathcal{K}(P_1) \perp\!\!\!\perp \mathcal{K}(P_2) := (\mathcal{K}(P_1) \sqcup \mathcal{K}(P_2)) / \sim$ (see [123, §4.1] for details). The map $P_1 \perp\!\!\!\perp P_2$ is called the *topological mating* of the polynomials P_1, P_2. Moreover, if $\mathcal{K}(P_1) \perp\!\!\!\perp \mathcal{K}(P_2)$ is homeomorphic to a 2-sphere, we say that the topological mating is *Moore-unobstructed*. We refer the reader to [123, Theorem 2.12] for the statement of Moore's theorem, which provides a general sufficient condition for the quotient of $\mathbb{S}^2$ under an equivalence relation to be a topological 2−sphere, and to [123, Proposition 4.12] for a useful application of Moore's theorem giving a sufficient condition for the topological mating of P_1, P_2 (as above) to be Moore-unobstructed (note that the conditions of Moore's theorem are not necessary, see [13, Example 13.18]). By [123, Proposition 4.3], if the topological mating of P_1, P_2 is not Moore obstructed (i.e., if $\mathcal{K}(P_1) \perp\!\!\!\perp \mathcal{K}(P_2) \cong \mathbb{S}^2$), then $P_1 \perp\!\!\!\perp P_2$ is topologically conjugate to an orientation-preserving branched covering of $\mathbb{S}^2$. The following definition relates the topological mating to rational maps of $\widehat{\mathbb{C}}$. We refer the reader to [33] for the notion of *Thurston equivalence* appearing below.

Definition 10.24 ([123, Definition 4.4]) Let the topological mating of $P_1 \perp\!\!\!\perp P_2$ be Moore-unobstructed, and $h : \mathcal{K}(P_1) \perp\!\!\!\perp \mathcal{K}(P_2) \to \mathbb{S}^2$ be a homeomorphism.

1. The polynomials P_1, P_2 are called *combinatorially mateable* if they are postcritically finite and if the branched covering $h \circ P_1 \perp\!\!\!\perp P_2 \circ h^{-1} : \mathbb{S}^2 \to \mathbb{S}^2$ is Thurston equivalent to a rational map R.
2. The polynomials P_1, P_2 are called *conformally/geometrically mateable* if the homeomorphism h can be so chosen that $R = h \circ P_1 \perp\!\!\!\perp P_2 \circ h^{-1} : \mathbb{S}^2 \to \mathbb{S}^2$ is a rational map and h is conformal on the interior of $\mathcal{K}(P_1) \perp\!\!\!\perp \mathcal{K}(P_2)$.

Conversely, a rational map R is said to be combinatorially (respectively, conformally) a mating if there exist polynomials P_1, P_2 satisfying the corresponding property above with $R = h \circ P_1 \perp\!\!\!\perp P_2 \circ h^{-1}$.

The following equivalent definition of conformal mating is often useful in practice (see [123, §4.7] for other definitions). In fact, this definition can be easily adapted for the other frameworks of combination theorems that we will discuss in this section (compare Definition 10.30).

Definition 10.25 ([123, Definition 4.14]) A rational map $R : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ of degree $d \geq 2$ is said to be the *conformal mating* of two degree d monic, centered, polynomials P_1 and P_2 with connected and locally connected filled Julia sets if and only if there exist continuous maps

$$\psi_1 : \mathcal{K}(P_1) \to \widehat{\mathbb{C}} \text{ and } \psi_2 : \mathcal{K}(P_2) \to \widehat{\mathbb{C}},$$

conformal on int $\mathcal{K}(P_1)$, int $\mathcal{K}(P_2)$, respectively, such that

1. $\psi_1(\mathcal{K}(P_1)) \bigcup \psi_2(\mathcal{K}(P_2)) = \widehat{\mathbb{C}}$,
2. $\psi_i \circ P_i = R \circ \psi_i$, for $i \in \{1, 2\}$, and
3. $\psi_1(z) = \psi_2(w)$ if and only if $z \sim w$, where $\sim$ is the equivalence relation defined above.

With the above notions of mating in place, we can now mention the first major results on mateability of complex polynomials. In fact, these provided the first main application of Thurston's theorem on the topological characterization for rational maps. The following theorem completely answers the question of conformal mateability of postcritically finite quadratic polynomials (see [31, 32] for a detailed study of the Mandelbrot set, or [105] for a quick introduction).

Theorem 10.31 ([81, 127, 136]) *Let $P_1(z) = z^2 + c_1$ and $P_2(z) = z^2 + c_2$ be two postcritically finite quadratic polynomials. Then P_1 and P_2 are conformally mateable if and only if c_1 and c_2 do not belong to conjugate limbs of the Mandelbrot set.*

Among other early works on matings of postcritically finite polynomials, we ought to mention the work of Shishikura and Lei which highlighted additional complexities that are absent in the quadratic setting, but arise for cubic rational maps [137].

In [82], Lei described the dynamics of postcritically finite cubic Newton maps (these maps, which are obtained by plugging in complex polynomials in Newton's classical root-finding method, form an important and well-studied class of rational maps), and exhibited in the process the fact that a large subclass of such maps are matings. (See also the more recent work [6] for a description of certain postcritically infinite cubic Newton maps as matings.)

The next theorem, due to Yampolsky and Zakeri, was the first existence result for conformal matings of polynomials that are not 'close cousins' of postcritically finite ones. A quadratic polynomial P is said to have a *bounded type Siegel fixed point* if it has a fixed point z_0 with $P'(z_0) = e^{2\pi i\theta}$ such that the continued fraction expansion of $\theta \in \mathbb{R}/\mathbb{Z}$ has uniformly bounded partial fractions.

Theorem 10.32 ([150]) *Suppose P_1, P_2 are quadratic polynomials which are not anti-holomorphically conjugate and each of which has a bounded type Siegel fixed point. Then P_1 and P_2 are conformally mateable.*

The question of unmating a rational map; i.e., deciding whether a given rational map appears as the mating of two polynomials (and if so, whether such a decomposition is unique) has also been studied by various authors. For a general combinatorial characterization of hyperbolic, postcritically finite rational maps arising as matings, see [104, Theorem 4.2]. The situation is a bit more subtle for postcritically finite, non-hyperbolic rational maps, as discussed in the same paper. However, the next theorem gives a positive answer to the unmating question for a class of postcritically finite, non-hyperbolic rational maps:

Theorem 10.33 ([103, Theorem 1.1]) *Let $R : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be a postcritically finite rational map such that its Julia set is the whole sphere. Then every sufficiently high iterate $R^{\circ n}$ of R arises as a mating (i.e., is topologically conjugate to the topological mating of two polynomials).*

We refer the reader to the excellent survey article [104] for more on this topic.

Matings of geometrically finite polynomials (i.e., a polynomial whose postcritical set intersects the Julia set in a finite set, or equivalently, if every critical point is either preperiodic, or attracted to an attracting or parabolic cycle) were studied by Haïssinsky and Lei using techniques of David homeomorphisms. They showed that two geometrically finite polynomials P_1 and P_2 with connected Julia sets and parabolic periodic points are mateable if and only if the postcritically finite polynomials $\mathcal{T}(P_1), \mathcal{T}(P_2)$ canonically associated to P_1, P_2 (such that $\mathcal{T}(P_i)$ and P_i have topologically conjugate Julia set dynamics, $i = 1, 2$) are mateable [55, Theorem D] (cf. [91, Theorem 5.2]).

Mating of Anti-holomorphic Polynomials Let us now mention a class of anti-holomorphic polynomials (anti-polynomials for short) for which a complete solution to the conformal mating problem is known. These are the so-called *critically fixed* anti-polynomials; i.e., anti-polynomials that fix all of their critical points. The proof of the following theorem crucially uses [125, Theorem 3.2], which in many situations, facilitates the application of Thurston's topological characterization of rational maps.

Theorem 10.34 ([85, Theorem 1.3]) *Let P_1 and P_2 be two (marked) anti-polynomials of equal degree $d \geq 2$, where P_1 is critically fixed and P_2 is postcritically finite, hyperbolic. Then there is an anti-rational map R that is the conformal mating of P_1 and P_2 if and only if there is no Moore obstruction.*

In the opposite direction, the question of unmating critically fixed anti-rational maps was also settled in [85, Theorem 1.2], and examples of *shared matings* were demonstrated (cf. [128]).

Remark 10.35

(1) Combined with [85, Lemma 4.22], Theorem 10.34 yields an effective procedure to decide conformal mateability of a critically fixed anti-polynomial P_1 and a postcritically finite, hyperbolic anti-polynomial P_2. This is particularly useful in applying Theorem 10.45 below to concrete examples.

(2) It is worth mentioning that the above mating (respectively, unmating) results for critically fixed anti-polynomials (respectively, anti-rational maps) serve as a precise philosophical counterpart of the double limit theorem for (geometrically finite) Kleinian reflection groups in the complex dynamics world (see the discussion before [85, Theorem 1.3] and [85, §4.3]).

To conclude, we list a few relevant works that we did not touch upon in this survey: [19, 25, 27, 37, 94, 135]. A good part of the mating theory discussed above carries over to the setting of Thurston maps (i.e., postcritically finite, orientation-preserving branched coverings of $\mathbb{S}^2$), for which we encourage the reader to consult [7, 13]. Several beautiful visual illustrations of polynomial matings can be found in [24]. For a list of open questions on polynomial matings, we refer the reader to [20].

10.8 Combining Rational Maps and Kleinian Groups

In this section, we will expound recently developed frameworks for combining polynomials (respectively, anti-polynomials) with Kleinian (respectively, reflection) groups.

10.8.1 Mating Anti-polynomials with Reflection Groups

The story of mating anti-polynomials with Kleinian reflection groups began with the study of a new class of anti-holomorphic dynamical systems given by *Schwarz reflection maps* associated with *quadrature domains*. We will recall the definitions of these objects, and sketch the simplest examples of the mating phenomenon in Sect. 10.8.1.1. To put these examples in a general framework, we will introduce in Sect. 10.8.1.2 a class of Kleinian reflection groups (called *necklace reflection groups*), that are central to the mating construction. Further, we will associate a map (called the *Nielsen map*) to each necklace reflection group that is *orbit equivalent* to the group. In Sect. 10.8.1.3, we formalize the notion of conformal mating of a necklace group and an anti-polynomial. Section 10.8.1.4 summarizes some of the main results of [75, 78, 79], where various explicit examples of Schwarz reflection maps were shown to be conformal matings of necklace groups and anti-polynomials. Finally in Sect. 10.8.1.5, we state a general combination theorem for necklace groups and anti-polynomials proved in [91].

10.8.1.1 Schwarz Reflection Maps and Motivating Examples

By definition, a domain $\Omega \subsetneq \widehat{\mathbb{C}}$ satisfying $\infty \notin \partial\Omega$ and $\Omega = \operatorname{int} \overline{\Omega}$ is a *quadrature domain* if there exists a continuous function $\sigma : \overline{\Omega} \to \widehat{\mathbb{C}}$ such that σ is anti-meromorphic in Ω and $\sigma(z) = z$ on the boundary $\partial\Omega$. Such a function σ is unique (if it exists), and is called the *Schwarz reflection map* associated with Ω. (See [3], [77] and the references therein.)

It is well known that except for a finite number of *singular* points (cusps and double points), the boundary of a quadrature domain consists of finitely many disjoint real analytic curves [132]. Every non-singular boundary point has a neighborhood where the local reflection in $\partial\Omega$ is well-defined. The (global) Schwarz reflection σ is an anti-holomorphic continuation of all such local reflections.

Round disks on the Riemann sphere are the simplest examples of quadrature domains. Their Schwarz reflections are just the usual circle reflections. Further examples can be constructed using univalent polynomials or rational functions. In fact, simply connected quadrature domains admit a simple characterization.

Proposition 10.36 ([3, Theorem 1]) *A simply connected domain $\Omega \subsetneq \widehat{\mathbb{C}}$ with $\infty \notin \partial\Omega$ and int $\overline{\Omega} = \Omega$ is a quadrature domain if and only if the Riemann uniformization $f : \mathbb{D} \to \Omega$ extends to a rational map on $\widehat{\mathbb{C}}$. The Schwarz reflection map σ of Ω is given by $f \circ (1/\overline{z}) \circ (f|_{\mathbb{D}})^{-1}$.*

In this case, if the degree of the rational map f is d, then $\sigma : \sigma^{-1}(\Omega) \to \Omega$ is a (branched) covering of degree $(d-1)$, and $\sigma : \sigma^{-1}(int\ \Omega^c) \to int\ \Omega^c$ is a (branched) covering of degree d.

$$\begin{array}{ccc} \overline{\mathbb{D}} & \xrightarrow{\ f\ } & \overline{\Omega} \\ {\scriptstyle 1/\overline{z}}\downarrow & & \downarrow{\scriptstyle \sigma} \\ \widehat{\mathbb{C}} \setminus \mathbb{D} & \xrightarrow[\ f\ ]{} & \widehat{\mathbb{C}}. \end{array}$$

In [77], questions on equilibrium states of certain 2-dimensional Coulomb gas models were answered using iteration of Schwarz reflection maps associated with quadrature domains. It transpired from their work that these maps give rise to dynamical systems that are interesting in their own right. The general situation is as follows. Given a disjoint collection of quadrature domains, we call the complement of their union a *droplet*. Removing the double points and cusps from the boundary of a droplet yields the *desingularized droplet* or the *fundamental tile*. One can then look at a partially defined anti-holomorphic dynamical system σ that acts on (the closure of) each quadrature domain as its Schwarz reflection map. Under this dynamical system, the Riemann sphere $\widehat{\mathbb{C}}$ admits a dynamically invariant partition. The first one is an open set called the *escaping/tiling set*, it is the set of all points that eventually escape to the fundamental tile (on the interior of which σ is not defined). The second invariant set is the *non-escaping* set, the complement of the tiling set or equivalently, the set of all points on which σ can be iterated forever. When the tiling set contains no critical points of σ, it is often the case that the dynamics of σ on its non-escaping set resembles that of an anti-polynomial on its filled Julia set, while the $\sigma-$action on the tiling set exhibits features of reflection groups.

This is precisely the case for the Schwarz reflection map of the exterior of a deltoid curve: this map is conformally conjugate to the anti-polynomial $\overline{z}^2$ on its non-escaping set, and conformally conjugate to a suitable piecewise circular reflection map associated with the ideal triangle reflection group on its tiling set. In this sense, this map is a conformal mating of $\overline{z}^2$ and the ideal triangle reflection group [78, §5] (see Theorem 10.39 and Fig. 10.2).

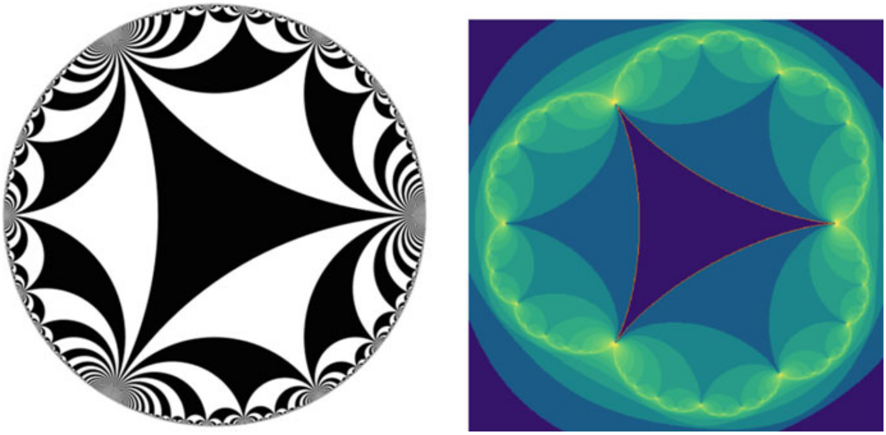


Fig. 10.2 Left: The tessellation of $\mathbb{D}$ for the ideal triangle reflection group. Right: The dynamical plane of the Schwarz reflection map associated with the quadrature domain $\Omega_0 = f_0(\mathbb{D})$ (the exterior of a deltoid curve), where $f_0(z) = 1/z + z^2/2$. The dynamics on the exterior of the bright green fractal curve is conformally conjugate to $\overline{z}^2|_{\mathbb{D}}$, while the dynamics on the interior is conformally equivalent to the Nielsen map of $\mathbf{\Gamma}_3$

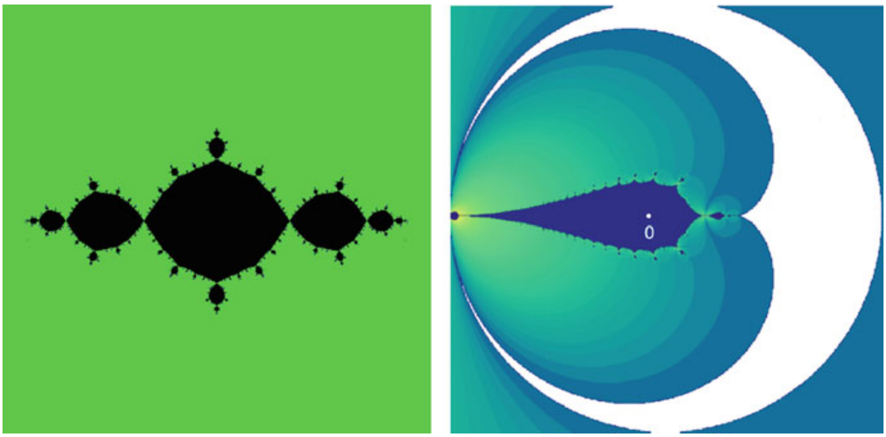


Fig. 10.3 Under the bijection χ of Theorem 10.41, the postcritically finite quadratic anti-polynomial $\overline{z}^2 - 1$ corresponds to F_a with $a = 0$. Left: The filled Julia set of $\overline{z}^2 - 1$. Right: The part of the non-escaping set of F_0 inside the cardioid (in dark blue) with the critical point 0 marked. Both maps have a critical cycle of period 2

The above example was extended in [79] by studying the Schwarz reflection maps associated with a fixed cardioid and a family of circumscribing circles. Such Schwarz reflection maps were shown to be conformal matings of generic quadratic anti-holomorphic polynomials with the ideal triangle reflection group (see Theorem 10.41 for the precise statement and Fig. 10.3 for a specific example).

While the above examples produce matings of a rigid group with quadratic anti-polynomials, conformal matings of a large class of Kleinian reflection groups (called *necklace groups*) with the anti-polynomial $\overline{z}^d$ were constructed in [75], and these matings were realized as Schwarz reflection maps arising from a natural space of 'univalent rational maps'.

10.8.1.2 Necklace Reflection Groups

A circle packing is a connected collection of oriented circles in $\mathbb{C}$ with disjoint interiors (where the interior is determined by the orientation). Up to a Möbius map, we can always assume that no circle of the circle packing contains ∞ in its interior; i.e., the interior int C of each circle C in the circle packing can be assumed to be the bounded complementary component of C. Combinatorially, a circle packing can be described by its *contact graph*, where we associate a vertex to each circle, and connect two vertices by an edge if and only if the two associated circles touch. By the Koebe-Andreev-Thurston circle packing theorem [142, Corollary 13.6.2], every connected, simple, planar graph is the contact graph of some circle packing.

Definition 10.26 A *necklace reflection group* is a group generated by reflections in the circles of a finite circle packing whose contact graph is 2-*connected* and *outerplanar*; i.e., the contact graph remains connected if any vertex is deleted, and has a face containing all the vertices on its boundary (Fig. 10.4).

Note that since a necklace reflection group is a discrete subgroup of the group of all Möbius and anti-Möbius automorphisms of $\widehat{\mathbb{C}}$, definitions of limit set and domain of discontinuity can be easily extended to necklace reflection groups. By [85, Proposition 3.4], the limit set of a necklace reflection group is connected. Moreover, for a necklace reflection group Γ generated by reflections in the circles $C_1, \cdots, C_{d+1}$, the set

$$\mathcal{F}_\Gamma := \widehat{\mathbb{C}} \setminus \left(\bigcup_{i=1}^{d+1} \text{int } C_i \bigcup_{j \neq k} (C_j \cap C_k) \right)$$

is a fundamental domain for the $\Gamma-$action on $\Omega(\Gamma)$ [75, Proposition 7] (Fig. 10.5).

The domain of discontinuity of a necklace group has a simply connected, invariant component. Thus, such groups play the role of *Bers slice closure* Kleinian groups in the world of reflection groups.

To a necklace reflection group Γ, one can associate a piecewise anti-Möbius reflection map ρ_Γ that plays an important role in the mating construction.

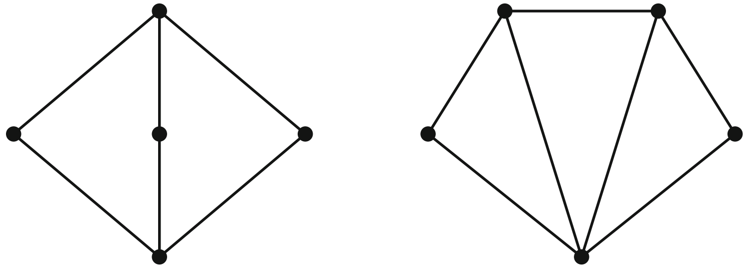


Fig. 10.4 Left: A 2-connected graph that is not outerplanar. Right: A 2-connected, outerplanar graph. A circle packing realizing this graph and the limit set of the associated necklace reflection group are shown in Fig. 10.5

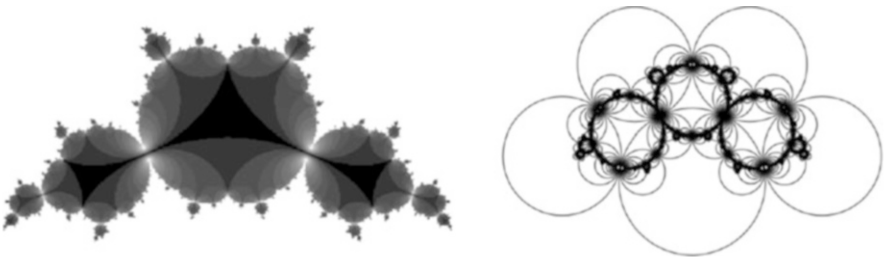


Fig. 10.5 Left: The dynamical plane of the Schwarz reflection map associated with some $f \in \partial\Sigma_4^*$. Right: The limit set of the corresponding necklace reflection group Γ_f

Definition 10.27 ([75, Definition 14], [91, Definition 6.6]) Let Γ be a necklace reflection group generated by reflections $\{r_i\}_{i=1}^{d+1}$ in circles $\{C_i\}_{i=1}^{d+1}$. We define the associated *Nielsen map* ρ_Γ by:

$$\rho_\Gamma : \bigcup_{i=1}^{d+1} \overline{\text{int } C_i} \to \widehat{\mathbb{C}}, \quad z \longmapsto r_i(z) \text{ if } z \in \overline{\text{int } C_i}.$$

The next proposition underscores the intimate dynamical connection between a necklace group Γ and its Nielsen map ρ_Γ.

Proposition 10.37 ([75, Proposition 16]) *Let Γ be a necklace reflection group. The map ρ_Γ is orbit equivalent to Γ on $\widehat{\mathbb{C}}$; i.e., for any two points $z, w \in \widehat{\mathbb{C}}$, there exists $g \in \Gamma$ with $g(z) = w$ if and only if there exist non-negative integers n_1, n_2 such that $\rho_\Gamma^{\circ n_1}(z) = \rho_\Gamma^{\circ n_2}(w)$.*

The simplest examples of necklace reflection groups are regular ideal polygon reflection groups.

Definition 10.28 Consider the Euclidean circles $\mathbf{C}_1, \cdots, \mathbf{C}_{d+1}$ where $\mathbf{C}_j$ intersects $\mathbb{S}^1$ at right angles at the roots of unity $\exp\left(\frac{2\pi i\cdot(j-1)}{d+1}\right)$, $\exp\left(\frac{2\pi i\cdot j}{d+1}\right)$. (By [146, Part II, Chapter 5, Theorem 1.2], the group generated by reflections in these circles is discrete.) We denote this group by $\mathbf{\Gamma}_{d+1}$.

Note that the Nielsen map $\rho_{\mathbf{\Gamma}_{d+1}}$ of the regular ideal polygon reflection group $\mathbf{\Gamma}_{d+1}$ restricts to an expansive degree d orientation-reversing covering of $\mathbb{S}^1$. By [26], there exists a homeomorphism $\mathcal{E}_d$ of the circle that conjugates $\rho_{\mathbf{\Gamma}_{d+1}}$ to $\overline{z}^d$. The conjugacy $\mathcal{E}_d$ serves as a connecting link between reflection groups and quadratic anti-polynomials.

10.8.1.3 Conformal Mating of Anti-polynomials and Necklace Groups

The precise meaning of conformal matings of the Nielsen map of a necklace group and an anti-polynomial is given below. The definition is an adaptation of the classical definition of conformal matings of two polynomials.

Let Γ be a necklace group generated by reflections in circles $C_1, \cdots, C_{d+1}$. The unbounded component of the domain of discontinuity $\Omega(\Gamma)$ is $\Gamma-$invariant [75, Proposition 15], and we denote it by $\Omega_\infty(\Gamma)$. We also set $\mathcal{K}(\Gamma) := \mathbb{C} \setminus \Omega_\infty(\Gamma)$. According to [75, Proposition 22], the restriction of ρ_Γ to $\Omega_\infty(\Gamma)$ is conformally conjugate to the $\rho_{\mathbf{\Gamma}_{d+1}}-$action on $\widehat{\mathbb{C}} \setminus \overline{\mathbb{D}}$, and (the inverse of) this conformal conjugacy continuously extends to yield a semiconjugacy $\phi_\Gamma : \mathbb{S}^1 \to \Lambda(\Gamma) = \partial\mathcal{K}(\Gamma)$ between $\rho_{\mathbf{\Gamma}_{d+1}}|_{\mathbb{S}^1}$ and $\rho_\Gamma|_{\Lambda(\Gamma)}$ such that $\phi_\Gamma(1)$ is the point of tangential intersection of C_1 and C_{d+1}. Recall also that $\mathcal{E}_d : \mathbb{S}^1 \to \mathbb{S}^1$ is a topological conjugacy between $\rho_{\mathbf{\Gamma}_{d+1}}|_{\mathbb{S}^1}$ and $z \mapsto \overline{z}^d|_{\mathbb{S}^1}$.

Let P be a monic, centered, anti-polynomial of degree d such that $\mathcal{J}(P)$ is connected and locally connected. Denote by $\phi_P : \mathbb{D}^* \to \mathcal{B}_\infty(P)$ the Böttcher coordinate for P such that $\phi'_P(\infty) = 1$. We note that since $\partial\mathcal{K}(P) = \mathcal{J}(P)$ is locally connected by assumption, it follows that ϕ_P extends to a semiconjugacy between $z \mapsto \overline{z}^d|_{\mathbb{S}^1}$ and $P|_{\mathcal{J}(P)}$.

The equivalence relation below specifies a gluing of $\mathcal{K}(\Gamma)$ with $\mathcal{K}(P)$ along their boundaries. The presence of the topological conjugacy $\mathcal{E}_d$ in the definition of the equivalence relation ensures that the maps ρ_Γ and P fit together to produce a continuous map on the resulting topological 2-sphere (when there is no Moore obstruction).

Definition 10.29 We define the equivalence relation $\sim$ on $\mathcal{K}(\Gamma) \sqcup \mathcal{K}(P)$ generated by $\phi_\Gamma(t) \sim \phi_P(\overline{\mathcal{E}_d(t)})$ for all $t \in \mathbb{S}^1$.

The following definition essentially says that an anti-holomorphic map F (defined on a subset of the Riemann sphere) is a conformal mating of Γ and P if there are continuous semi-conjugacies from $\mathcal{K}(\Gamma), \mathcal{K}(P)$ into the dynamical plane of F (conformal on the interiors) such that the images fill up the whole sphere and intersect only along their boundaries as prescribed by the equivalence relation $\sim$ (compare Definition 10.25).

Definition 10.30 ([91, Definition 10.16]) Let Γ be a necklace group as above, and let P be a monic, centered anti-polynomial such that $\mathcal{J}(P)$ is connected and locally connected. Further, let $\Omega \subsetneq \widehat{\mathbb{C}}$ be an open set, and $F : \overline{\Omega} \to \widehat{\mathbb{C}}$ be a continuous map that is anti-meromorphic on Ω. We say that F is a *conformal mating* of Γ with P if there exist continuous maps

$$\psi_P : \mathcal{K}(P) \to \widehat{\mathbb{C}} \text{ and } \psi_\Gamma : \mathcal{K}(\Gamma) \to \widehat{\mathbb{C}},$$

conformal on int $\mathcal{K}(P)$, int $\mathcal{K}(\Gamma)$, respectively, such that

1. $\psi_P(\mathcal{K}(P)) \bigcup \psi_\Gamma(\mathcal{K}(\Gamma)) = \widehat{\mathbb{C}}$,
2. $\Omega = \widehat{\mathbb{C}} \setminus \psi_\Gamma(\overline{\mathcal{F}_\Gamma})$,
3. $\psi_P \circ P = F \circ \psi_P$ on $\mathcal{K}(P)$,
4. $\psi_\Gamma \circ \rho_\Gamma = F \circ \psi_\Gamma$ on $\mathcal{K}(\Gamma) \setminus \text{int}\,\mathcal{F}_\Gamma$, and
5. $\psi_\Gamma(z) = \psi_P(w)$ if and only if $z \sim w$ where $\sim$ is as in Definition 10.29.

Remark 10.38 *For the purposes of mating necklace groups with anti-polynomials, it is important to work with labeled circle packings, or equivalently, to regard the space of necklace groups as a space of representations of the ideal polygon reflection group* $\mathbf{\Gamma}_{d+1}$. *While we have suppressed this abstraction for ease of exposition, we refer the reader to [75, §2.2] or [91, §10.1], where necklace groups are organized in Bers slices of* $\mathbf{\Gamma}_{d+1}$. *Although this point of view may seem like an artificial complication at a first glance, the language of representations turns out to be an unavoidable technicality in the mating theory. Roughly speaking, different representations give rise to different ways of gluing the limit set of a necklace group with the Julia set of an anti-polynomial, and the choice of gluing determines whether or not a conformal mating exists (compare [91, Remark 10.21]).*

10.8.1.4 Examples of the Mating Phenomenon

By studying the dynamics and parameter spaces of specific families of Schwarz reflection maps, one can often recognize such maps as matings of anti-polynomials and necklace reflection groups. This strategy was successfully implemented in [75, 78, 79]. We collect some results from these papers in this subsection.

Example 10.1 (The Deltoid Reflection) We will start with the simplest instance of the mating phenomenon; namely, the conformal mating of the anti-polynomial $\overline{z}^2$ and the ideal triangle reflection group $\mathbf{\Gamma}_3$.

Theorem 10.39 ([78, Theorem 1.1]) *The map* $f_0(z) = 1/z + z^2/2$ *is injective on* $\overline{\mathbb{D}}$, *and hence* $\Omega_0 := f_0(\mathbb{D})$ *is a simply connected quadrature domain. The associated Schwarz reflection map* σ_0 *is the unique conformal mating of* $\overline{z}^2$ *and* $\mathbf{\Gamma}_3$.

Remark 10.40 *A* welding homeomorphism *is a homeomorphism of the circle that arises as the composition of a conformal map from the unit disk onto the interior region of a Jordan curve with a conformal map from the exterior of this Jordan curve* *onto the exterior of the unit disk. A complex-analytic corollary of Theorem 10.39 is that the circle homeomorphism $\mathcal{E}_2$ is a welding homeomorphism. That the same is true for each $\mathcal{E}_d$ ($d \geq 2$) follows from a straightforward higher degree generalization of Theorem 10.39 worked out in [80, Appendix B] (also compare Theorem 10.44 below). We refer the reader to [91, Theorem 5.1] for a general conformal welding result for circle homeomorphisms conjugating suitable covering maps of the circle.*

Example 10.2 (The Circle and Cardioid Family) To describe Schwarz reflection maps that are conformal matings of other quadratic anti-polynomials with the ideal triangle reflection group, we need to recall the Circle and Cardioid family which was introduced in [78, §6]. We consider the fixed cardioid

$$\heartsuit := \left\{ w = z/2 - z^2/4 : \ |z| < 1 \right\},$$

and for each complex number $a \in \mathbb{C} \setminus (-\infty, -1/12)$, let $B(a, r_a)$ be the smallest open disk containing $\heartsuit$ centered at a (in other words, $\{w : |w - a| = r_a\}$ is a circumcircle of the cardioid; see [79, Figure 2]). Let $\Omega_a := \heartsuit \cup \overline{B}(a, r_a)^c$ (where $\overline{B}(a, r_a)$ is the closed disk $\{w : |w - a| \leq r_a\}$), and $T_a := \Omega_a^c$. We now define a piecewise Schwarz reflection dynamical system $F_a : \overline{\Omega_a} \to \widehat{\mathbb{C}}$ as,

$$w \mapsto \begin{cases} \sigma(w) & \text{if } w \in \overline{\heartsuit}, \\ \sigma_a(w) & \text{if } w \in B(a, r_a)^c, \end{cases}$$

where σ is the Schwarz reflection of $\heartsuit$, and σ_a is reflection with respect to the circle $\partial B(a, r_a)$. The family

$$\mathcal{S} := \left\{ F_a : \overline{\Omega_a} \to \widehat{\mathbb{C}} : a \in \mathbb{C} \setminus (-\infty, -1/12) \right\}$$

is referred to as the C&C family.

For any $a \in \mathbb{C} \setminus (-\infty, -1/12)$, ∂T_a has two singular points; namely, the double point α_a where $\partial B(a, r_a)$ touches $\partial \heartsuit$, and the cusp point $\frac{1}{4}$. Both of them are fixed points of F_a. The *fundamental tile* of F_a is defined as $T_a^0 := T_a \setminus \{\alpha_a, \frac{1}{4}\}$. A parameter $a \in \mathbb{C} \setminus (-\infty, -1/12)$ (equivalently, the corresponding map $F_a \in \mathcal{S}$) is said to be *postcritically finite* if the unique (simple) critical point 0 of F_a has a finite forward orbit that does not meet T_a^0. The following mating description for postcritically finite maps in $\mathcal{S}$ was given in [79].

Theorem 10.41 ([78, §8], [79, Theorems 1.1, 1.2]) *There exists a bijection χ between postcritically finite maps in $\mathcal{S}$ and (the Möbius conjugacy classes of) postcritically finite quadratic anti-polynomials $\overline{z}^2 + c$ (excluding $\overline{z}^2$) such that the postcritically finite map $F_a \in \mathcal{S}$ is a conformal mating of the ideal triangle reflection group $\mathbf{\Gamma}_3$ and the quadratic anti-polynomial $\overline{z}^2 + \chi(a)$.*

Remark 10.42 *For conformal matings of $\mathbf{\Gamma}_3$ with more general quadratic anti-polynomials, see [79, Theorem 1.1]. Further, a combinatorial model of the* connectedness locus *of the family $\mathcal{S}$ is given in [79, Theorem 1.4] in terms of the* Tricorn*, which is the connectedness locus of quadratic anti-polynomials (for a quick introduction to the Tricorn, see [79, §2], and for its detailed topological properties, see [62–64, 120]).*

Example 10.3 (The Space Σ_d^*) The family of 'univalent rational maps'

$$\Sigma_d^* := \left\{ g(z) = z + \frac{a_1}{z} + \cdots + \frac{a_d}{z^d} : a_d = -\frac{1}{d} \text{ and } g|_{\widehat{\mathbb{C}}\setminus\overline{\mathbb{D}}} \text{ is conformal} \right\}$$

was introduced in [76] and studied extensively in [74, 75] in terms of the associated Schwarz reflection maps.

Remark 10.43 *The family Σ_d^* is closely related to the classically studied space Σ of suitably normalized schlicht functions on $\widehat{\mathbb{C}} \setminus \overline{\mathbb{D}}$, see [36, §4.7, §9.6].*

Combining the pinching deformation theory for Σ_d^* (developed in [74]) with tools from holomorphic dynamics, it was proved in [75] that (Fig. 10.5):

Theorem 10.44 ([75, Theorem A]) *There is a bijection $f \mapsto \Gamma_f$ between Σ_d^* and the space of necklace reflection groups of rank $d+1$ (up to a natural equivalence) such that the Schwarz reflection map associated with $f \in \Sigma_d^*$ is a conformal mating of the anti-polynomial $\overline{z}^d$ with the corresponding necklace group Γ_f.*

10.8.1.5 The General Theorem

We conclude our discussion of combinations of necklace reflection groups and anti-polynomials with a general existence theorem:

Theorem 10.45 ([91, Lemma 10.17, Theorem 10.20]) *Let P be a monic, post-critically finite, hyperbolic anti-polynomial of degree d, and let Γ be a necklace group. Then, P and Γ are conformally mateable if and only if $\mathcal{K}(P) \sqcup \mathcal{K}(\Gamma)/\sim$ is homeomorphic to $\mathbb{S}^2$ (where $\sim$ is the equivalence relation from Definition 10.29).*

Moreover, if $F : \overline{\Omega} \to \widehat{\mathbb{C}}$ is a conformal mating of Γ and P, then each component of Ω is a simply connected quadrature domain, and F is the piecewise defined Schwarz reflection map associated with these quadrature domains.

The hard part of the above theorem is to show that if $\mathcal{K}(P) \sqcup \mathcal{K}(\Gamma)/\sim$ is homeomorphic to $\mathbb{S}^2$, then a conformal mating of P and Γ exists. In fact, the condition that $\mathcal{K}(P) \sqcup \mathcal{K}(\Gamma)/\sim\, \cong \mathbb{S}^2$ guarantees the existence of a topological mating on a 2-sphere, but promoting the topological mating to an anti-holomorphic map lies at the heart of the difficulty. This goal is achieved in two steps. One first uses Thurston's topological characterization theorem to construct a hyperbolic anti-rational map R that is a conformal mating of P and another postcritically finite (in fact, critically fixed), hyperbolic anti-polynomial P_Γ such that the Julia dynamics of P_Γ is topologically conjugate to the limit set dynamics of the Nielsen map ρ_Γ. The existence of such an anti-polynomial P_Γ follows from [75] or [85], while conformal mateability of P and P_Γ follows from the general mateability criterion given in Theorem 10.34 (in fact, the condition $\mathcal{K}(P) \sqcup \mathcal{K}(\Gamma)/\sim \cong \mathbb{S}^2$ is equivalent to saying that the topological mating of P and P_Γ is Moore-unobstructed, so Theorem 10.34 can be applied to produce R). Finally, to turn R into a conformal mating of P and Γ, one needs to glue Nielsen maps of ideal polygon reflection groups in suitable invariant Fatou components of R. The fact that all fixed points of R on its Julia set are hyperbolic while those of a Nielsen map are parabolic prohibits the use of purely quasiconformal tools to carry out this task. This problem is tackled by employing surgery techniques involving David homeomorphisms: generalizations of quasiconformal homeomorphisms.

Remark 10.46 *Although Theorem 10.45 guarantees the existence of conformal matings of suitable anti-polynomials and necklace reflection groups, in general, it may be hard to find explicit Schwarz reflection maps realizing such conformal matings. However, in certain low complexity situations, the second statement of the theorem (that the conformal matings are piecewise Schwarz reflection maps associated with simply connected quadrature domains) allows one to use Proposition 10.36 and the desired dynamical properties to explicitly characterize the conformal matings (see Fig. 10.6 for an illustration, and [91, §11] for various worked out examples).*

10.8.2 Mating Polynomials with Kleinian Groups

This subsection is a summary of [114], where a new setup for combination theorems of complex polynomials and Kleinian surface groups was designed using the notion of orbit equivalence.

10.8.2.1 The Fuchsian Case

A foundational problem that arises in trying to make sense of what it means to combine a polynomial P with a Kleinian group Γ is that on one side of the picture we have the semigroup $\langle P \rangle$ generated by P, while on the other side we have a non-commutative group Γ generated by more than one element. To formulate a precise notion of mateability between Fuchsian groups and complex polynomials (with Jordan curve Julia sets), one needs to address this inherent discord between these two objects, and this leads to the notion of *mateable* circle maps: *single* maps A that capture essential dynamical and combinatorial features of Fuchsian groups acting on $\mathbb{S}^1$. Further, A should also be dynamically compatible with polynomial maps. Before giving a precise definition of mateable maps, let us outline the underlying motivation: the following features are required of a mateable map $A : \mathbb{S}^1 \to \mathbb{S}^1$.

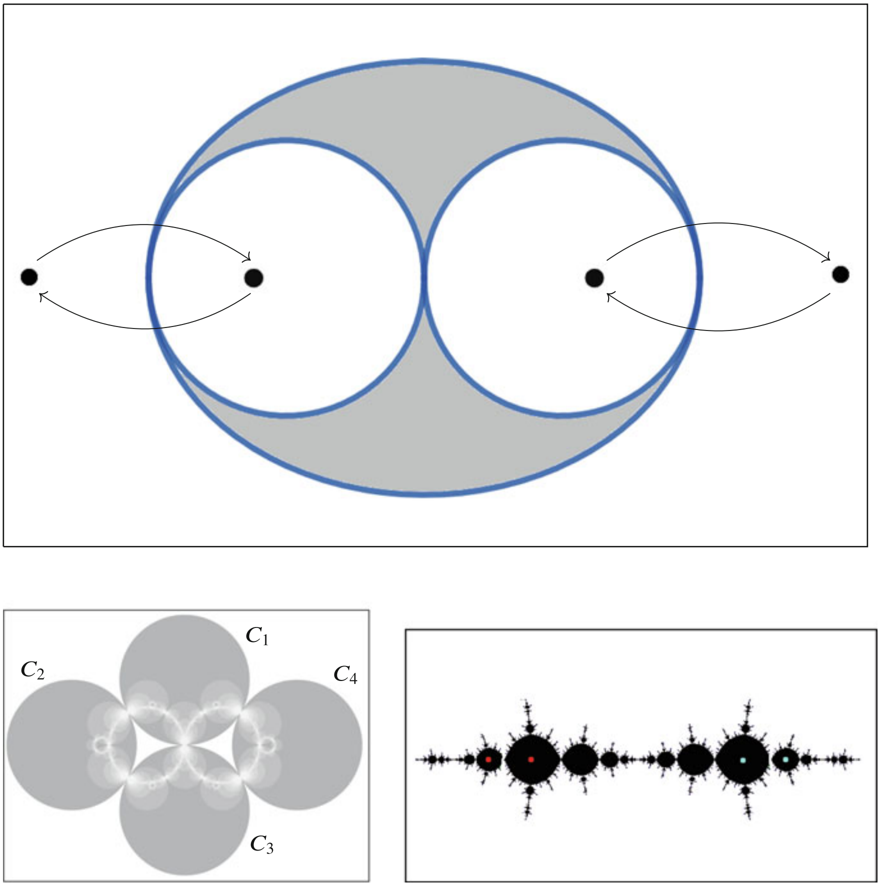


Fig. 10.6 Bottom left: The circles C_i generate a necklace reflection group Γ. Bottom right: The dynamical plane of $P(z) = \overline{z}^3 - \frac{3i}{\sqrt{2}}\overline{z}$; each critical point of which forms a 2-cycle. Top: The conformal mating of P and the necklace group Γ is given by the piecewise Schwarz reflection map associated with the disjoint union of three quadrature domains: the exterior of an ellipse, and two round disks contained in the interior of the ellipse. Each of the two critical points of F forms a 2-cycle (See [91, §11.2] for proofs of these statements.)

1. A must be dynamically compatible with a Fuchsian group Γ. This leads to
   a. orbit equivalence between A and Γ.
   b. A has to be piecewise Fuchsian.
2. A must be dynamically compatible with complex polynomials. Hence we demand the existence of a topological conjugacy between A and the polynomial $z^d|_{\mathbb{S}^1}$ (where $d \geq 2$ is the degree of A),
3. A must be combinatorially compatible with z^d leading to a Markov condition,

4. A must be conformally compatible with z^d requiring absence of asymmetrically hyperbolic periodic break-points of A (this is a weaker version of the C^1-condition).

To fulfill the above requirements, we make the following definition. We denote the group of conformal automorphisms of the unit disk $\mathbb{D}$ by $\mathrm{Aut}(\mathbb{D})$.

Definition 10.31 ([114, Definitions 2.7, 2.16])

1. A map $A : \mathbb{S}^1 \to \mathbb{S}^1$ is called *piecewise Möbius* if there exist $k \in \mathbb{N}$, closed arcs $I_j \subset \mathbb{S}^1$, and $g_j \in \mathrm{Aut}(\mathbb{D})$, $j \in \{1, \cdots, k\}$, such that
   (a) $\mathbb{S}^1 = \bigcup_{j=1}^{k} I_j$,
   (b) $\mathrm{int}\, I_m \cap \mathrm{int}\, I_n = \emptyset$ for $m \neq n$, and
   (c) $A|_{I_j} = g_j$.

   A piecewise Möbius map is called *piecewise Fuchsian* if $g_1, \cdots, g_k$ generate a Fuchsian group, which we denote by Γ_A.
2. A map $A : \mathbb{S}^1 \to \mathbb{S}^1$ is called *piecewise Fuchsian Markov* if it is a piecewise Fuchsian expansive covering map (of degree at least two) such that the pieces (intervals of definition) of A form a Markov partition for $A : \mathbb{S}^1 \to \mathbb{S}^1$.
3. A piecewise Fuchsian Markov map A is said to be *mateable* if A is orbit equivalent to the Fuchsian group Γ_A generated by its pieces, and none of the periodic break-points of A is asymmetrically hyperbolic.

We refer the reader to [114, §2] for the definition of the term 'symmetrically hyperbolic', and for a detailed discussion on the necessity of each of the requirements in the definition of a mateable map. We also note that the expansivity condition above ensures that a mateable map is topologically conjugate to the polynomial z^d (for some $d \geq 2$).

Remark 10.47 *In the anti-holomorphic setting, the role of mateable maps was played by Nielsen maps of necklace reflection groups (see Sect. 10.8.1.2).*

The simplest example of a mateable map is given by the classical *Bowen–Series map* [16, 17]. While such a map can be defined for arbitrary Fuchsian groups equipped with suitable fundamental domains, they are typically discontinuous. However, it turns out that for Fuchsian groups uniformizing spheres with punctures (possibly with one/two order two orbifold points), the Bowen–Series map is a covering map of the circle satisfying the defining properties of a mateable map [114, §2].

Higher Bowen–Series Maps More examples of mateable maps are given by *higher Bowen–Series maps* of punctured sphere Fuchsian groups (see [114, §4] for their definition and basic properties). As suggested by the name, there are close connections between higher Bowen–Series maps and Bowen–Series maps. Indeed, the higher Bowen–Series map of a Fuchsian group uniformizing $S_{0,k}$ (a sphere with

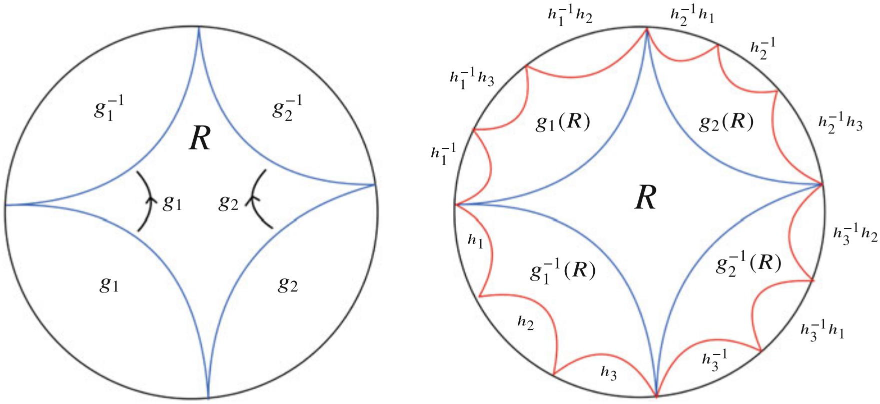


Fig. 10.7 Left: R is a fundamental domain of a Fuchsian group Γ uniformizing $S_{0,3}$ with side pairing transformations $g_1^{\pm 1}, g_2^{\pm 1}$. The canonical extension of the Bowen–Series map of Γ (associated with R) is defined on $\overline{\mathbb{D}} \setminus \text{int } R$ in terms of $g_1^{\pm 1}, g_2^{\pm 1}$ as shown in the figure. Right: The second iterate of the Bowen–Series map of Γ is a higher Bowen–Series map of the index two subgroup $\Gamma' = \langle g_1^2, g_1 g_2, g_1 g_2^{-1} \rangle \leq \Gamma$, which uniformizes $S_{0,4}$. Its canonical extension is defined on the region enclosed by $\mathbb{S}^1$ and the red (hyperbolic) geodesics in terms of $h_1 = g_1^2, h_2 = g_1 g_2, h_3 = g_1 g_2^{-1}$ as shown in the figure. It maps the boundary of the red polygon onto the boundary of R. The degree of the Bowen–Series map of Γ (as a circle covering) is 3, so the degree of the higher Bowen–Series map of Γ' is 9

k punctures) can be represented as the second iterate of the Bowen–Series map of a Fuchsian group uniformizing a sphere with roughly k/2 punctures and zero/one/two order two orbifold points [114, Corollary 5.6] (see Fig. 10.7). Alternatively, a higher Bowen–Series map of a Fuchsian group is obtained by 'gluing together' several Bowen–Series maps of the same Fuchsian group with overlapping fundamental domains [114, Proposition 4.5] (see Fig. 10.8).

Every piecewise Fuchsian Markov map A of the circle can be conformally extended to a canonically defined subset of $\overline{\mathbb{D}}$ (see [114, §2.2]). This extension is termed the *canonical extension of* A. The following result, which is a conformal combination theorem for punctured sphere Fuchsian groups and hyperbolic polynomials with Jordan curve Julia sets, can be regarded as an analog of the Bers simultaneous uniformization theorem in the current setting.

Theorem 10.48 ([114, Theorem 3.7, Theorem 4.8]) *The canonical extensions of Bowen–Series maps and higher Bowen–Series maps of Fuchsian groups uniformizing punctured spheres (possibly with one/two orbifold points of order two) can be conformally mated with polynomials lying in principal hyperbolic components (of appropriate degree).*

Remark 10.49 *As in the anti-holomorphic case, there is a key qualitative difference between the dynamics of Bowen–Series (respectively, higher Bowen–Series) maps on* $\mathbb{S}^1$ *and the dynamics of polynomials (lying in principal hyperbolic components)*

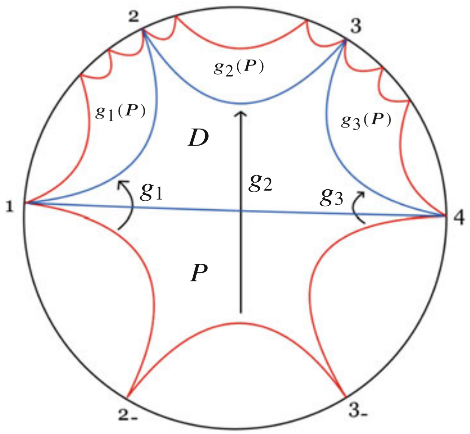


Fig. 10.8 The quadrilaterals D and P with ideal vertices at $1, 2, 3, 4$ and $1, 2_-, 3_-, 4$ (respectively) together form a fundamental domain of a Fuchsian group Γ uniformizing $S_{0,4}$ with side pairing transformations $g_1^{\pm 1}, g_2^{\pm 1}, g_3^{\pm 1}$. The corresponding higher Bowen–Series map A of Γ acts on the anti-clockwise arc from 1 to 4 as the Bowen–Series map of Γ associated with the fundamental domain $D \cup P$; while on the clockwise arc from j to $j+1$, A equals the Bowen–Series map of Γ associated with the fundamental domain $D \cup g_j(P)$ $(j = 1, 2, 3)$. The degree of a Bowen–Series map of Γ (associated with any fundamental domain) is 5, while the degree of a higher Bowen–Series map of Γ is 9

on their Julia set; namely, the former has parabolic fixed points on $\mathbb{S}^1$ *while all fixed points of the latter on their Julia sets are repelling. Consequently, the topological conjugacy between a Bowen–Series (respectively, higher Bowen–Series) map and such a polynomial is not quasisymmetric. This forces one to abandon classical quasiconformal techniques (used in the proof of the Bers simultaneous uniformization theorem), and apply David homeomorphisms to prove Theorem 10.48.*

Moduli Space of Fuchsian Matings In the torsion-free case, the only topological surfaces that Theorem 10.48 succeeds to combine with complex polynomials are punctured spheres (see [114, 6.35] for the definition of moduli space of matings between a topological surface and hyperbolic complex polynomials with Jordan curve Julia sets). This naturally raises the following questions.

1. Do mateable maps exist for higher genus surfaces (possibly with punctures)? More precisely, does there exist a mateable map A with $\mathbb{D}/\Gamma_A \cong S_{g,k}$, for $g \geq 1$?
2. Are Bowen–Series and higher Bowen–Series maps the only mateable maps associated with punctured spheres?

In this generality, the above questions remain open. However, [114, Theorems 6.18, 6.33] give a complete description of mateable maps satisfying some natural 2-point conditions over and above orbit equivalence. It turns out that under these additional hypotheses, punctured spheres are the only topological surfaces that can be combined with complex polynomials (see [114, Theorem 6.36] for a complete description of the interiors of such constrained moduli space of matings). A major part of the proofs of these theorems is to determine the topology of the surface $\mathbb{D}/\Gamma_A$ from the dynamical properties of a mateable map A, and this is accomplished by analyzing certain patterns and laminations associated with mateable maps.

10.8.2.2 The Case of Bers Boundary Groups

We proceed to discuss the structure of the boundaries of the moduli spaces of Fuchsian matings arising from Theorem 10.48.

For definiteness, let us fix a base Fuchsian group Γ_0 uniformizing $S_{0,k}$ ($k \geq 3$). For each Γ lying on the boundary of the Bers slice $\mathcal{B}(\Gamma_0)$, there exists a continuous map $\phi_\Gamma : \mathbb{S}^1 \to \Lambda_\Gamma$, called the *Cannon–Thurston map* after [22], that semi-conjugates the action of Γ_0 to that of Γ [29, 110–112]. In fact, the data of the *ending lamination* can be recovered from the Cannon–Thurston map. More precisely, the group Γ can be obtained by 'pinching a lamination' on the surface $\mathbb{D}/\Gamma_0$ (while keeping the hyperbolic structure on the surface $(\widehat{\mathbb{C}} \setminus \overline{\mathbb{D}})/\Gamma_0$ unchanged), and the endpoints of the corresponding geodesic lamination on $\mathbb{D}$ generate a Γ_0-invariant equivalence relation on $\mathbb{S}^1$ which agrees with the one defined by the fibers of the Cannon–Thurston map ϕ_Γ.

Assume further that A_{Γ_0} is a Bowen–Series (respectively, higher Bowen–Series) map of Γ_0. To extend the notion of mateability to a group $\Gamma \in \partial\mathcal{B}(\Gamma_0)$, one needs the limit set of Γ to carry a continuous, piecewise complex-analytic self-map A_Γ (that is orbit equivalent to Γ) defined by the following commutative diagram:

$$\begin{array}{ccc} \mathbb{S}^1 & \xrightarrow{A_{\Gamma_0}} & \mathbb{S}^1 \\ {\scriptstyle \phi_\Gamma}\downarrow & & \downarrow{\scriptstyle \phi_\Gamma} \\ \Lambda_\Gamma & \xrightarrow[A_\Gamma]{} & \Lambda_\Gamma \end{array}$$

See [114, §7.1] for details and an alternative description of A_Γ as a uniform limit of Bowen–Series (respectively, higher Bowen–Series) maps. The map A_Γ, if it exists, is called the *Bowen–Series* (respectively, *higher Bowen–Series*) map of Γ and can be thought of as a mateable map associated with a Bers boundary group.

It turns out that the existence of such a map A_Γ imposes severe restrictions on the laminations that can be pinched. The next theorem says that only finitely many possibilities exist. We call such laminations *admissible* (see Fig. 10.9 for an example of an admissible lamination in the Bowen–Series case).

For $\Gamma \in \partial\mathcal{B}(\Gamma_0)$, we denote the unique Γ-invariant component of the domain of discontinuity $\Omega(\Gamma)$ by $\Omega_\infty(\Gamma)$, and set $K(\Gamma) := \widehat{\mathbb{C}} \setminus \Omega_\infty(\Gamma)$. If Γ admits a Bowen–Series (respectively, higher Bowen–Series) map $A_\Gamma : \Lambda_\Gamma \to \Lambda_\Gamma$, then this map can be extended as a continuous, piecewise Möbius map to a canonical closed set $K(\Gamma) \setminus \text{int } R_\Gamma$, where R_Γ is a 'pinched' fundamental domain for the $\Gamma-$action

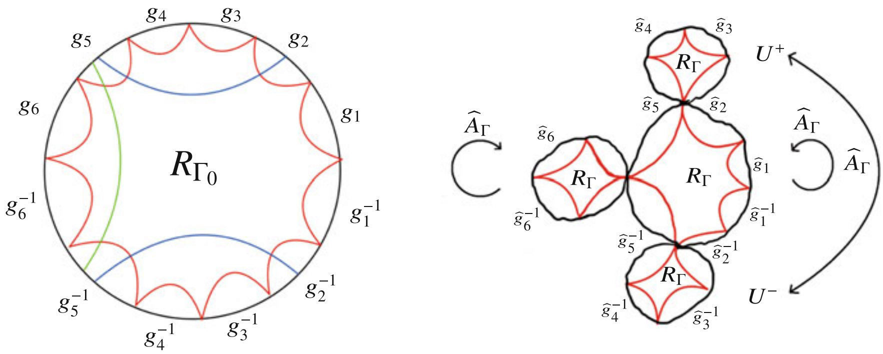


Fig. 10.9 Left: R_{Γ_0} is a fundamental domain of a Fuchsian group Γ_0 uniformizing $S_{0,7}$ with side pairing transformations $g_1^{\pm 1}, \cdots, g_6^{\pm 1}$. The geodesic lamination $\mathcal{L}^*$ on $\mathbb{D}/\Gamma_0$ consisting of two simple, closed curves corresponding to the elements $g_5, g_2 g_5^{-1} \in \Gamma_0$ is admissible for the Bowen–Series map A_{Γ_0}. The blue and green geodesics are the connected components of the Γ_0-lift of $\mathcal{L}^*$ that intersect R_{Γ_0}. Right: A cartoon of the limit set of a Bers boundary group Γ, which is obtained by pinching $\mathcal{L}^*$. The Γ-action on $\Omega(\Gamma) \setminus \Omega_\infty(\Gamma)$ admits a pinched fundamental domain R_Γ, and the canonical extension $\widehat{A}_\Gamma$ of the Bowen–Series map of Γ is defined on $K(\Gamma) \setminus \operatorname{int} R_\Gamma$. Two of the components of $\Omega(\Gamma)$ intersecting R_Γ are invariant under $\widehat{A}_\Gamma$, while the other two components $U^\pm$ form a 2-cycle. The first return map of $\widehat{A}_\Gamma$ on $U^\pm$ is conformally conjugate to higher Bowen–Series maps of punctured sphere Fuchsian groups. The Möbius maps defining $\widehat{A}_\Gamma$ are also marked, where $\widehat{g}_i$ is the image of g_i under the representation $\Gamma_0 \to \Gamma$

on $\Omega(\Gamma) \setminus \Omega_\infty(\Gamma)$ determined by R_{Γ_0} (see [114, §7.3]). This canonical extension is denoted by $\widehat{A}_\Gamma$. The following theorem also demonstrates conformal mateability of groups $\Gamma \in \partial\mathcal{B}(\Gamma_0)$ admitting Bowen–Series/ higher Bowen–Series maps with polynomials lying in principal hyperbolic components (of suitable degree).

Theorem 10.50 ([114, Lemma 7.3, Lemma 7.5, Theorem 7.19]) *Let Γ_0 be a Fuchsian group uniformizing $S_{0,k}$. Then, there are only finitely many quasiconformal conjugacy classes of groups $\Gamma \in \partial\mathcal{B}(\Gamma_0)$ for which the Cannon–Thurston map of Γ semi-conjugates the Bowen–Series (respectively, higher Bowen–Series) map of Γ_0 to a self-map of $\Lambda(\Gamma)$ that is orbit equivalent to Γ. These Kleinian groups arise out of pinching finitely many disjoint, simple, closed curves (on the surface $\mathbb{D}/\Gamma_0$) out of an explicit finite list. In particular, all such groups Γ are geometrically finite.*

Let $\Gamma \in \partial\mathcal{B}(\Gamma_0)$ be a group that admits a Bowen–Series (respectively, higher Bowen–Series) map A_Γ. Then the canonical extension $\widehat{A}_\Gamma : K(\Gamma) \setminus \operatorname{int} R_\Gamma \to K(\Gamma)$ can be conformally mated with polynomials lying in the principal hyperbolic component of degree $2k-3$ (respectively, $(k-1)^2$).

We refer the reader to [114, Remark 7.20] for a precise definition of conformal mateability of canonical extensions of the Bowen–Series/ higher Bowen–Series maps with polynomials lying in principal hyperbolic components (the definition is analogous to Definition 10.30).

The finiteness part of Theorem 10.50 underscores the incompatibility between group invariant geodesic laminations and polynomial laminations (see [72] for details on polynomial laminations) by establishing that the equivalence relation on $\mathbb{S}^1$ induced by a group invariant geodesic lamination on $\mathbb{D}$ is seldom invariant under A_{Γ_0} (since $A_{\Gamma_0}|_{\mathbb{S}^1}$ is topologically conjugate to $z^d|_{\mathbb{S}^1}$ for some $d \geq 2$, invariance under A_{Γ_0} should be thought of as z^d-invariance).

The proof of existence of a conformal mating between $\widehat{A_\Gamma} : K(\Gamma) \setminus \text{int } R_\Gamma \to K(\Gamma)$ and polynomials lying in principal hyperbolic components has two main steps. The first one is to topologically realize the action of $A_\Gamma|_{\Lambda_\Gamma}$ by the dynamics of a postcritically finite polynomial P_Γ on its Julia set, which is the content of [114, Theorem 7.16]. Once this is achieved, one needs to replace the dynamics of P_Γ on periodic Fatou components by the action of Bowen–Series/ higher Bowen–Series maps of suitable punctured sphere Fuchsian groups. This involves a rather delicate surgery technique using David homeomorphisms.

Acknowledgements The authors would like to thank Mikhail Lyubich for helpful conversations and useful suggestions. They also thank Athanase Papadopoulos for a careful reading of the manuscript and suggestions.

References

1. I. Agol, The virtual Haken conjecture. Doc. Math. **18**, 1045–1087 (2013). With an appendix by Agol, Daniel Groves, and Jason Manning
2. I. Agol, D. Groves, J.F. Manning, An alternate proof of Wise's malnormal special quotient theorem. Forum Math. Pi **4**, paper no. e1, 54 pp. (2016)
3. D. Aharonov, H.S. Shapiro, Domains on which analytic functions satisfy quadrature identities. J. Analyse Math. **30**, 39–73 (1976)
4. E. Alibegović, A combination theorem for relatively hyperbolic groups. Bull. Lond. Math. Soc. **37**(3), 459–466 (2005)
5. M. Aschenbrenner, S. Friedl, H. Wilton, *3-manifold Groups*. EMS Series of Lectures in Mathematics (European Mathematical Society (EMS), Zürich, 2015)
6. M. Aspenberg, P. Roesch, Newton maps as matings of cubic polynomials. Proc. Lond. Math. Soc. (3) **113**(1), 77–112 (2016)
7. L. Bartholdi, D. Dudko, Algorithmic aspects of branched coverings IV/V. Expanding maps. Trans. Amer. Math. Soc. **370**(11), 7679–7714 (2018)
8. N. Bergeron, D.T. Wise, A boundary criterion for cubulation. Amer. J. Math. **134**(3), 843–859 (2012)
9. L. Bers, Simultaneous uniformization. Bull. Amer. Math. Soc. **66**, 94–97 (1960)
10. L. Bers, Uniformization by Beltrami equations. Commun. Pure Appl. Math. **14**, 215–228 (1961)
11. M. Bestvina, R-trees in topology, geometry, and group theory, in *Handbook of Geometric Topology* (North-Holland, Amsterdam, 2002), pp. 55–91
12. M. Bestvina, M. Feighn, A combination theorem for negatively curved groups. J. Differ. Geom. **35**, 85–101 (1992)
13. M. Bonk, D. Meyer, *Expanding Thurston Maps*, volume 225 of Mathematical Surveys and Monographs (American Mathematical Society, Providence, 2017)
14. B.H. Bowditch, The Cannon-Thurston map for punctured surface groups. Math. Z. **255**, 35–76 (2007)
15. B.H. Bowditch, Relatively hyperbolic groups. Int. J. Algebra Comput. **22**, 1250016, 66 pp. (2012)

16. R. Bowen, Hausdorff dimension of quasicircles. Inst. Hautes Études Sci. Publ. Math. **50**, 11–25 (1979)
17. R. Bowen, C. Series, Markov maps associated with Fuchsian groups. Inst. Hautes Études Sci. Publ. Math. **50**, 153–170 (1979)
18. H. Brolin. Invariant sets under iteration of rational functions. Ark. Mat. **6**, 103–144 (1965)
19. X. Buff, A.L. Epstein, S. Koch, Twisted matings and equipotential gluings. Ann. Fac. Sci. Toulouse Math. (6) **21**(5), 995–1031 (2012)
20. X. Buff, A.L. Epstein, S. Koch, D. Meyer, K. Pilgrim, M. Rees, T. Lei, Questions about polynomial matings. Ann. Fac. Sci. Toulouse Math. (6) **21**(5), 1149–1176 (2012)
21. S. Bullett, C. Penrose, Mating quadratic maps with the modular group. Invent. Math. **115**, 483–511 (1994)
22. J. Cannon, W.P. Thurston, Group invariant peano curves. Geom. Topol. **11**, 1315–1355 (2007)
23. L. Carleson, T.W. Gamelin, *Complex Dynamics* (Springer, Berlin, 1993)
24. A. Chéritat, Polynomial matings on the Riemann sphere. https://www.math.univ-toulouse.fr/~cheritat/MatMovies/
25. A. Chéritat, Tan Lei and Shishikura's example of non-mateable degree 3 polynomials without a Levy cycle. Ann. Fac. Sci. Toulouse Math. (6) **21**(5), 935–980 (2012)
26. E.M. Coven, W.L. Reddy, Positively expansive maps of compact manifolds, in *Global theory of dynamical systems (Proc. Internat. Conf., Northwestern Univ., Evanston, Ill., 1979)*, volume 819 of Lecture Notes in Math. (Springer, Berlin, 1980), pp. 96–110
27. G. Cui, W. Peng, T. Lei, On a theorem of Rees-Shishikura. Ann. Fac. Sci. Toulouse Math. (6) **21**(5), 981–993 (2012)
28. F. Dahmani, Combination of convergence groups. Geom. Topol. **7**, 933–963 (2003)
29. S. Das, M. Mj, Semiconjugacies between relatively hyperbolic boundaries. Groups Geom. Dyn. **10**, 733–752 (2016)
30. A. De Zotti, Some connections of complex dynamics, in *Advancements in Complex Analysis–From Theory to Practice* (Springer, Cham, 2020), pp. 519–536
31. A. Douady, J.H. Hubbard, *Étude dynamique des polynômes complexes. Partie I*, volume 84 of Publications Mathématiques d'Orsay [Mathematical Publications of Orsay]. (Université de Paris-Sud, Département de Mathématiques, Orsay, 1984)
32. A. Douady, J.H. Hubbard, *Étude dynamique des polynômes complexes. Partie II*, volume 85 of Publications Mathématiques d'Orsay [Mathematical Publications of Orsay]. (Université de Paris-Sud, Département de Mathématiques, Orsay, 1985). With the collaboration of P. Lavaurs, T. Lei and P. Sentenac
33. A. Douady, J.H. Hubbard, A proof of Thurston's topological characterization of rational functions. Acta Math. **171**, 263–297 (1993)
34. S. Dowdall, R.P. Kent, C.J. Leininger, Pseudo-anosov subgroups of fibered 3-manifold groups. Groups Geom. Dyn. **8**(4), 1247–1282 (2014)
35. A. Douady, Systèmes dynamiques holomorphes, in *Séminaire Bourbaki*, vol. 1982/1983. Astérisque, 105–106 (Soc. Math. France, Paris, 1983), pp. 39–63
36. P.L. Duren, *Univalent Functions*, vol. 259 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] (Springer, New York, 1983)
37. A. Epstein, T. Sharland, A classification of bicritical rational maps with a pair of period two superattracting cycles. Ann. Fac. Sci. Toulouse Math. (6) **21**(5), 907–934 (2012)
38. B. Farb, Relatively hyperbolic groups. Geom. Funct. Anal. **8**, 810–840 (1998)
39. A. Fathi, M. Laudenbach, V. Poenaru, Travaux de Thurston sur les surfaces. Asterisque **66–67**, 1–284 (1979)
40. P. Fatou, Sur les équations fonctionnelles. Bull. Soc. Math. France **47**, 161–271 (1919)
41. P. Fatou, Sur les équations fonctionnelles. Bull. Soc. Math. France **48**, 33–94 (1920)
42. P. Fatou, Sur les équations fonctionnelles. Bull. Soc. Math. France **48**, 208–314 (1920)
43. P. Fatou, Sur l'itération des fonctions transcendantes entières. Acta Math. **47**(4), 337–370 (1926)

44. P. Fatou, Notice sur les travaux scientifiques de M. P. Fatou. Astronome titulaire de l'observatoire de Paris, 5–29 (1929). https://www.math.purdue.edu/~eremenko/dvi/fatou-b.pdf
45. F. Gautero, Geodesics in trees of hyperbolic and relatively hyperbolic spaces. Proc. Edinburgh. Math. Soc. **59**(3), 701–740 (2016)
46. F. Gautero, M. Heusener, Cohomological characterization of relative hyperbolicity and combination theorem. Publ. Mat. **53**(2), 489–514 (2009)
47. F. Gautero, R. Weidmann, An algebraic combination theorem for graphs of relatively hyperbolic groups. Preprint, University of Nice (2011)
48. S.M. Gersten, Cohomological lower bounds for isoperimetric functions on groups. Topology **37**(5), 1031–1072 (1998)
49. M. Gromov, Hyperbolic groups, in *Essays in Group Theory*, ed. by S. Gersten, MSRI Publ., vol. 8 (Springer, 1985), pp. 75–263
50. S. Gupta, H. Seshadri, Complex geometry of Teichmüller domains, in *Handbook of Teichmüller Theory. Vol. VII*, ed. by A. Papadopoulos, volume 30 of IRMA Lect. Math. Theor. Phys. (Eur. Math. Soc., Zürich, 2020), pp. 63–88
51. M.F. Hagen, D.T. Wise, Cubulating hyperbolic free-by-cyclic groups: the general case. Geom. Funct. Anal. **25**(1), 134–179 (2015)
52. M.F. Hagen, D.T. Wise, Cubulating hyperbolic free-by-cyclic groups: the irreducible case. Duke Math. J. **165**(9), 1753–1813 (2016)
53. F. Haglund, D.T. Wise, Special cube complexes. Geom. Funct. Anal. **17**(5), 1551–1620 (2008)
54. F. Haglund, D.T. Wise, A combination theorem for special cube complexes. Ann. Math. (2) **176**(3), 1427–1482 (2012)
55. P. Haïssinsky, T. Lei, Convergence of pinching deformations and matings of geometrically finite polynomials. Fundam. Math. **181**(2), 143–188 (2004)
56. U. Hamenstädt, Word hyperbolic extensions of surface groups. preprint, arXiv:math/0505244 (2005)
57. J. Hempel, *3-Manifolds* (AMS Chelsea Publishing, Providence, 2004). Reprint of the 1976 original
58. G.C. Hruska, Relative hyperbolicity and relative quasiconvexity for countable groups. Algebr. Geom. Topol. **10**(3), 1807–1856 (2010)
59. T. Hsu, D.T. Wise, Cubulating graphs of free groups with cyclic edge groups. Amer. J. Math. **132**(5), 1153–1188 (2010)
60. T. Hsu, D.T. Wise, Cubulating malnormal amalgams. Invent. Math. **199**(2), 293–331 (2015)
61. J. Hubbard, *Teichmüller theory and applications to geometry, topology, and dynamics. Vol. 1* (Matrix Editions, Ithaca, 2006). Teichmüller theory, With contributions by Adrien Douady, William Dunbar, Roland Roeder, Sylvain Bonnot, David Brown, Allen Hatcher, Chris Hruska and Sudeb Mitra, With forewords by William Thurston and Clifford Earle
62. J. Hubbard, D. Schleicher, Multicorns are not path connected, in *Frontiers in Complex Dynamics*, volume 51 of Princeton Math. Ser. (Princeton Univ. Press, Princeton, 2014), pp. 73–102
63. H. Inou, S. Mukherjee, Non-landing parameter rays of the multicorns. Invent. Math. **204**(3), 869–893 (2016)
64. H. Inou, S. Mukherjee, Discontinuity of straightening in antiholomorphic dynamics: I. Trans. Amer. Math. Soc. **374**(9), 6445–6481 (2021)
65. G. Julia, Mémoire sur l'itération des fonctions rationnelles. J. Math. Pures Appl. **1**, 47–246 (1918)
66. G. Julia, Mémoire sur la permutabilité des fractions rationnelles. Ann. Sci. École Norm. Sup. (3) **39**, 131–215 (1922)
67. Y. Kabaya, Deformation of quasi-fuchsian groups (2016). https://www.math.kyoto-u.ac.jp/~kabaya/galleries/quasifuchsian_deformation/quasifuchsian_deformation.html
68. J. Kahn, V. Markovic, Immersing almost geodesic surfaces in a closed hyperbolic three manifold. Ann. Math. (2) **175**(3), 1127–1190 (2012)

69. I. Kapovich, The combination theorem and quasiconvexity. Int. J. Algebra Comput. **11**(2), 185–216 (2001)
70. M. Kapovich, *Hyperbolic Manifolds and Discrete Groups*, volume 183 of Progress in Mathematics (Birkhäuser, Boston, 2001)
71. M. Kapovich, P. Sardar, *Trees of Hyperbolic Spaces*. preprint, arXiv:2202.09526 (2022)
72. J. Kiwi, ℝeal laminations and the topological dynamics of complex polynomials. Adv. Math. **184**(2), 207–267 (2004)
73. F. Klein, Neue Beiträge zur Riemann'schen Functionentheorie. Math. Ann. **21**(2), 141–218 (1883)
74. K. Lazebnik, N.G. Makarov, S. Mukherjee, Univalent polynomials and Hubbard trees. Trans. Amer. Math. Soc. **374**(7), 4839–4893 (2021)
75. K. Lazebnik, N.G. Makarov, S. Mukherjee, Bers slices in families of univalent maps. Math. Z. (2021). https://doi.org/10.1007/s00209-021-02871-y
76. S.-Y. Lee, N. Makarov, Sharpness of connectivity bounds for quadrature domains (2014, November). https://arxiv.org/abs/1411.3415v1
77. S.Y. Lee, N.G. Makarov, Topology of quadrature domains. J. Amer. Math. Soc. **29**(2), 333–369 (2016)
78. S.-Y. Lee, M. Lyubich, N.G. Makarov, S. Mukherjee, Dynamics of Schwarz reflections: the mating phenomena (2018). https://arxiv.org/abs/1811.04979v2
79. S.-Y. Lee, M. Lyubich, N.G. Makarov, S. Mukherjee, Schwarz reflections and the Tricorn (2018). https://arxiv.org/abs/1812.01573v1
80. S.-Y. Lee, M. Lyubich, N.G. Makarov, S. Mukherjee, Schwarz reflections and anti-holomorphic correspondences. Adv. Math. **385**, paper no. 107766, 88 pp. (2021)
81. T. Lei, Matings of quadratic polynomials. Ergodic Theory Dyn. Syst. **12**(3), 589–620 (1992)
82. T. Lei, Branched coverings and cubic Newton maps. Fundam. Math. **154**(3), 207–260 (1997)
83. L. Li, K. Ohshika, X. Wang, On Klein-Maskit combination theorem in space. I. Osaka J. Math. **46**(4), 1097–1141 (2009)
84. L. Li, K. Ohshika, X. Wang, On Klein-Maskit combination theorem in space II. Kodai Math. J. **38**(1), 1–22 (2015)
85. R. Lodge, Y. Luo, S. Mukherjee, Circle packings, kissing reflection groups and critically fixed anti-rational maps. Forum Math. Sigma **10**, paper no. e3, 38 pp. (2022)
86. Y. Luo, Limits of rational maps, R-trees and barycentric extension. Adv. Math. **393**, paper no. 108075, 46 pp. (2021)
87. Y. Luo, On geometrically finite degenerations I: boundaries of main hyperbolic components (2021). https://arxiv.org/abs/2101.07880
88. Y. Luo, On geometrically finite degenerations II: convergence and divergence (2021). https://arxiv.org/abs/2102.00357
89. M. Lyubich, *Conformal Geometry and Dynamics of Quadratic Polynomials, vol I-II* (2022, in preparation). http://www.math.stonybrook.edu/~mlyubich/book.pdf
90. M. Lyubich, Y. Minsky, Laminations in holomorphic dynamics. J. Differential Geom. **47**, 17–94 (1997)
91. M. Lyubich, S. Merenkov, S. Mukherjee, D. Ntalampekos, David extension of circle homeomorphisms, welding, mating, and removability (2020). https://arxiv.org/abs/2010.11256v2
92. J. Manning, M. Mj, M. Sageev, Cubulating surface-by-free groups. Preprint, arxiv:1908.03545 (2019), p. 46
93. A. Martin, D. Osajda, A combination theorem for combinatorially non-positively curved complexes of hyperbolic groups, in *Mathematical Proceedings of the Cambridge Philosophical Society* (2021)
94. I. Mashanova, V. Timorin, Captures, matings and regluings. Ann. Fac. Sci. Toulouse Math. (6) **21**(5), 877–906 (2012)
95. B. Maskit, Construction of Kleinian groups, in *Proc. Conf. Complex Analysis (Minneapolis, 1964)* (Springer, Berlin, 1965), pp. 281–296
96. B. Maskit, On Klein's combination theorem. Trans. Amer. Math. Soc. **120**, 499–509 (1965)

97. B. Maskit, On Klein's combination theorem. II. Trans. Amer. Math. Soc. **131**, 32–39 (1968)
98. B. Maskit, On Klein's combination theorem. III, in *Advances in the Theory of Riemann Surfaces (Proc. Conf., Stony Brook, N.Y., 1969)*, Ann. of Math. Studies, No. 66 (Princeton Univ. Press, Princeton, 1971), pp. 297–316
99. B. Maskit, *Kleinian Groups*. Springer Grundlehren 287 (1980)
100. B. Maskit, On Klein's combination theorem. IV. Trans. Amer. Math. Soc. **336**(1), 265–294 (1993)
101. C.T. McMullen, Iteration on Teichmuller space. Invent. Math. **9**, 425–454 (1990)
102. C.T. McMullen, *Renormalization and 3-Manifolds which Fiber over the Circle* (Princeton University Press, Princeton, 1998)
103. D. Meyer, Expanding Thurston maps as quotients (2009). https://arxiv.org/abs/0910.2003
104. D. Meyer, Unmating of rational maps: sufficient criteria and examples, in *Frontiers in Complex Dynamics*, volume 51 of Princeton Math. Ser. (Princeton Univ. Press, Princeton, 2014), pp. 197–233
105. J. Milnor, Periodic orbits, externals rays and the Mandelbrot set: an expository account, in *Géométrie complexe et systèmes dynamiques (Orsay, 1995)*, volume 261 (2000), pp. xiii, 277–333. Astérisque
106. J. Milnor, Pasting together Julia sets: a worked out example of mating. Exp. Math. **13**(1), 55–92 (2004)
107. J. Milnor, *Dynamics in One Complex Variable*, 3rd edn. Annals of Mathematics Studies, vol. 160 (Princeton University Press, Princeton, 2006)
108. M. Mitra, Cannon-Thurston maps for hyperbolic group extensions. Topology **37**(3), 527–538 (1998)
109. M. Mitra, Cannon-Thurston maps for trees of hyperbolic metric spaces. J. Differential Geom. **48**(1), 135–164 (1998)
110. M. Mj, Cannon-Thurston maps for surface groups. Ann. Math. (2) **179**(1), 1–80 (2014)
111. M. Mj, Ending laminations and Cannon-Thurston maps, with an appendix by S. Das and M. Mj. Geom. Funct. Anal. **24**(1), 297–321 (2014)
112. M. Mj, Cannon-Thurston maps for Kleinian groups. Forum Math. Pi **5**, paper no. e1, 49 pp. (2017)
113. M. Mj, Tight trees and model geometries of surface bundles over graphs. J. Lond. Math. Soc. (2) **102**(3), 1178–1222 (2020)
114. M. Mj, S. Mukherjee, Combining rational maps and Kleinian groups via orbit equivalence (2021). https://arxiv.org/abs/2103.00835v1
115. M. Mj, L. Reeves, A combination theorem for strong relative hyperbolicity. Geom. Topol. **12**(3), 1777–1798 (2008)
116. M. Mj, P. Sardar, A combination theorem for metric bundles. Geom. Funct. Anal. **22**(6), 1636–1707 (2012)
117. J.W. Morgan, P.B. Shalen, Valuations, trees, and degenerations of hyperbolic structures. I. Ann. Math. (2) **120**(3), 401–476 (1984)
118. J.W. Morgan, P.B. Shalen, Degenerations of hyperbolic structures. II. Measured laminations in 3-manifolds. Ann. Math. (2) **127**(2), 403–456 (1988)
119. J.W. Morgan, P.B. Shalen, Degenerations of hyperbolic structures. III. Actions of 3-manifold groups on trees and Thurston's compactness theorem. Ann. Math. (2) **127**(3), 457–519 (1988)
120. S. Mukherjee, S. Nakane, D. Schleicher, On multicorns and unicorns II: bifurcations in spaces of antiholomorphic polynomials. Ergodic Theory Dyn. Syst. **37**(3), 859–899 (2017)
121. J.P. Otal, Thurston's hyperbolization of Haken manifolds, in *Surveys in Differential Geometry, Vol. III (Cambridge, MA, 1996)* (1998), pp. 77–194
122. J.P. Otal, *The Hyperbolization Theorem for Fibered 3-manifolds*, volume 7 of SMF/AMS Texts and Monographs. (American Mathematical Society/Société Mathématique de France, Providence/Paris, 2001). Translated from the 1996 French original by Leslie D. Kay
123. C.L. Petersen, D. Meyer, On the notions of mating. Ann. Fac. Sci. Toulouse Math. (6) **21**(5), 839–876 (2012)

124. K.M. Pilgrim, *Combinations of Complex Dynamical Systems*. Lecture Notes in Mathematics (Springer, 2003)
125. K.M. Pilgrim, T. Lei, Combining rational maps and controlling obstructions. Ergodic Theory Dyn. Syst. **18**(1), 221–245 (1998)
126. P. Przytycki, D.T. Wise, Mixed 3-manifolds are virtually special. J. Amer. Math. Soc. **31**(2), 319–347 (2018)
127. M. Rees, A partial description of parameter space of rational maps of degree two. I. Acta Math. **168**(1–2), 11–87 (1992)
128. M. Rees, Multiple equivalent matings with the aeroplane polynomial. Ergodic Theory Dyn. Syst. **30**(4), 1239–1257 (2010)
129. M. Rees, One hundred years of complex dynamics. Proc. A. **472**(2185), 20150453, 16 (2016)
130. P. Roesch, Introduction [Workshop on Polynomial Matings]. Ann. Fac. Sci. Toulouse Math. (6) **21**(5), iii–x (2012). Held in Toulouse, June 8–11, 2011
131. M. Sageev, Ends of group pairs and non-positively curved cube complexes. Proc. Lond. Math. Soc. (3) **71**(3), 585–617 (1995)
132. M. Sakai, Regularity of a boundary having a Schwarz function. Acta Math. **166**(3–4), 263–297 (1991)
133. P. Scott, C.T.C. Wall, Topological methods in group theory, in *Homological Group Theory*, ed. by C.T.C. Wall, London Math. Soc. Lecture Notes Series, vol. 36 (Cambridge Univ. Press, Cambridge, 1979)
134. Z. Sela, Acylindrical accessibility for groups. Invent. Math. **129**(3), 527–565 (1997)
135. T. Sharland, Matings of cubic polynomials with a fixed critical point, Part I: Thurston obstructions. Conform. Geom. Dyn. **23**, 205–220 (2019)
136. M. Shishikura, On a theorem of M. Rees for matings of polynomials, in *The Mandelbrot Set, Theme and Variations*, volume 274 of London Math. Soc. Lecture Note Ser. (Cambridge Univ. Press, Cambridge, 2000), pp. 289–305
137. M. Shishikura, T. Lei, A family of cubic rational maps and matings of cubic polynomials. Exp. Math. **9**(1), 29–53 (2000)
138. C.L. Siegel, Iteration of analytic functions. Ann. Math. (2) **43**, 607–612 (1942)
139. R.K. Skora, Splittings of surfaces. J. Amer. Math. Soc. **9**(2), 605–616 (1996)
140. D. Sullivan, Quasiconformal homeomorphisms and dynamics I: Solution of the Fatou-Julia problem on wandering domains. Ann. Math. (2) **122**, 401–418 (1985)
141. D. Sullivan, C.T. McMullen, Quasiconformal homeomorphisms and dynamics III: The Teichmüller space of a holomorphic dynamical system. Adv. Math. **135**, 351–395 (1998)
142. W.P. Thurston, *The Geometry and Topology of 3-Manifolds*. Princeton University Notes (1980)
143. W.P. Thurston, Hyperbolic structures on 3-manifolds,I: deformations of acylindrical manifolds. Ann. Math. (2) **124**, 203–246 (1986)
144. W.P. Thurston, Hyperbolic structures on 3-manifolds,II: surface groups and 3-manifolds which fiber over the circle. Preprint, arXiv:math.GT/9801045 (1986)
145. W.P. Thurston, Hyperbolic structures on 3-manifolds,III: deformations of 3-manifolds with incompressible boundary. Preprint (1986)
146. È.B. Vinberg, O.V. Shvartsman, Discrete groups of motions of spaces of constant curvature, in *Geometry, II*, volume 29 of Encyclopaedia Math. Sci. (Springer, Berlin, 1993), pp. 139–248
147. D.T. Wise, The residual finiteness of negatively curved polygons of finite groups. Invent. Math. **149**(3), 579–617 (2002)
148. D.T. Wise, *From Riches to Raags: 3-manifolds, Right-angled Artin Groups, and Cubical Geometry*, volume 117 of CBMS Regional Conference Series in Mathematics. Published for the Conference Board of the Mathematical Sciences, Washington, DC (American Mathematical Society, Providence, 2012)
149. D.T. Wise, *The Structure of Groups with a Quasiconvex Hierarchy*. Annals of Mathematics Studies, vol. 209 (Princeton University Press, Princeton, 2021)
150. M. Yampolsky, S. Zakeri, Mating Siegel quadratic polynomials. J. Amer. Math. Soc. **14**(1), 25–78 (2001)

Chapter 11
On the Pullback Relation on Curves Induced by a Thurston Map

Kevin M. Pilgrim

Abstract Via taking connected components of preimages, a Thurston map $f : (S^2, P_f) \to (S^2, P_f)$ induces a pullback relation on the set of isotopy classes of curves in the complement of its postcritical set P_f. We survey known results about the dynamics of this relation, and pose some questions.

Keywords Thurston map · Teichmüller space · Curves

2010 Mathematics Subject Classification 37F10

11.1 Introduction

An orientation-preserving branched covering $f : S^2 \to S^2$ of degree at least two is a *Thurston map* if its *postcritical set* $P_f = \cup_{n>0} f^n(C_f)$ is finite, where C_f is the finite set of branch (critical) points at which f fails to be locally injective.

A fundamental theorem in complex dynamics is Thurston's Characterization and Rigidity Theorem [14]. It asserts that apart from a well-understood set of counterexamples, (1) a Thurston map f is conjugate-up-to-isotopy-relative-to-P_f (or "equivalent") to a rational map R if and only if it has no "obstructions", and (2) if f has no obstructions, R is unique, up to holomorphic conjugacy. An obstruction is a multicurve (a finite collection of pairwise disjoint simple closed curves in $S^2 - P_f$, up to isotopy) with certain invariance properties.

Suppose $P \subset S^2$ is finite. The set of isotopy classes relative to P of Thurston maps f for which $P_f \subset P$ admits the structure of a countable semigroup under composition; we denote this by $\mathrm{BrMod}(S^2, P)$. Pre- and post-composition with homeomorphisms fixing P gives this semigroup the additional structure of a biset over the mapping class group $\mathrm{Mod}(S^2, P)$. In this way, $\mathrm{BrMod}(S^2, P)$ may be

K. M. Pilgrim (✉)
Indiana University, Bloomington, IN, USA
e-mail: pilgrim@indiana.edu

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_11

fruitfully thought of as a generalization of the mapping class group. This perspective is useful in developing intuition for the range of potential behavior of and structure theory for Thurston maps.

The mapping class group of a surface acts naturally on the countably infinite set of isotopy classes of curves on the surface. Even better, it acts on the associated curve complex; see [18]. It is natural to try to do something similar for Thurston maps. Since the set P_f contains the branch values of f, the restriction $f : S^2 - f^{-1}(P_f) \to S^2 - P_f$ is a covering map. It follows that a component $\widetilde{\gamma}$ of the inverse image $f^{-1}(\gamma)$ of a simple closed curve γ in $S^2 - P_f$ is a simple closed curve in $S^2 - f^{-1}(P_f)$. Since P_f is forward-invariant, we have an inclusion $S^2 - f^{-1}(P_f) \hookrightarrow S^2 - P_f$, so the curve $\widetilde{\gamma}$ is again a simple closed curve in $S^2 - P_f$. Abusing terminology, we'll call $\widetilde{\gamma}$ a *preimage* of γ, or sometimes say γ *lifts*, or *pulls back*, to $\widetilde{\gamma}$. By lifting isotopies, we obtain a pullback relation $\stackrel{f}{\leftarrow}$ on the set $\overline{\mathcal{C}}$ of such simple closed curves up to isotopy. The curve γ might have several preimages, so we obtain an induced relation instead of a function. A preimage of an inessential curve is again inessential. Similarly, a preimage of a peripheral curve—one which is essential and isotopic into any small neighborhood of a single point in P_f—is either again peripheral, or is inessential. We call inessential and peripheral curves *trivial*, and note that the set of trivial curves is invariant under the pullback relation.

When $\#P_f = 4$, the pullback relation induces—almost—a function on the set of nontrivial curves. On the one hand, distinct nontrivial curves in this case must intersect. On the other hand, distinct components of $f^{-1}(\gamma)$ are in general disjoint. It follows that there can be at most one class of nontrivial preimage, and we almost get a function in this case. Why "almost"? Typical examples have the property that for some curve, each of its preimages are trivial. So while the mapping class group acts naturally on e.g. the infinite diameter curve complex, it is less clear how to construct a nice complex related to curves on which a Thurston map acts via pullback. This relative lack of preserved structure makes answering even basic questions challenging.

This survey presents some known results about the dynamical behavior of taking iterated preimages of curves under a given Thurston map. It assumes the basic vocabulary related to Thurston maps from [14], and the reader may find [21] also useful for more detailed explanations and references to some terminology encountered along the way.

Here are some highlights, to convince you that this is interesting. When $f(z) = z^2+c$ is the so-called *Douady Rabbit* polynomial, where c is chosen so that $\Im(c) > 0$ and the origin has period 3, under iteration every curve pulls back to either a trivial curve, or to the prominent 3-cycle. But when $f(z) = z^2 + i$, every curve pulls back eventually to a trivial curve. See [29]. In these two cases, we see that there is a *finite global attractor* for the pullback relation. That is, there is a finite set of curves such that under iterated pullback, each curve either iterates into this set, or to the trivial curve; see the next section for a precise definition. Among obstructed Thurston maps, though, it is easy to manufacture examples with wandering curves and infinitely many fixed curves. A basic conjecture is

Conjecture 11.1.1 If f is rational and not a flexible Lattès example, then the pullback relation on curves has a finite global attractor.

The flexible Lattès examples are ubiquitous counterexamples to general statements in complex dynamics. They arise as follows. Suppose $\Lambda < \mathbb{C}$ is a lattice, and $X = \mathbb{C}/\Lambda$ is the corresponding complex torus, regarded as both a Riemann surface and an abelian group. Via the corresponding Weierstrass $\wp$-function, the quotient $X/(x \sim -x)$ is the Riemann sphere $\widehat{\mathbb{C}}$. For an integer $n \geq 2$, the endomorphism $x \mapsto n \cdot x$ of X descends to a degree n^2 self-map f of $\widehat{\mathbb{C}}$ with $\#P_f = 4$. It is easily seen that the pullback relation on nontrivial curves is the identity map. For fixed n, the quasiconformal conjugacy class of f is independent of the choice of Λ, hence the term "flexible". More general Lattès examples arise as quotients of other holomorphic endomorphisms of complex tori by more complicated finite groups; in this case, the lattice Λ must be more special. For all Lattès examples f, we have $\#P_f \leq 4$.

Here, informally, is the source of the tension in Conjecture 11.1.1. Equipping $\widehat{\mathbb{C}} - P_f$ with its hyperbolic metric, we may represent each element of $\mathcal{C}$ by a unique geodesic. Pulling back and lifting the metric to $\widehat{\mathbb{C}} - f^{-1}P_f$, the lifted curve $\widetilde{\gamma}$ may unwind and become up to $\deg(f)$ times as long as γ. But when including the curve $\widetilde{\gamma}$ back into $\widehat{\mathbb{C}} - P_f$, the Schwarz-Pick lemma implies the length of $\widetilde{\gamma}$ shrinks. It is unclear which force—lengthening or shortening—dominates in the long run. And since there exist expanding non-rational examples with wandering curves (see Sect. 4.4 below), the exact mechanism that would imply the conjecture remains mysterious.

11.2 Conventions and Notation

Throughout, f denotes a Thurston map, P its postcritical set, and d its degree. To avoid repeated mention of special cases, unless otherwise stated, f has hyperbolic orbifold and $\#P \geq 4$. In particular, f is not a Lattès example. We denote by

- $\simeq$ the equivalence relation of isotopy-through-Thurston-maps-with-fixed-postcritical-set P;
- $\mathcal{C}$, the countably infinite set of isotopy classes of unoriented, essential, simple, nonperipheral curves in $S^2 - P$ (we will often call such elements simply "curves", abusing terminology); on $\mathcal{C}$ we have the *geometric intersection number* $\iota(\alpha, \beta)$ which counts the minimum number of intersection points among representatives;
- o, the union of the $\#P+1$ isotopy classes of unoriented, simple, closed, peripheral and inessential curves in $S^2 - P$, i.e. the trivial ones;
- $\overline{\mathcal{C}} := \mathcal{C} \cup \{o\}$;
- $\stackrel{f}{\leftarrow}$, the pullback relation on $\overline{\mathcal{C}}$ induced by $\gamma \stackrel{f}{\leftarrow} \delta \subset f^{-1}(\gamma)$, where $[\gamma] \in \overline{\mathcal{C}}$ and δ is a component of $f^{-1}(\gamma)$;
- $\overline{\mathcal{A}}$ and $\mathcal{A}$, the set of curves contained in cycles of $\stackrel{f}{\leftarrow}$ in $\overline{\mathcal{C}}$ and $\mathcal{C}$, respectively;

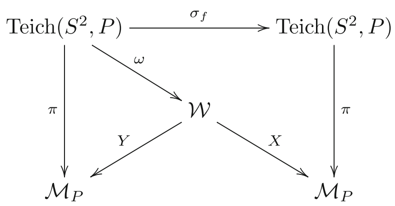


Fig. 11.1 The fundamental diagram

- $\mathcal{W} \subset \mathcal{C}$, the set of "wandering" curves γ_0, namely, those for which there is an infinite sequence $\gamma_n, n \geq 0$, of distinct nontrivial curves satisfying $\gamma_n \overset{f}{\leftarrow} \gamma_{n+1}, n \geq 0$;
- the relation $\overset{f}{\leftarrow}$ has a *finite global attractor* if $\mathcal{W}$ is empty and $\mathcal{A}$ is finite;
- the moduli space $\mathcal{M}_P = \{P \hookrightarrow \widehat{\mathbb{C}}\}/\text{Aut}(\widehat{\mathbb{C}})$ the space of injections of P into the Riemann sphere, up to post-composition with Möbius transformations;
- Teich(S^2, P), the Teichmüller space of the sphere marked at the set P;
- σ_f : Teich(S^2, P) $\to$ Teich(S^2, P), the holomorphic self-map obtained by pulling back complex structures; it is the lift to the universal cover of an algebraic *correspondence on moduli space* $X \circ Y^{-1}$, where Y is a finite cover and X is holomorphic. See Fig. 11.1.
- Mod(S^2, P) and PMod(S^2, P), the mapping class group and pure mapping class group, respectively.
- The covering Y is induced by a finite-index subgroup $H_f <$ PMod(S^2, P) characterized by the property that $hf \simeq f\tilde{h}$ for some $\tilde{h} \in$ PMod(S^2, P); the map $h \mapsto \tilde{h} =: \phi_f(h)$ is the *virtual endomorphism* of PMod(S^2, P) induced by f.
- A *nearly Euclidean Thurston map* (NET map) is one for which $\#P_f = 4$ and each critical point has local degree two; see [11] and also [16] for a survey.

11.3 Non-dynamical Properties of $\overset{f}{\leftarrow}$

A Thurston map $f : (S^2, P) \to (S^2, P)$ may factor as a composition of maps of pairs

$$(S^2, P) \overset{f_1}{\to} (S^2, C) \overset{f_2}{\to} (S^2, P)$$

where each f_i is admissible in the sense that its set of branch values is contained in the distinguished subset appearing in its codomain. This motivates studying properties of so-called admissible branched covers $f : (S^2, A) \to (S^2, B)$ with $f(A) \subset B \supset f(C_f)$ where domain and codomain are no longer identified; this perspective was introduced by S. Koch. Instead of a pullback self-relation on curves $\mathcal{C}$, we have a pullback relation $\mathcal{C}_B \overset{f}{\leftarrow} \mathcal{C}_A$ from classes of curves in $S^2 - B$ to classes in $S^2 - A$. The virtual endomorphism becomes a virtual homomorphism $\phi_f : \mathrm{PMod}(S^2, B) \to \mathrm{PMod}(S^2, A)$.

11.3.1 Known General Results

Thinking non-dynamically first, we have the following known results about the pullback relation $\stackrel{f}{\leftarrow}$.

1. When $\#A = \#B$, each nonempty fiber of $\stackrel{f}{\leftarrow}$ is dense in the Thurston boundary; in particular, each nontrivial fiber is infinite [21]. Here, by the fiber over β, we mean $\{\alpha : \alpha \stackrel{f}{\leftarrow} \beta\}$.
2. The relation $\stackrel{f}{\leftarrow}$ can be trivial in the sense that the only pairs are of the form $\gamma \stackrel{f}{\leftarrow} \delta$ where δ is trivial. Equivalently, σ_f is constant. See [21], correcting an argument appearing originally in [8].
3. The relation $\stackrel{f}{\leftarrow}$ satisfies a Lipschitz-type inequality related to intersection numbers: $\iota(\widetilde{\alpha}, \widetilde{\beta}) \leq d \cdot \iota(\alpha, \beta)$ whenever $\alpha \stackrel{f}{\leftarrow} \widetilde{\alpha}, \beta \stackrel{f}{\leftarrow} \widetilde{\beta}$. To see this, we represent the pair α, β by minimally intersecting curves, and note that the full preimage of their intersection has at most the indicated cardinality.

   The study of the interaction between intersection numbers and the geometry of σ_f seems to be just beginning. Implicit use of such interactions appears in the analysis of NET maps by W. Floyd, W. Parry, and this author; see [11, 17]. Parry develops this intersection theory further in [26]. Intersection theory has been fruitfully applied to control the possible locations of obstructions in the formulation of surgery operations; see [6, 13, 30].
4. The set of nonempty multicurves is in natural bijective correspondence with boundary strata in the augmented Teichmüller space, which by a theorem of Masur is known to be the completion of Teichmüller space in the Weil-Petersson metric [24]. A result of Selinger [32] shows that $\sigma_f : \mathrm{Teich}(S^2, B) \to \mathrm{Teich}(S^2, A)$ extends to this completion, sending the stratum corresponding to a multicurve Γ to the stratum corresponding to the multicurve $f^{-1}(\Gamma)$. It follows that analytical tools for studying σ_f can be used to study properties of the combinatorial relation $\stackrel{f}{\leftarrow}$ [21, 29]. There is thus a rich interplay between the analytic and algebro-geometric properties of the correspondence on moduli space, and the combinatorial properties of the pullback relation; see [31], for example.
5. Associated to a nonempty multicurve $\Gamma \subset S^2 - B$ is the free abelian group $\mathrm{Tw}(\Gamma)$ of products of powers of Dehn twists about the curves in Γ. It is easy to see that $\phi_f(\mathrm{Tw}(\Gamma)) < \mathrm{Tw}(f^{-1}\Gamma)$. Hence the pullback relation on curves can be encoded using the associated induced virtual homomorphism. It follows that group-theoretic tools can also be used to study properties of the pullback relation on curves; see [29] and [20].

Question 11.3.1 If the pullback relation $\stackrel{f}{\leftarrow}$ is not trivial, must it be surjective?

It seems very likely that the answer is no, for the following reason. The Composition Trick, introduced in the next subsection, should allow one to build examples where the image of σ_f has positive dimension and codimension, so that its the closure of its image misses many WP boundary strata.

11.3.2 Mechanisms for Triviality of $\stackrel{f}{\leftarrow}$

There seem to be three or four mechanisms via which $\stackrel{f}{\leftarrow}$ can be trivial.

1. **Composition trick.** The map $f : (S^2, A) \to (S^2, B)$ may factor through (S^2, C) with $\#C = 3$ (C. McMullen, [8]). Such maps have the property that σ_f and $\stackrel{f}{\leftarrow}$ are trivial.
2. **NET maps.** A. Saenz [23] found an example of a Thurston map f for which σ_f is constant but for which f does not decompose as in the Composition Trick. Here is his example, from a different point of view.

   Let $X = \mathbb{C}/\Lambda$ be a complex torus, regarded again as a Riemann surface and an abelian group. There are 8 distinct points of order 3; under the involution $x \mapsto -x$ these 8 points descend to a set A of 4 points on $\widehat{\mathbb{C}}$ whose cross-ratio is, miraculously, constant as Λ varies. Now take X to be the square torus and let $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ be the degree 9 flexible Lattès map induced by the tripling map on X, and let B be the image of the set of four fixed-points of $x \mapsto -x$ under projection to $\widehat{\mathbb{C}}$. Note that $f(A)$ is a single element of B, corresponding to the image in $\widehat{\mathbb{C}}$ of the identity element. As Λ varies, the conformal shape of B varies, but that of A does not. Thus σ_f is constant and so $\stackrel{f}{\leftarrow}$ is trivial. One can see this triviality directly by observing that the action of $\mathrm{PSL}_2(\mathbb{Z})$ is transitive on curves (since it acts transitively on extended rationals regarded as slopes), that A is invariant under this action (since points of order 3 are invariant under the induced group-theoretic automorphisms), and that the horizontal curve has all preimages inessential or peripheral (as drawing a single easy picture shows). W. Parry (personal communication) has classified which NET maps with σ_f constant arise via the Composition Trick. He reports the empirical finding that an exhaustive search of low-complexity examples reveals that Saenz' example is among a very small handful of five sporadic cases.
3. **Other sporadic examples.** Let f be the unique (up to pre- and post-composition by independent automorphisms) degree four rational map with three double critical points (c_1, c_2, c_3) mapping to necessarily distinct critical values (v_1, v_2, v_3). Choose w a point distinct from the v_i's, let $B = \{v_1, v_2, v_3, w\}$ and $A = R^{-1}(w) = \{z_1, z_2, z_3, z_4\}$. Then the j-invariant (obtained from the cross-ratio by applying a certain degree six rational function) of the z_i's is constant in w, whence σ_f is constant and so $\stackrel{f}{\leftarrow}$ is trivial. To see this, note that as w approaches

some v_i, three of the z_i's equidistribute around and converge to c_i, and the remaining one converges to some other point, call it c_i'. Normalizing so $c_i = 0$ and $c_i' = \infty$ and scaling via multiplication with a nonzero complex constant shows that the conformal shape of the fiber $R^{-1}(w)$ converges to that of the cube roots of unity together with the point at infinity. Thus the j-invariant of $R^{-1}(w)$ is a bounded holomorphic function, hence constant.

4. **Combinations of the above.**

Question 11.3.2 Do there exist examples f with σ_f constant and $\deg(f)$ prime?

Cui G. has thought about the general case, see [12]. It is natural to look for the simplest such examples. By cutting along a maximal multicurve, one may restrict to the case $\#A = \#B = 4$; let's call these "minimal". It is natural to look for examples which do not factor as in the Composition Trick; let's call these "primitive".

Question 11.3.3 What are the minimal primitive branched covers $f : (S^2, A) \to (S^2, B)$ for which $\stackrel{f}{\leftarrow}$ is trivial?

11.3.3 Computation of $\stackrel{f}{\leftarrow}$

Though the set of curves $\mathcal{C}$ is complicated, it is conveniently described by a variety of coordinate systems. For example, Dehn-Thurston coordinates record intersection numbers of curves with edges in a fixed triangulation with vertex set P [15]. Train tracks and measured foliations give other methods. But expressing the pullback relation in these coordinates can be very complicated. The effect on Dehn-Thurston coordinates of pulling back from $S^2 - P_f$ to $S^2 - f^{-1}(P_f)$ is indeed easy to compute, since one can just lift intersection numbers. But the formula for the induced "puncture erasing map" is hard to write down in closed form and leads to continued-fraction-like cases.

When $\#P_f = 4$, though, the set of curves $\mathcal{C}$ can be encoded by "slopes" in the extended rationals $\mathbb{Q} \cup \{1/0\}$, the pullback relation $\stackrel{f}{\leftarrow}$ is a almost a function, and things are a bit easier but are still quite complicated.

If f is an NET map, there is an algorithm that computes the image of a slope under pullback. This can be done easily by hand, and has been implemented. NET maps can be easily encoded by combinatorial input. W. Parry has written a computer program that implements this algorithm. The website http://intranet.math.vt.edu/netmaps/, maintained by W. Floyd, contains a database of tens of thousands of examples. For NET maps, it appears that this ability to calculate the pullback map on curves (and related invariants, such as the degree by which preimages map, and how many preimages there are) leads to an effective algorithm for determining whether a given example is, or is not, equivalent to a rational map; see [17].

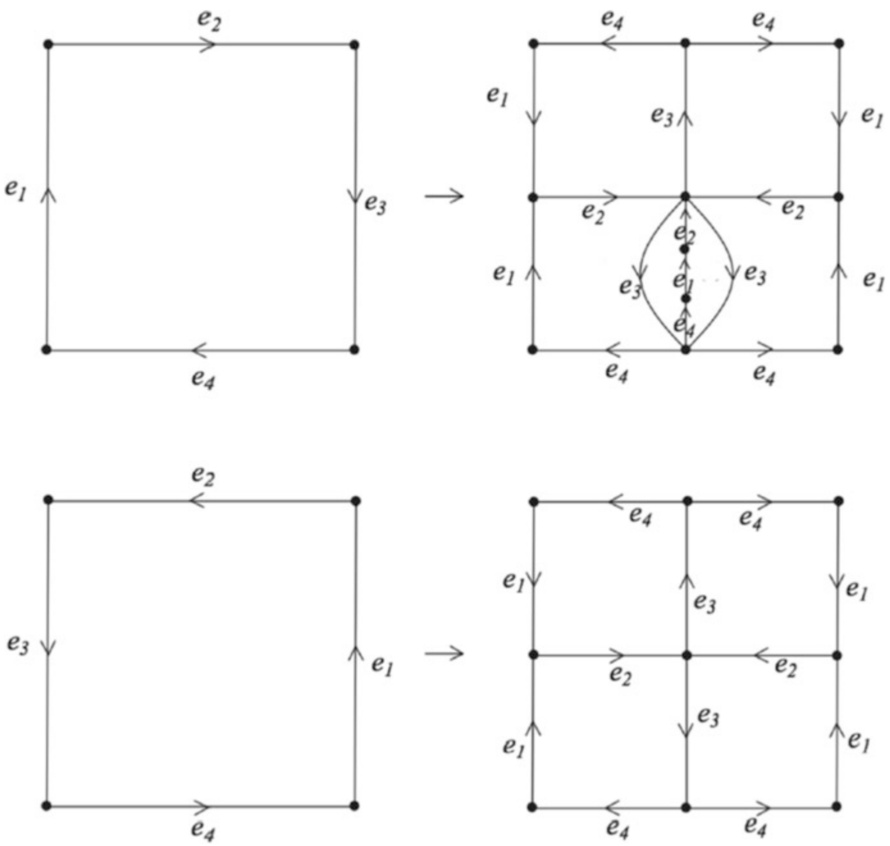


Fig. 11.2 The sphere is the union of the two squares at left along their boundaries as indicated to form a square "pillowcase". Each big square at right is identified with the square just to its left by a translation, so that the pillowcases are identified. The figure shows a cell structure in domain and codomain. The map f goes in the opposite direction to the indicated arrow and defines a cellular degree 5 map from the pillowcase to itself. The four corners of the pillowcase form the postcritical set. *Figure by W. Floyd*

When $\#P_f = 4$ and f is the subdivision map of a subdivision rule of the square pillowcase (like in Fig. 11.2), W. Parry has written a program for computing the image of a slope under pullback (personal communication).

Question 11.3.4 Are there any interesting settings in which one can effectively compute $\xleftarrow{f}$ when $\#P_f \geq 5$?

11.3.4 When Each Curve Has a Nontrivial Preimage

The example studied by Lodge [22] is, nondynamically speaking, the unique generic cubic, in the following sense. We have $f : (S^2, A) \to (S^2, B)$ where A is the set of four simple critical points and $B = f(A)$ is the corresponding set of four critical values. Each nontrivial curve has exactly one nontrivial curve in its preimage.

Question 11.3.5 Suppose $f : (S^2, A) \to (S^2, B)$ has the property that A consists of $2d-2$ simple critical points, $B = f(A)$, and $f|A$ is injective, so that B consists of the $2d-2$ critical values. Does each nontrivial curve in $S^2 - B$ have a nontrivial preimage?

Up to pre- and post-composition with independent homeomorphisms there is a unique such map [4]. If each curve γ has a nontrivial preimage, so does $h(\gamma)$, where $h \in H_f$ lifts under f. Since H_f is a finite-index subgroup of $\mathrm{Mod}(S^2, P_f)$, determining the answer to this question is a finite computation.

In the case of four postcritical points, we have the following result from [17, Thm. 4.1].

Theorem 11.3.6 *If* $\#P = 4$*, the pullback function on curves is either surjective, or trivial.*

The key insight: when $\#P = 4$, looking at the correspondence on moduli space, the space $\mathcal{W}$ is a Riemann surface whose set of ends consists of finitely many cusps, the map $X : \mathcal{W} \to \mathcal{M}_P$ is holomorphic, and $\mathcal{M}_P$ is the triply-punctured sphere; thus if X is nonconstant, then each cusp of $\mathcal{M}_P$ is the image of a cusp of $\mathcal{W}$.

11.4 Dynamical Properties

11.4.1 General Properties

Here, we discuss some examples and known results about the possible dynamical behavior of $\stackrel{f}{\leftarrow}$.

1. **Example:** *Every curve iterates to the trivial curve.* This happens for $z^2 + i$. Here is one way to see this. Examining the possibilities for how the bounded region enclosed by a curve meets the finite postcritical set $\{i, i-1, -i\}$, one sees that a curve must eventually become trivial unless it surrounds both $-i$ and $i-1$. For this type of curve α, there is at most one nontrivial curve β with $\alpha \stackrel{f}{\leftarrow} \beta$ and β a curve of the same type. Moreover, $\deg(\alpha \stackrel{f}{\leftarrow} \beta) = 1$. Equipping the complement of the postcritical set with the hyperbolic metric, the Schwarz-Pick Lemma shows that the length of a geodesic representative of β is strictly shorter than that of α. Iterating this process, it follows that such a curve cannot be periodic or wandering under $\stackrel{f}{\leftarrow}$: curves in its orbit cannot get too complicated,

since otherwise they would have to get long, so they must eventually become a different type of curve and thus become trivial upon further iteration.

The "airplane" quadratic polynomial $f(z) = z^2 + c$, with the origin periodic of period 3 and $\mathrm{Im}(c) = 0$, is another such example [20].

2. Another very natural question is

   Question 11.4.1 Does there exist an example of a Thurston map f for which the pullback relation induced by f is nontrivial but that induced by some iterate f^n is trivial?

3. **Theorem** If f is rational and non-Lattès, then there are only finitely many multicurves for which $f^{-1}\Gamma = \Gamma$; see [29, Thm. 1.5]. The proof uses the decomposition theory.
4. *Conjecture 11.1.1* If f is rational and not a flexible Lattès example then the pullback relation on curves has a finite global attractor.

   There is partial progress on this conjecture for special families of Thurston maps.

   (a) Kelsey and Lodge [20] verify this for all quadratic non-Lattès maps with four postcritical points.
   (b) Hlushchanka [19] verifies this for rational maps each of whose critical points is fixed. Such maps up to holomorphic conjugacy are classified by connected planar multigraphs, up to planar isomorphism, and the global attractor can be explicitly identified in terms of the multigraph.
   (c) If the virtual endomorphism ϕ_f on the mapping class group is contracting, then $\xleftarrow{f}$ has a finite global attractor [29, Thm. 1.4]. Here, contracting means that for some generating set of $\mathrm{PMod}(S^2, P)$, there exists $0 < \rho < 1, n \in \mathbb{N}$, and $C \geq 1$ such that $||\phi_f^n(g)|| < \rho||g||$ whenever $||g|| > C$. Nekrashevych [25, Thm. 7.2] establishes this contraction property in the case of hyperbolic polynomials.
   (d) A topological polynomial is a Thurston map f which is maximally ramified at some fixed-point, which we call infinity. It is convenient to redefine the postcritical set P_f so as to omit infinity. Belk, Lanier, Margalit, and Winarski [3] associate to a topological polynomial f a contractible simplicial complex $\mathcal{T}$ whose vertices are planar trees with endpoints in P_f, up to planar isomorphism, and a simplicial map $\lambda_f : \mathcal{T} \to \mathcal{T}$ induced by lifting. For general complex polynomials, the associated Hubbard trees are fixed vertices, and the uniqness of the Hubbard tree for iterates of f leads to a contraction property of λ_f that implies the existence of a finite global attractor for the pullback relation on curves.

       Question 11.4.2 For a general Thurston map, is there a natural simplicial action on a contractible simplicial complex?

   (e) If the correspondence on moduli space (in the direction of σ_f) has a nonempty invariant compact subset, then ϕ_f is contracting, so there is a finite global attractor. If moduli space admits an incomplete metric which is (i) uniformly contracted by σ_f, and (ii) whose completion is homeomorphic to that of the WP metric, then the trivial curve is a finite global attractor [21, Thm. 7.2]. The latter occurs for $f(z) = z^2+i$; the correspondence on moduli space is the inverse of a Lattès map with three postcritical points and Julia set the whole sphere, which expands the Euclidean orbifold metric.

(f) Intersection theory provides some insight. Elementary arguments show that if α, β are any two curves in $\widehat{\mathbb{C}} - P$ then

$$\mathrm{mod}(\alpha)\mathrm{mod}(\beta) \lesssim 1/\iota(\alpha, \beta)^2; \tag{11.1}$$

here mod$(\cdot)$ refers to the analytic modulus of the family of paths in the given homotopy class. Now consider the linear map $\lambda_f : \mathbb{R}[\mathcal{C}] \to \mathbb{R}[\mathcal{C}]$ defined on basis vectors γ by

$$\gamma \mapsto \sum_{\delta \subset f^{-1}(\gamma)} \frac{1}{\deg(\gamma \stackrel{f}{\leftarrow} \delta)}[\delta]$$

where $[\delta]$ denotes the class of δ in $\mathcal{C}$ if it is nontrivial, and zero otherwise. Obstructions correspond to certain invariant subspaces for which the restriction has an eigenvalue outside the open unit disk. If f is rational, $\gamma \in \mathcal{C}$, A is an embedded annulus homotopic to γ with mod$(A) \geq m$, and $f^{-1}(\gamma) \cap \mathcal{C} = \{\gamma_1, \ldots, \gamma_k\}$, then the annulus A lifts to a collection of annuli with corresponding vector of moduli bounded below by $\lambda_f(m \cdot \gamma)$.

If f is rational, then for any curve γ, we have $||\lambda_f^n(\gamma)||_\infty \to 0$ as $n \to \infty$. For otherwise, one has either an obstruction arising from cycles of $\stackrel{f}{\leftarrow}$ (of possibly a generalized sort, in which the curves might intersect), or a wandering curve with a uniform lower bound on the moduli of path families in the corresponding coordinates. The former possibility is ruled out by the rationality assumption, and the latter by Eq. (11.1).

11.4.2 Bounds on the Size of the Attractor

Since up to conjugacy there are only finitely many non-flexible Lattès rational maps with a given degree d and size $\#P$ of postcritical set, the cardinality of the finite global attractor $\mathcal{A}$, if one exists, must be bounded by some constant depending on d and $\#P$. I know very little about the behavior of this function.

1. Certainly $\#\mathcal{A}$ can be large if $\#P$ is large: for $n \geq 2$ the "$1/n$-rabbit" quadratic polynomial will have an n-cycle of curves. Other examples can be constructed by perturbing flexible Lattès examples f as follows. By pushing down from the complex torus, we can find, for any integer $m \geq 1$, a collection $\gamma_1, \ldots, \gamma_m$ of analytic curves in $\widehat{\mathbb{C}}$ for which $f(\gamma_i) = \gamma_i$, $i = 1 \ldots m$. The union of these curves

is a hyperbolic set, which is therefore stable under sufficiently small perturbations of the map f. A result of X. Buff and T. Gauthier [7, Cor. 3] implies that f is a limit of a sequence $f_1, f_2, \ldots$ of hyperbolic rational Thurston maps each with the maximum number $2d-2$ of attracting cycles. Combining these observations, we conclude some f_k with k sufficiently large is a Thurston map with m fixed curves.

2. In composite degrees, $\#\mathcal{A}$ can be small (say zero), by taking e.g. examples with σ_f constant. Using McMullen's compositional trick one can easily build both hyperbolic rational maps and rational maps with Julia set the whole sphere having the property that σ_f is constant and $\#P_f$ is arbitrarily large.
3. Results of G. Kelsey and R. Lodge [20] show that for quadratic rational maps f with $\#P_f = 4$, we have $\#\mathcal{A} \le 4$.

   The bound might be explained as follows. The map f corresponds (not quite bijectively) to a repelling fixed-point p of a correspondence $g = Y \circ X^{-1}$ on moduli space. In the nonexceptional cases, this is actually a rational map $g : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$. There appears to be a natural bijection between invariant (multi)curves for f and periodic internal rays joining points in periodic superattracting cycles of g (these lie at infinity in moduli space) to p. This seems to be true also for critically fixed polynomials with three finite critical points.

Question 11.4.3 If f is rational and Γ is an invariant multicurve, does there exist a real-analytic σ_f-invariant curve connecting the unique fixed-point of σ_f in Teich$(\widehat{\mathbb{C}}, P)$ to a fixed-point in a WP-boundary stratum corresponding to Γ? More generally, does something similar occur for periodic multicurves?

For quadratics with four postcritical points, the analysis of [20] seems to confirm this. But in higher degrees with $\#P_f = 4$, the situation is more complicated; see e.g. [22].

In higher dimensions, another first natural example to try is the case of f a critically fixed polynomial with four finite simple critical points. One would need to show the existence of internal rays in two complex dimensions. This example is beautifully symmetric and possesses many invariant complex lines that might make the problem more tractable. See [9, section 3].

11.4.3 Examples with Symmetries

Maps with nontrivial symmetries provide a source of non-rational examples without a finite global attractor. We denote by $\mathrm{Mod}(f) = \{h : hf \simeq fh\}$. We recall four facts:

1. The pure mapping class group $P\mathrm{Mod}(S^2, P)$ has no elements of finite order, so neither does $P\mathrm{Mod}(f)$.
2. If f is rational, $P\mathrm{Mod}(f)$ is trivial, unless f is a flexible Lattès example, in which case it is isomorphic to the free group on two generators.

3. Suppose f has an obstruction $\Gamma = (\gamma_1, \ldots, \gamma_m)$ satisfying $f^{-1}(\Gamma) = \Gamma$ as subsets of $\mathcal{C}$, and with the property that the linear map $\lambda_f : \mathbb{R}[\Gamma] \to \mathbb{R}[\Gamma]$ has 1 among its eigenvalues, with a corresponding nonnegative integer eigenvector $(a_1, \ldots, a_m)$. Let T_i denote the Dehn twist about γ_i. Then some power of $T_1^{a_1} \cdots T_m^{a_m}$ gives an element of Mod(f) [28].
4. Thurston maps are like mapping classes. If f is obstructed, there is a canonical decomposition by cutting along a certain invariant multicurve [27, 32]. The "pieces" might contain cycles of degree one: mapping classes, each with its own centralizer. The fact that the decomposition is canonical means that the centralizers of the pieces will embed into Mod(f). Using this idea one can create examples of Thurston maps with a variety of prescribed behaviors. For example, if f is the identity on some sufficiently large piece, then clearly $\mathcal{A}$ contains infinitely many fixed curves. If f is a pseudo-Anosov map on some other sufficiently large piece, then there are infinitely many distinct orbits of wandering curves.
5. L. Bartholdi and D. Dudko give an explicit example of f with Mod(f) infinitely generated [1].

11.4.4 Maps with the Same Fundamental Invariants

As motivation, recall that if L is a flexible Lattès example with postcritical set P, then σ_L acts as the identity, and the pullback relation on curves is the identity function. So if f is now an arbitrary Thurston map with the same postcritical set P, and if $P_{L\circ f} = P$, then $\sigma_{L\circ f} = \sigma_f$ and the pullback relation on curves for f and $L \circ f$ are the same.

Question 11.4.4 When do two Thurston maps have the same pullback relation on curves?

11.4.5 Expanding vs. Nonexpanding Maps

The examples in Sect. 11.4.3 are not isotopic to expanding maps. There exist Levy cycles—cycles of $\stackrel{f}{\leftarrow}$ in $\mathcal{C}$ in which each curve maps by degree 1. Levy cycles are essentially the only obstructions to the existence of an expanding representative, as the main result of [2] shows. The example below shows that there exist expanding obstructed maps without finite global attractors.

Blow up the degree four flexible Lattès example along the lower middle vertical edge to get a Thurston map f; see Fig. 11.2. The "rim" of the square pillowcasethe common boundary of the two squares at left is an invariant Jordan curve containing P_f: that is, f is the subdivision map of a finite subdivision rule. This example has no Levy cycles: for otherwise, the Levy cycle forms an obstruction. For maps with four postcritical points, there is precisely one obstruction. In this example, it is the vertical cure, which is not a Levy cycle. Appealing to the characterization of expanding maps in [2], we conclude there is such an example that is expanding on the whole sphere with respect to a complete length metric. We remark that one can rule out Levy cycles also as follows: this map satisfies the combinatorial expansion properties of both Bonk-Meyer [5] and Cannon-Floyd-Parry [10]. Appealing to either one of these works, we conclude there is a map isotopic to this example in which the diameters of the tiles goes to zero upon iterated subdivision. This implies that this example has no Levy cycles.

The vertical curve is an obstruction with multiplier 1, and the horizontal curve is invariant. Let T be the Dehn twist about this vertical curve, so that $fT^2 \simeq T^2 f$. This immediately implies (1) $\mathcal{A}_f$ is infinite, since the orbit of the horizontal curve under the infinite cyclic group generated by T^2 will consist of f-invariant curves, and (2) if we put $g = Tf$, then the horizontal curve γ wanders. To see this, note that $g^{-n}(\gamma) = (Tf)^{-n}(\gamma) = T^{-n} f^{-n}(\gamma) = T^{-n}\gamma$ as required. I do not know if Tf is isotopic to an expanding map. Letting L denote the flexible Lattès map induced by doubling on the torus, however, we have that LTf is expanding. To see this, we observe that under L every preimage of a curve maps by degree two. Thus LTf has no Levy cycles and has a wandering curve.

Question 11.4.5 Suppose $\text{Mod}(f)$ is trivial. Could there exist infinitely many periodic curves? Could there exist wandering curves?

Acknowledgements This work was partially supported by Simons Foundation collaboration grants 4429419 and 615022.

References

1. L. Bartholdi, D. Dudko, Algorithmic aspects of branched coverings II/V: sphere bisets and their decompositions (2016). https://arxiv.org/pdf/1603.04059.pdf
2. L. Bartholdi, D. Dudko, Algorithmic aspects of branched coverings IV/V. Expanding maps. Trans. Am. Math. Soc. **370**, 7679–7714 (2018)
3. J. Belk, J. Lanier, D. Margalit, R.R.Winarski, Recognizing topological polynomials by lifting trees (2019). https://arxiv.org/pdf/1906.07680.pdf
4. I. Berstein, A.L. Edmonds, On the construction of branched coverings of low-dimensional manifolds. Trans. Am. Math. Soc. **247**, 87–124 (1979)
5. M. Bonk, D. Meyer, *Expanding Thurston Maps*. Mathematical Surveys and Monographs, vol. 225 (American Mathematical Society, Providence, 2017)
6. M. Bonk, M. Hlushchanka, A. Iseli, Eliminating Thurston obstructions and controlling dynamics on curves (2021). arXiv:2105.06938
7. X. Buff, T.Gauthier, Perturbations of flexible Lattès maps. Bull. Soc. Math. France **141**, 603–614 (2013)
8. X. Buff, A. Epstein, S. Koch, K. Pilgrim, On Thurston's pullback map, in *Complex Dynamics* (A K Peters, Wellesley, 2009), pp. 561–583
9. X. Buff, A.L. Epstein, S. Koch, Böttcher coordinates. Indiana Univ. Math. J. **61**, 1765–1799 (2012)

10. J.W. Cannon, W.J. Floyd, W.R. Parry, Finite subdivision rules. Conform. Geom. Dyn. **5**, 153–196 (2001)
11. J.W. Cannon, W.J. Floyd, W.R. Parry, K.M. Pilgrim, Nearly Euclidean Thurston maps. Conform. Geom. Dyn. **16**, 209–255 (2012)
12. G. Cui, Rational maps with (constant) pullback map. http://www.math.ac.cn/kyry/cgz/201501/W020150121410977178324.pdf
13. G. Cui, W. Peng, L. Tan, Renormalizations and wandering Jordan curves of rational maps. Commun. Math. Phys. **344**, 67–115 (2016)
14. A. Douady, J. Hubbard, A proof of Thurston's topological characterization of rational functions. Acta. Math. **171**, 263–297 (1993)
15. A. Fathi, F. Laudenbach, V. Poénaru, *Thurston's Work on Surfaces*. Mathematical Notes,vol. 48 (Princeton University Press, Princeton, 2012). Translated from the 1979 French original by Djun M. Kim and Dan Margalit
16. W. Floyd, G. Kelsey, S. Koch, R. Lodge, W. Parry, K.M. Pilgrim, E. Saenz, Origami, affine maps, and complex dynamics. Arnold Math. J. **3**, 365–395 (2016)
17. W. Floyd, W. Parry, K.M. Pilgrim, Rationality is decidable for nearly Euclidean Thurston maps. Geom. Dedicata **213**, 487–512 (2021). https://doi.org/10.1007/s10711-020-00593-9. https://mathscinet.ams.org/mathscinet/search/publications.html?fmt=bibtex&pg1=MR&s1=4278340
18. W.J. Harvey, Boundary structure of the modular group, in *Riemann Surfaces and Related Topics: Proceedings of the 1978 Stony Brook Conference (State Univ. New York, Stony Brook, 1978)*. Annals of Mathematics Studies, vol. 97 (Princeton University Press, Princeton, 1981), pp. 245–251
19. M. Hlushchanka, Tischler graphs of critically fixed rational maps and their applications (2019). https://arxiv.org/pdf/1904.04759.pdf
20. G. Kelsey, R. Lodge, Quadratic Thurston maps with few postcritical points (2017). https://arxiv.org/abs/1704.03929
21. S. Koch, K.M. Pilgrim, N. Selinger, Pullback invariants of Thurston maps. Trans. Am. Math. Soc. **368**, 4621–4655 (2016)
22. R. Lodge, Boundary values of the Thurston pullback map. Conform. Geom. Dyn. **17**, 77–118 (2013)
23. E.A.S. Maldonado, On nearly Euclidean thurston maps. PhD. Thesis, Virginia Polytechnic University (2012)
24. H. Masur, Extension of the Weil-Petersson metric to the boundary of Teichmuller space. Duke Math. J. **43**, 623–635 (1976)
25. V. Nekrashevych, Combinatorial models of expanding dynamical systems. Ergodic Theory Dyn. Syst. **34**, 938–985 (2014)
26. W. Parry, NET map slope functions (2018). arXiv:1811.01274
27. K.M. Pilgrim, Canonical Thurston obstructions. Adv. Math. **158**, 154–168 (2001)
28. K.M. Pilgrim, *Combinations of Complex Dynamical Systems*. Lecture Notes in Mathematics, vol. 1827 (Springer, Berlin, 2003)
29. K.M. Pilgrim, An algebraic formulation of Thurston's characterization of rational functions. Ann. Fac. Sci. Toulouse Math. **21**, 1033–1068 (2012)
30. K.M. Pilgrim, L. Tan, Combining rational maps and controlling obstructions. Ergodic Theory Dyn. Syst. **18**, 221–245 (1998)
31. R. Ramadas, Algebraic stability of meromorphic maps descended from Thurston's pullback maps. Trans. Am. Math. Soc. **374**, 565–587 (2021)
32. N. Selinger, Thurston's pullback map on the augmented Teichmüller space and applications. Invent. Math. **189**, 111–142 (2012)

Chapter 12
The Pullback Map on Teichmüller Space Induced from a Thurston Map

William Floyd

Abstract William Thurston's topological characterization of rational maps has had an enormous impact in complex dynamics. The proof depends heavily on a pullback map σ_f on Teichmüller space associated with a postcritically finite branched covering $f\colon S^2 \to S^2$. In this chapter we describe Thurston's characterization theorem, and briefly discuss some more recent developments on understanding the pullback map σ_f.

Keywords Thurston map · Pullback map · Teichmüller space

AMS Classification (2010) 30F60, 37F20, 57M12

When is a branched covering $f\colon S^2 \to S^2$ equivalent, in some useful sense, to a rational map? If it is equivalent to a rational map, how many (up to conjugacy) rational maps is it equivalent to? And given the map f, how do you tell? For a general branched covering f, these are extremely difficult questions. A natural reduction is to assume that f is postcritically finite, that is that the forward orbit P_f of the set of critical points is finite.

For a postcritically finite branched covering $f\colon S^2 \to S^2$, William Thurston answered these questions in 1982. To make the statements clearer, we assume that f has a hyperbolic orbifold; this avoids a small number of cases which are well understood. The first step in answering the questions is to involve Teichmüller theory. Given f as above, let $\mathcal{T}_f := \mathcal{T}_{P_f}$ be the Teichmüller space of the orbifold (S^2, P_f) associated with f. There is a pullback map $\sigma_f\colon \mathcal{T}_f \to \mathcal{T}_f$. In Theorem 12.1 Thurston proves that f is combinatorially equivalent to a rational map if and only if σ_f has a fixed point. Moreover, there is a bijection between fixed points of σ_f and conjugacy classes of rational maps equivalent to f. Thurston then develops an obstruction theory and uses it to characterize when f is equivalent to a rational map.

W. Floyd (✉)
Department of Mathematics, Virginia Tech, Blacksburg, VA, USA
e-mail: floyd@math.vt.edu

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_12

12.1 Thurston's Characterization Theorem

Let $f\colon S^2 \to S^2$ be a continuous map. The map f is a *branched covering* if it is orientation preserving and for each $x \in S^2$ there are local charts about x and $f(x)$ (sending x and $f(x)$ to 0) for which f becomes the map $x \mapsto x^k$ for some positive integer k. The integer k is independent of the choice of charts; it is called the *local degree* of f at x and is denoted $\deg_x(f)$. A point $x \in S^2$ is called a *critical point* if $\deg_x(f) > 1$ and is called a *postcritical point* if there exist a critical point $y \in S^2$ and a positive integer n such that $f^{\circ n}(y) = x$. The *critical set* Ω_f is the union of the critical points, and the *postcritical set* P_f is the union of the postcritical points. The map f is *postcritically finite* or *critically finite* if P_f is finite. If f is postcritically finite and the degree of f is at least two (so f isn't a homeomorphism), then f is a *Thurston map*.

For each $x \in S^2$, let $D_f(x) = \{k \in \mathbb{Z}_+ :$ there exist a positive integer n and a point $y \in S^2$ such that $f^{\circ n}(y) = x$ and $\deg_y(f^{\circ n}) = k\}$. That is, $D_f(x)$ is the set of local degrees of iterates of f at preimages of x. Define $\nu_f\colon S^2 \to \mathbb{Z}_+ \cup \{\infty\}$ by

$$\nu_f(x) = \begin{cases} \mathrm{lcm}(D_f(x)) & \text{if } D_f(x) \text{ is finite,} \\ \infty & \text{if } D_f(x) \text{ is infinite.} \end{cases}$$

Let $\mathcal{O}_f$ be the orbifold (S^2, ν_f). The postcritical set P_f is the set of distinguished points (points $x \in S^2$ with $\nu_f(x) > 1$) of $\mathcal{O}_f$. The *Euler characteristic of the orbifold* $\mathcal{O}_f$ is

$$\chi(\mathcal{O}_f) = 2 - \sum_{x \in P_f} \left(1 - \frac{1}{\nu_f(x)}\right).$$

The orbifold Euler characteristic differs from the Euler characteristic of the underlying space in that the contribution of a vertex x is $1/\nu_f(x)$ instead of 1. It is standard, see for example Peter Scott's paper [29, Section 2], that $\mathcal{O}_f$ is hyperbolic if and only if $\chi(\mathcal{O}_f) < 0$, and $\mathcal{O}_f$ is Euclidean if and only if $\chi(\mathcal{O}_f) = 0$.

Two Thurston maps $f, g\colon S^2 \to S^2$ are *combinatorially equivalent* or *Thurston equivalent* if there are orientation-preserving homeomorphisms $h_0, h_1\colon (S^2, P_f) \to (S^2, P_g)$ such that $h_0 \circ f = g \circ h_1$ and h_0 and h_1 are isotopic rel P_f. A Thurston map f is *realizable* by a rational map if it is combinatorially equivalent to a rational map.

A fundamental contribution of Thurston's to complex dynamics is his topological characterization theorem for rational maps. It solves the problem of determining when a Thurston map is combinatorially equivalent to a rational map. A special case of the problem of when a Thurston map is realizable by a rational map came up in an earlier paper [21, Theorem 12.1] of John Milnor and Thurston on iterating maps of an interval; the proof of the existence of the required polynomials there is by a different method.

Thurston's approach to the problem is through the Teichmüller space of the associated orbifold. Let f be a Thurston map, and let $P = P_f$. The *Teichmüller space* $\mathcal{T}_P := \mathcal{T}(S^2, P) = \mathcal{T}(\mathcal{O}_f)$ of the orbifold $\mathcal{O}_f$ is the space of complex structures on $\mathcal{O}_f \setminus P$, up to isotopy. Since there is a single complex structure on S^2, we can view a point in $\mathcal{T}_P$ as a finite set $B \subset \widehat{\mathbb{C}}$ together with an orientation-preserving homeomorphism $\phi\colon (S^2, P) \to (\widehat{\mathbb{C}}, B)$. In this point of view, two orientation-preserving homeomorphisms $\phi_1\colon (S^2, P) \to (\widehat{\mathbb{C}}, B_1)$ and $\phi_2\colon (S^2, P) \to (\widehat{\mathbb{C}}, B_2)$ are equivalent if there is a Möbius transformation $h\colon (\widehat{\mathbb{C}}, B_1) \to (\widehat{\mathbb{C}}, B_2)$ such that $h \circ \phi_1$ is isotopic to ϕ_2 rel P. The Teichmüller space $\mathcal{T}_P$ is a complex manifold (for example, see Hubbard's book [11, Theorem 6.5.1]). The *moduli space* $\mathcal{M}_P := \mathcal{M}(\mathcal{O}_f)$ is the space of injections of P into $\widehat{\mathbb{C}}$, modulo the equivalence of postcomposition with a Möbius map. The Teichmüller space $\mathcal{T}_P$ is the universal covering space of the moduli space $\mathcal{M}_P$; the covering map π is defined by $\pi([\phi]) = [\phi|_P]$ (for example, see the Douady-Hubbard paper [8, Section 3]).

A complex structure on $\mathcal{O}_f \setminus P$ lifts under f to a complex structure on $\mathcal{O}_f \setminus f^{-1}(P)$, and this extends to a complex structure on $\mathcal{O}_f \setminus P$. This map on complex structures descends to the quotient spaces to give the *pullback map* $\sigma_f\colon \mathcal{T}_P \to \mathcal{T}_P$. It is straightforward to show that σ_f is analytic (see [8, Proposition 2.1]), and that σ_f only depends on the combinatorial equivalence class of f. Using the coderivative of σ_f, one can show that $\|D\sigma_f(z)\| \leq 1$ and if $\mathcal{O}_f$ is hyperbolic then $\|D\sigma_f^{\circ 2}(z)\| < 1$ for all $z \in \mathcal{T}_P$ (see [8, Proposition 3.3] or [12, Proposition 10.7.3]) and so $\sigma_f^{\circ 2}$ is a weak contraction.

If f is a rational Thurston map, then f fixes the complex structure on $\widehat{\mathbb{C}}$ and the pullback map σ_f has a fixed point. The combinatorial equivalence relation on Thurston maps fits well enough together with the equivalence relation in the definition of the Teichmüller space that one has the following theorem.

Theorem 12.1 (Thurston) *Let f be a Thurston map. Then f is combinatorially equivalent to a rational map if and only if σ_f has a fixed point.*

Proof First suppose that f is combinatorially equivalent to a rational map g. Then there are orientation-preserving homeomorphisms $\phi_1, \phi_2\colon (S^2, P_f) \to (\widehat{\mathbb{C}}, P_g)$ such that the diagram

$$\begin{array}{ccc} (S^2, P_f) & \xrightarrow{\phi_2} & (\widehat{\mathbb{C}}, P_g) \\ f \downarrow & & \downarrow g \\ (S^2, P_f) & \xrightarrow{\phi_1} & (\widehat{\mathbb{C}}, P_g) \end{array}$$

commutes, and ϕ_1 and ϕ_2 are isotopic rel P_f. Then $\sigma_f([\phi_1]) = [\phi_2]$. By the definition of combinatorial equivalence, $[\phi_2] = [\phi_1]$ and $[\phi_1]$ is a fixed point of σ_f.

Now suppose that $[\phi]$ is a fixed point of σ_f. Then there are finite sets $A_1, A_2 \subset \widehat{\mathbb{C}}$ and an orientation-preserving homeomorphism $\tilde{\phi}\colon (S^2, P_f) \to (\widehat{\mathbb{C}}, A_2)$ such that the diagram

$$\begin{array}{ccc} (S^2, P_f) & \xrightarrow{\tilde{\phi}} & (\widehat{\mathbb{C}}, A_2) \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g} \\ (S^2, P_f) & \xrightarrow{\phi} & (\widehat{\mathbb{C}}, A_1) \end{array}$$

commutes, where $g = \phi \circ f \circ \tilde{\phi}^{-1}$ is analytic. Since $[\tilde{\phi}] = \sigma_f([\phi]) = [\phi]$, there is a Möbius transformation $h\colon (\widehat{\mathbb{C}}, A_1) \to (\widehat{\mathbb{C}}, A_2)$ such that the diagram

$$\begin{array}{ccc} & & (\widehat{\mathbb{C}}, A_1) \\ & \overset{\phi}{\nearrow} & \downarrow{\scriptstyle h} \\ (S^2, P_f) & \xrightarrow[\tilde{\phi}]{} & (\widehat{\mathbb{C}}, A_2) \end{array}$$

commutes up to isotopy rel P_f. Then the diagram

$$\begin{array}{ccc} (S^2, P_f) & \xrightarrow{h^{-1}\circ\tilde{\phi}} & (\widehat{\mathbb{C}}, A_1) \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle g\circ h} \\ (S^2, P_f) & \xrightarrow{\phi} & (\widehat{\mathbb{C}}, A_1) \end{array}$$

shows that $g \circ h$ is a rational map combinatorially equivalent to f. □

It follows from the proof that the fixed points of σ_f correspond to conjugacy classes of rational maps that are combinatorialy equivalent to f. If f is a Thurston map with a Euclidean orbifold and $P = P_f$, then, unless $\mathcal{O}_f$ is a (2, 2, 2, 2) orbifold (a rectangular pillowcase), $\mathcal{T}_P$ is a single point and f is combinatorially equivalent to a rational map.

If a Thurston map f has a hyperbolic orbifold, then $\sigma_f^{\circ 2}$ is a weak contraction and so σ_f cannot have more than one fixed point. We get the following as a corollary.

Theorem 12.2 (Thurston Rigidity) *Suppose $f\colon S^2 \to S^2$ is a Thurston map such that the orbifold $\mathcal{O}_f$ is hyperbolic. If g and h are rational maps that are combinatorially equivalent to f, then g and h are conjugate by a Möbius transformation.*

We continue to suppose that f is a Thurston map with a hyperbolic orbifold and let $P = P_f$. Choose a point $\tau_0 \in \mathcal{T}_P$, and define the sequence $\{\tau_n\}$ in $\mathcal{T}_P$ recursively by $\tau_i = \sigma_f(\tau_{i-1})$ for all positive integers i. Since σ_f^2 is a weak contraction, $\{\tau_n\}$ will converge if and only if it stays in a compact subset of $\mathcal{T}_P$. By Mumford's theorem [23] (see also Hubbard's book [11, Theorem 7.3.3] for the statement in genus 0) on compact subsets of the moduli spaces, this will happen as long as there is a lower bound on the lengths of simple closed geodesics in hyperbolic surfaces represented by the points τ_n.

One way that the sequence $\{\tau_n\}$ can fail to converge is if there is an annular obstruction (Thurston obstruction). Here is the terminology. A *multicurve* in $\mathcal{O}_f$ is a finite set $\Gamma = \{\gamma_1, \ldots, \gamma_k\}$ of pairwise disjoint, unoriented, simple closed curves in $\mathcal{O}_f \setminus P_f$ such that i) each γ_i is essential (does not bound a disk containing at most one element of P_f) and ii) if $i, j \in \{1, \ldots, k\}$ and $i \neq j$, then γ_i is not isotopic to γ_j in $\mathcal{O}_f \setminus P_f$. A multicurve Γ is *invariant* if for each $\gamma \in \Gamma$, each element of $f^{-1}(\gamma)$ is either non-essential or isotopic in $\mathcal{O}_f \setminus P_f$ to an element of Γ. If $\Gamma = \{\gamma_1, \ldots, \gamma_k\}$ is an invariant multicurve, let A_Γ be the $k \times k$ matrix with i, j-entry

$$A_{ij} = \sum_\alpha \frac{1}{\deg(f\colon \alpha \to \gamma_j)},$$

where the sum is over curves $\alpha \subset f^{-1}(\gamma_j)$ that are isotopic in $\mathcal{O}_f \setminus P_f$ to γ_i. If Γ is an invariant multicurve, the spectral radius of A_Γ (the eigenvalue of largest norm) is called the *multiplier*. We denote it by $\lambda(\Gamma)$. If $\mathcal{O}_f$ is hyperbolic and Γ is an invariant multicurve whose multiplier is at least one, then Γ is called an *annular obstruction* or a *Thurston obstruction*.

Here is the motivation. Suppose $f\colon S^2 \to S^2$ is a Thurston map such that $\mathcal{O}_f$ is hyperbolic. Let $P = P_f$, and suppose that $\Gamma = \{\gamma_1, \ldots, \gamma_k\}$ is an annular obstruction for f. Let $\tau_0 \in \mathcal{T}_P$, and define the sequence $\{\tau_n\}$ recursively as above. Choose a family of pairwise disjoint annuli in $\mathcal{O}_f \setminus P_f$ with core curves $\gamma_1, \ldots, \gamma_k$. For each $i \in \{1, \ldots, k\}$, let v_i be the conformal modulus of the annulus above containing γ_i in a hyperbolic surface determined by τ_0. Let $V = (v_1, \ldots v_k)^t$. It follows by induction and the subadditivity of conformal moduli that for each positive integer n and each $i \in \{1, \ldots, k\}$, in a hyperbolic surface determined by τ_n there is an annulus with core curve γ_i whose conformal modulus is at least the i^{th} component of $A^n V$. In a hyperbolic surface, the conformal modulus of an annulus is bounded above by $2\pi^2$ times the reciprocal of the length of a simple closed geodesic homotopic to a core curve. If the multiplier of Γ is greater than one, then the norms $\|A^n V\|$ are not bounded above, so $\{\tau_n\}$ does not stay in a compact subset of $\mathcal{T}_P$ and hence does not converge. If the multiplier is one, it follows from a theorem of Kurt Strebel [34, Theorem 21.7] that f cannot be a rational Thurston map unless $\mathcal{O}_f$ is a (2, 2, 2, 2) orbifold. If $\mathcal{O}_f$ is a (2, 2, 2, 2) orbifold, then f is double covered by a torus endomorphism $F\colon T_f \to T_f$. In this case, let $A_f\colon H_1(T_f, \mathbb{Z}) \to H_1(T_f, \mathbb{Z})$ be the induced map on the first homology group.

Let f be a Thurston map with a hyperbolic orbifold. If f is combinatorially equivalent to a rational map, then it doesn't have an annular obstruction. Thurston's characterization theorem states that the converse is true. For completeness, we include the statement for the case that f has a Euclidean orbifold.

Theorem 12.3 (Thurston's Characterization Theorem) *Let $f\colon S^2 \to S^2$ be a Thurston map. If $\mathcal{O}_f$ is hyperbolic, then f is combinatorially equivalent to a rational map if and only if $\lambda(\Gamma) < 1$ for every invariant multicurve Γ. If $\mathcal{O}_f$ is Euclidean, then f is combinatorially equivalent to a rational map unless $\mathcal{O}_f$ is a $(2, 2, 2, 2)$ orbifold and a matrix representing A_f has distinct real eigenvalues.*

Let f be a Thurston map whose orbifold $\mathcal{O}_f$ is hyperbolic, let $P = P_f$, and consider a sequence $\{\tau_i\}$ in $\mathcal{T}_P$ of iterates under the pullback map σ_f. If $\{\tau_i\}$ does not converge to a fixed point of σ_f, then for every $\epsilon > 0$ there must exist an index i such that some simple closed geodesic in a hyperbolic structure on $\mathcal{O}_f$ representing τ_i has length less than ϵ. However, this does not imply that the lengths of the geodesics in that free homotopy class will converge to zero under further iteration. For a smaller choice of ϵ, the length of the geodesic in a different free homotopy class could be short. The essence of the proof of the hard direction of Thurston's theorem is that if $\{\tau_i\}$ does not converge, then for some $\epsilon > 0$ and some index i, the ϵ-short geodesics give an invariant multicurve which is an annular obstruction. Here ϵ depends on the largest possible multiplier less than one for an invariant multicurve for f. (Since there are only finitely many possible matrices A_Γ for an invariant multicurve for f, there is a largest possible multiplier less than one.)

While this is a fundamental and very useful theorem, it can be difficult to apply. There are infinitely many multicurves, and it's difficult to predict or know which ones are candidates for annular obstructions.

We'll next briefly describe the application of Thurston's characterization theorem to matings, after first defining Levy cycles since they are needed for that application. For an excellent discussion of several other applications of Thurston's characterization theorem, see the article [3, Section 3] by Xavier Buff, Cui Guizhen and Tan Lei.

12.1.1 Levy Cycles

In his thesis [18], Silvio Levy defines an important class of multicurves which are now called Levy cycles. A *Levy cycle* is a multicurve $\Gamma = \{\gamma_1, \dots, \gamma_k\}$ such that for each $i \in \{1, \dots, k\}$, there is a component of $f^{-1}(\gamma_i)$ that is isotopic to γ_{i-1} (where we define $\gamma_0 := \gamma_k$) and maps to γ_i with degree 1. Levy proves that if a Thurston map with degree 2 is obstructed, then it has a Levy cycle. In [2, Theorem 5.5] Ben Bielefeld, Yuval Fisher, and John Hubbard prove that if a topological polynomial is obstructed, then any annular obstruction contains a Levy cycle. In each case this is a big reduction, since ruling out the existence of a Levy cycle is much easier than ruling out the existence of an annular obstruction.

12.1.2 An Application to Matings

A primary motivation for Thurston's characterization theorem was the discovery of matings by Douady and Hubbard in 1982 [37, Preface] and their mating conjecture on when two quadratic polynomials are mateable. We start with some terminology. Let $d \in \mathbb{Z}$ with $d \geq 2$, and let $p_1\colon \mathbb{C}_1 \to \mathbb{C}_1$ and $p_2\colon \mathbb{C}_2 \to \mathbb{C}_2$ be monic complex polynomials of degree d, where $\mathbb{C}_1 = \mathbb{C} = \mathbb{C}_2$. Let K_1 be the filled Julia set of p_1 and let K_2 be the filled Julia set of p_2. For each $i \in \{1, 2\}$, the Böttcher map $\phi_i\colon \mathbb{C} \setminus K_i \to \mathbb{C} \setminus \overline{\mathbb{D}}$ is a biholomorphism that conjugates the restriction of p_i on the complement of its filled Julia set to the map $z \mapsto z^d$ on the complement of the closed unit disk. The preimages of rays $\{re^{i\theta} : r \in (1, \infty)\}$ under the Böttcher map are called *external rays*, and their closures are called *closed rays*. Let $\widetilde{\mathbb{C}_1}$ and $\widetilde{\mathbb{C}_2}$ be the standard compactification $\widetilde{\mathbb{C}}$ of $\mathbb{C}$ obtained by adding a circle $\{\infty \cdot e^{2\pi i\theta} : \theta \in \mathbb{R}/\mathbb{Z}\}$ which we call the *equator* at infinity. For $i \in \{1, 2\}$, p_i extends continuously to $\widetilde{\mathbb{C}_i}$ by defining $p_i(\infty \cdot e^{2\pi i\theta}) = \infty \cdot e^{2d\pi i\theta}$. Let $S = \widetilde{\mathbb{C}_1} \coprod \widetilde{\mathbb{C}_2}/\sim$, where $\sim$ is the equivalence relation generated by identifying $\infty \cdot e^{2\pi i\theta} \in \widetilde{\mathbb{C}_1}$ with $\infty \cdot e^{-2\pi i\theta} \in \widetilde{\mathbb{C}_2}$ for all $\theta \in \mathbb{R}/\mathbb{Z}$. Then $p_1 \coprod p_2$ induces a continuous map $p\colon S \to S$ with $p([z]) = [p_i(z)]$ if $z \in \widetilde{\mathbb{C}_i}$. The map $p\colon S \to S$ is the *formal mating* of p_1 and p_2.

Now let $\sim$ be the equivalence relation on S generated by calling two points equivalent if they lie on the image in S of a closed ray in $\widetilde{\mathbb{C}_1}$ or $\widetilde{\mathbb{C}_2}$. Since any point in the equator is on two closed rays and a point in the Julia set of p_1 or the Julia set of p_2 may be on multiple closed rays, this equivalence relation can be very complicated. By a theorem of R. L. Moore [22], the quotient space will be a 2-sphere if there is more than one equivalence class, each equivalence class is closed and connected, and no equivalence class separates S into at least two components. If the quotient space $S/\sim$ is a 2-sphere, then the map $p_1 \coprod p_2$ induces a postcritically finite map on the quotient space. This map is a *topological mating* of p_1 and p_2, and p_1 and p_2 are *topologically mateable*.

There is a third construction called the *degenerate mating* or *essential mating*. Let Y' be the set of equivalence classes of $\sim$ that contain at least two points of $\Omega_f \cup P_f$, and let Y be the set of equivalence classes of $\sim$ that contain at least one point of $\Omega_f \cup P_f$ and have some iterated image under f which is an iterated image under f of an element of Y'. If $Y' = \emptyset$, then the degenerate mating is the formal mating. If $Y' \neq \emptyset$ and some element of Y has disconnected complement in the 2-sphere, then the degenerate mating is not defined (and the topological mating is also not defined since the quotient space isn't a 2-sphere). If $Y' \neq \emptyset$ and no element of Y has disconnected complement, then the quotient space S' of S obtained by collapsing each element of Y to a point is a 2-sphere; the map $f'\colon S' \to S'$ induced by f is not a branched covering because it collapses some nontrivial equivalence classes to points. It can be easily modified to a branched covering; this branched covering is the degenerate mating. For more details, see Mitsuhiro Shishikura's paper [32] or Tan Lei's paper [35] or [36].

If the topological mating exists and is topologically conjugate to a rational map g such that the conjugating map is holomorphic on the images in $S/\sim$ of the interiors of the filled-in Julia sets of p_1 and p_2, then g is a *geometric mating* of p_1 and p_2, and p_1 and p_2 are *geometrically mateable* or *mateable*.

Douady and Hubbard conjectured (see [7, Section III.3]) that two quadratic polynomials $f_c(z) = z^2 + c$ and $f_{c'}(z) = z^2 + c'$ are mateable if and only if c and c' are not in conjugate limbs of the Mandelbrot set. It is straightforward that f_c and $f_{c'}$ are not mateable if c and c' are in conjugate limbs. Using his theorem on Levy cycles, Levy made initial progress on the conjecture in [18, Chapter 5]. In her thesis [35] (see also [36]), Tan proves the hard direction of the following theorem. The proof depends crucially on Thurston's characterization theorem, and on the theorem that if a quadratic Thurston map has an annular obstruction then it has a Levy cycle. Shishikura gives a proof in [32] that if the degenerate mating is equivalent to a rational map then the topological mating is topologically conjugate to this rational map; his proof is based on an unpublished manuscript of Mary Rees.

Theorem 12.4 (Tan, Rees-Shishikura) *Suppose $f_c(z) = z^2 + c$ and $f_{c'}(z) = z^2 + c'$ are postcritically finite. The following are equivalent.*

1. *f_c and $f_{c'}$ are geometrically mateable.*
2. *f_c and $f_{c'}$ are topologically mateable.*
3. *c and c' are not in conjugate limbs of the Mandelbrot set.*

In degrees three and higher, an obstructed Thurston map need not have a Levy cycle. In [33, Section 2], Shishikura and Tan give an example of a cubic Thurston map which has an annular obstruction but does not have a Levy cycle. They also give an example of two monic cubic polynomials such that their topological mating exists but their geometric mating does not exist.

12.2 Further Developments

In the years since Thurston proved his characterization theorem, there has been a great deal of interest in getting a deeper understanding of the pullback map. In this section we briefly describe some of the progress that has been made.

12.2.1 Canonical Obstructions

A fixed Thurston map can have more than one annular obstruction. In [8, Appendix], Douady and Hubbard give an example with four different annular obstructions. Furthermore, the homotopy classes of curves in the four obstructions cannot be realized by pairwise disjoint simple closed curves.

In [26, Theorem 1.1], Kevin Pilgrim refines the arguments of [8] and shows that if a Thurston map with hyperbolic orbifold is obstructed then there is a canonical annular obstruction. Let f be a Thurston map with a hyperbolic orbifold and let

$P = P_f$. Given $\tau \in \mathcal{T}_P$ and a homotopy class α of an essential simple closed curve in $\mathcal{O}_f$, let $l_\tau(\alpha)$ be the length of a geodesic representing α in a hyperbolic metric on $\mathcal{O}_f$ representing the point τ. Choose a point $\tau_0 \in \mathcal{T}_P$, and as before define the sequence $\{\tau_i\}$ recursively by $\tau_i = \sigma_f(\tau_{i-1})$. Let Γ_c be the set of homotopy classes α of essential simple closed curves on $\mathcal{O}_f$ such that $\lim_{i\to\infty} l_{\tau_i}(\alpha) = 0$. Since the pullback map is distance nonincreasing, Γ_c does not depend on the initial point τ_0.

Theorem 12.5 (Pilgrim) *Let f be a Thurston map with hyperbolic orbifold, and let Γ_c be defined as above.*

i) If $\Gamma_c = \emptyset$, then f is combinatorially equivalent to a rational map.
ii) If $\Gamma_c \neq \emptyset$, then f is obstructed and there is an annular obstruction for f consisting of a representative of each element of Γ_c.

While knowing the existence of a canonical annular obstruction is useful, it is still difficult to (a) determine whether or not a Thurston map is obstructed and (b) find an annular obstruction (or find the canonical annular obstruction) for an obstructed Thurston map.

12.2.2 The Extension to a Boundary

If you already know Thurston's work on surface homeomorphisms, it's natural to wonder if the pullback map on Teichmüller space has a continuous extension to one of the boundaries of Teichmüller space. In his thesis [30, Theorem 4.1.1] and in the paper [31, Theorem 6.4], Nikita Selinger proves the following:

Theorem 12.6 (Selinger) *Let $f\colon S^2 \to S^2$ be a Thurston map with postcritical set P. Then σ_f extends continuously to the augmented Teichmüller space $\overline{\mathcal{T}}_P$ of $\mathcal{O}_f$.*

The augmented Teichmüller space $\overline{\mathcal{T}}_P$ is homeomorphic to the completion of $\mathcal{T}_P$ with respect to the Weil-Peterson metric. $\overline{\mathcal{T}}_P$ is a stratified space, with each stratum being homeomorphic to the Teichmüller space of a noded surface obtained by collapsing each element of a multicurve to a point.

The augmented Teichmüller space $\overline{\mathcal{T}}_P$ is in general neither compact nor locally compact, and a continuous map of $\overline{\mathcal{T}}_P$ to itself need not have a fixed point. However, the quotient of $\overline{\mathcal{T}}_P$ by the action of the mapping class group is homeomorphic to the Mumford-Deligne compactification of the moduli space.

Even without compactness or the fixed-point property, this is a powerful approach. Using the analysis in the proof, Selinger derives new proofs of Thurston's characterization theorem and of Pilgrim's canonical obstruction theorem. He also notes that in general the pullback map σ_f will not extend continuously to the Thurston boundary of $\mathcal{T}_P$.

12.2.3 The g-Map and the Hurwitz Space

In their seminal paper [1, Section 5.1] on the twisted rabbit problem, Laurent Bartholdi and Volodymyr Nekrashevych define a "g-map" for the rabbit polynomial. Let f be the rabbit polynomial, defined by $f(z) = z^2 + c$ where $f^{\circ 2}(c) = 0$ and $\Im(c) > 0$. Let $P = P_f$, let $\tau \in \mathcal{T}_P$, let $\tau' = \sigma_f(\tau)$, let $w_1 = \pi(\tau)$, and let $w_0 = \pi(\tau')$, where $\pi\colon\ \mathcal{T}_P \to \mathcal{M}_P$ is the universal covering map. They show that

$$w_1 = 1 - \frac{1}{w_0^2}.$$

Note that this is the projection to the moduli space of the correspondence $\sigma_f(\tau) \mapsto \tau$ and not of the function σ_f. For the polynomial $h(z) = z^2 + i$, they show that the correspondence $\sigma_h(\tau) \mapsto \tau$ projects to the moduli space to the function $F(w) = \left(\frac{2-w}{w}\right)^2$. They also compute the function F for a third quadratic polynomial $f(z) = z^2 + c$ corresponding to the critical value c having preperiod 2 and period 1.

In her two theses [14] and [15], Sarah Koch starts with these examples and builds a beautiful theory. Suppose $f\colon\ S^2 \to S^2$ is a Thurston map. We have a diagram

$$\begin{array}{ccc} \mathcal{T}_P & \xrightarrow{\sigma_f} & \mathcal{T}_P \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ \mathcal{M}_P & & \mathcal{M}_P \end{array} .$$

When is there a continuous map $g_f\colon\ \mathcal{M}_P \to \mathcal{M}_P$ such that the diagram

$$\begin{array}{ccc} \mathcal{T}_P & \xrightarrow{\sigma_f} & \mathcal{T}_P \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ \mathcal{M}_P & \xleftarrow[g_f]{} & \mathcal{M}_P \end{array}$$

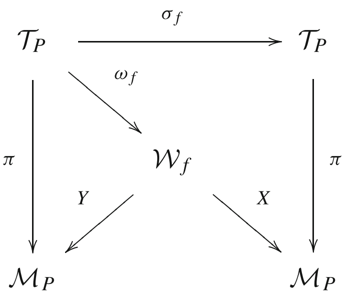
commutes? We get a related question by replacing $\mathcal{M}_P$ by its compactification $\mathbb{P}^n$, where $n = |P_f| - 3$. When is there a continuous map $g_f\colon\ \mathbb{P}^n \to \mathbb{P}^n$ such that the diagram

$$\begin{array}{ccc} \mathcal{T}_P & \xrightarrow{\sigma_f} & \mathcal{T}_P \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ \mathbb{P}^n & \xleftarrow[g_f]{} & \mathbb{P}^n \end{array}$$

commutes, where now we abuse notation and denote by π the postcomposition of the universal covering map with the inclusion of $\mathcal{M}_P$ into its compactification $\mathbb{P}^n$?

In [14, Theorem 7.7] she proves that if f is a unicritical topological polynomial, then there is a postcritically-finite endomorphism $g_f\colon \mathbb{P}^n \to \mathbb{P}^n$ that makes the diagram commute. Furthermore, for each such map the complement of the postcritical set is Kobayashi hyperbolic. The periodic cycles of g_f in $\mathcal{M}_P$ are all repelling, and each fixed point corresponds to a rational map with the same dynamic portrait as f. In [15, Theoren 4.0.1] she proves the same results for a topological polynomial f such that each critical point of f is periodic. Furthermore, if g_f exists then $\sigma_f(\mathcal{T}_P)$ is open and the restriction $\sigma_f\colon \mathcal{T}_P \to \sigma_f(\mathcal{T}_P)$ is a covering map.

Koch revisits the issue in [16]. Suppose $f\colon S^2 \to S^2$ is a Thurston map such that $|P_f| \geq 4$, and let $P = P_f$. A homeomorphism $h\colon (S^2, P) \to (S_2, P)$ with $h|_P = \mathrm{id}$ (that is, a representative of an element of the pure mapping class group) is called *liftable* if there is a homeomorphism $h'\colon (S^2, P) \to (S^2, P)$ such that $h'|_P = \mathrm{id}$ and $f \circ h' = h \circ f$. Let H_f be the subgroup of the pure mapping class group of (S^2, P) represented by liftable elements. Koch defines the *Hurwitz space* $\mathcal{W}_f$, which has the following properties: $\mathcal{W}_f$ is a complex manifold that is homeomorphic to $\mathcal{T}_P/H_f$; the universal covering map $\mathcal{T}_P \xrightarrow{\pi} \mathcal{M}_P$ factors as $\mathcal{T}_P \xrightarrow{\omega_f} \mathcal{W}_f \xrightarrow{Y} \mathcal{M}_P$, where ω_f is a holomorphic covering map and Y is a finite holomorphic covering map; and there is a holomorphic map $X\colon \mathcal{W}_f \to \mathcal{M}_P$ such that $\pi \circ \sigma_f = X \circ \omega_f$. Hence the following diagram commutes.



The Hurwitz space $\mathcal{W}_f$ depends only on the Hurwitz class of f. If f is a topological polynomial that is either unicritical or has all critical points periodic, then X is injective. If f has only two critical points and one of them is a postcritical point, then f is Hurwitz equivalent to a unicritical topological polynomial and so X is injective.

12.2.4 The Pullback Map Near a Fixed Point

In [4, Theorem 1.1], Buff, Adam Epstein, Koch, and Pilgrim consider the possibilities for the pullback map near a fixed point. They show that the following possibilities can each occur for a fixed point τ for the pullback map σ_f of a Thurston map $f\colon S^2 \to S^2$ with $|P_f| \geq 4$. As before, let $P = P_f$.

(1) τ is an attracting fixed point of σ_f, the image $\sigma_f(\mathcal{T}_P)$ is open and dense in $\mathcal{T}_P$, and σ_f is a covering map.
(2) τ is a superattracting fixed point of σ_f, σ_f is surjective, and σ_f is a ramified Galois covering map.
(3) σ_f is the constant map to the point τ.

They prove that case (1) occurs whenever f is a polynomial such that $|P_f| \geq 4$ and all of the critical points of f are periodic.

For case (2), they give the specific example $f(z) = \frac{3z^2}{2z^3+1}$. Here $P_f = \{0, 1, -1/2 \pm i\sqrt{3}/2\}$. All four postcritical points are critical; 0 and 1 are fixed, and the other two are mapped to each other. In his thesis [19, Section 4] and in the paper [20, Section 5], Russell Lodge studies this example in great depth. He computes the functions ω_f, X, and Y from the $\mathcal{W}$-space diagram, and computes the virtual endomorphism on the fundamental group and the virtual endomorphism on the pure mapping class group. He analyzes the slope function (the pullback map on the boundary points of $\overline{\mathcal{T}_P}$) and shows that the slope function has a finite global attractor consisting of a fixed point and a 2-cycle. Using this, he solves the twist problem for f.

The example that Buff, Epstein, Koch, and Pilgrim give for case (3), $f(z) = 2i\left(z^2 - \frac{1+i}{2}\right)^2$, is due to Curtis McMullen, as is the explanation for why it has a constant pullback map. The map f factors as $g \circ s$, where $g(z) = 2i\left(z - \frac{1+i}{2}\right)^2$ and $s(z) = z^2$. Furthermore, σ_f factors as $\sigma_f = \sigma_s \circ \sigma_g$, where $A = \{0, 1, \infty\}$, $\sigma_g\colon \mathcal{T}(\widehat{\mathbb{C}}, P_f) \to \mathcal{T}(\widehat{\mathbb{C}}, A)$, and $\sigma_s\colon \mathcal{T}(\widehat{\mathbb{C}}, A) \to \mathcal{T}(\widehat{\mathbb{C}}, P_f)$. Since $\mathcal{T}(\widehat{\mathbb{C}}, A)$ is a single point, σ_f is a constant map. In addition to this example, they give hypotheses, now often called *McMullen's condition*, on a factorization of a Thurston map f that implies that σ_f is a constant map. It was believed that these conditions might be the only way for a Thurston map f to have $|P_f| \geq 4$ and σ_f constant, but this is now known to be false. They also prove that σ_f is a constant map if and only if for every essential curve α in $S_2 \setminus P_f$, no component of $f^{-1}(\alpha)$ is essential.

In their introductory paper [6, Theorem 10.2] on nearly Euclidean Thurston (NET) maps, James Cannon, Floyd, Walter Parry, and Pilgrim give an algebraic formulation that characterizes when a NET map has a constant pullback map. A NET map is a Thurston map $f\colon S^2 \to S^2$ such that $|P_f| = 4$ and every critical point has local degree 2. They are close enough to Euclidean Thurston maps to be computationally tractable, but general enough to have many interesting examples.

In his thesis [27] (much of which appears in [28]), Edgar Saenz does extensive work on NET maps with constant pullback maps. He gives a NET map of degree 9 that has constant pullback map and does not satisfy McMullen's condition, and shows that for every odd multiple of 9 there is a NET map of that degree that has constant pullback map and does not satisfy McMullen's condition.

In [27, Appendix D] Saenz also gives a remarkable example in degree 4, the map $f(z) = \frac{z(z^3+2)}{2z^3+1}$. The map f isn't a NET map since each of the three critical points has local degree 3. An easy core arc argument shows that f has a constant pullback map. Since the critical points have degree 3, f cannot factor so as to satisfy McMullen's condition.

12.2.5 Other Pullback Invariants

In addition to the pullback map σ_f on Teichmüller space, there are several other pullback invariants that can be associated with a Thurston map $f\colon (S^2, P_f) \to (S^2, P_f)$. In [17], Koch, Pilgrim, and Selinger discuss these invariants and establish connections between them. Before describing their results, we start with some terminology.

Let P be a finite subset of S^2 with $|P| \geq 3$. As before $\mathcal{T}_P$ denotes the Teichmüller space $\mathcal{T}(S_2, P)$ and $\mathcal{M}_P$ denotes the moduli space $\mathcal{M}(S^2, P)$. Let $\mathcal{S}_P$ denote the set of homotopy classes of simple, unoriented, essential curves in $S_2 \setminus P$, and let $\mathbb{R}[\mathcal{S}_P]$ denote the free $\mathbb{R}$-module with basis $\mathcal{S}_P$. Let $G_P := \mathrm{PMod}(S^2, P)$, the pure mapping class group of (S^2, P).

Now suppose that $f\colon (S^2, P_f) \to (S^2, P_f)$ is a Thurston map with $|P_f| \geq 3$, and let $P = P_f$. Two pullback invariants of f are the pullback map $\sigma_f\colon \mathcal{T}_P \to \mathcal{T}_P$ and its continuous extension $\sigma_f\colon \overline{\mathcal{T}}_P \to \overline{\mathcal{T}}_P$. We can also define the *pullback relation*

$$\mathcal{S}_P \cup \{o\} \stackrel{f}{\leftarrow} \mathcal{S}_P \cup \{o\}$$

by $o \stackrel{f}{\leftarrow} o$, $[\alpha] \stackrel{f}{\leftarrow} o$ if a connected component of $f^{-1}(\alpha)$ is not essential, and $[\alpha] \stackrel{f}{\leftarrow} [\beta]$ if a connected component of $f^{-1}(\alpha)$ is in $[\beta]$. Much as we define the matrix A_Γ associated with an invariant multicurve Γ, there is a linear operator $\lambda_f\colon \mathbb{R}[\mathcal{S}_P] \to \mathbb{R}[\mathcal{S}_P]$ defined by

$$\lambda_f([\alpha]) = \sum_{\beta_i} d_i[\beta_i],$$

where the sum is over essential connected components of $f^{-1}(\alpha)$ and $d_i = 1/(\deg(f\colon \beta_i \to \alpha))$.

Let H_f be the subset of liftable elements of the pure mapping class group G_P. So a homeomorphism $h\colon (S^2, P) \to (S^2, P)$ which fixes each element of P represents an element of H_f if and only if there is a homeomorphism $\tilde{h}\colon (S^2, P) \to (S^2, P)$ fixing each element of P such that the diagram

$$\begin{array}{ccc} (S^2, P) & \xrightarrow{\tilde{h}} & (S^2, P) \\ {\scriptstyle f}\downarrow & & \downarrow{\scriptstyle f} \\ (S^2, P) & \xrightarrow[h]{} & (S^2, P) \end{array}$$

commutes. They prove that H_f has finite index in G_P and that there is a homomorphism $\phi_f\colon H_f \to G_P$. Hence there is a virtual homomorphism

$$\phi_f\colon G_P \dashrightarrow G_P.$$

Though I've given the pullback invariants in the dynamic setting for a map $f\colon (S^2, P) \to (S^2, P)$, Koch, Pilgrim, and Selinger work more generally in the nondynamic setting of a map $f\colon (S^2, A) \to (S^2, B)$. The requirements on A and B are that both are finite, $f(A) \subseteq B$, and B contains the set of critical values. If these conditions are satisfied, $f\colon (S^2, A) \to (S^2, B)$ is an *admissable cover*.

If $f\colon (S^2, A) \to (S^2, B)$ is an admissable cover, we have the pullback relation

$$\mathcal{S}_B \cup \{o\} \overset{f}{\leftarrow} \mathcal{S}_A \cup \{o\},$$

the linear transformation

$$\lambda_f\colon \mathbb{R}[\mathcal{S}_B] \to \mathbb{R}[\mathcal{S}_A],$$

the pullback map

$$\sigma_f\colon \mathcal{T}_B \to \mathcal{T}_A \text{ and its continuous extension } \sigma_f\colon \overline{\mathcal{T}}_B \to \overline{\mathcal{T}}_A,$$

the pullback correspondence

$$X \circ Y^{-1}\colon \mathcal{M}_B \rightrightarrows \mathcal{M}_A,$$

and the virtual homomorphism

$$\phi_f\colon G_B \dashrightarrow G_A.$$

Here are two of their theorems, one dealing with the case that the preimage of an essential curve always contains an essential component, and the other dealing with the case that the preimage of an essential curve never contains an essential component. In the dynamic setting that $A = B = P_f$, much of Theorem 12.8 appears in the Buff-Epstein-Koch-Pilgrim paper [4, Theorem 5.1].

Theorem 12.7 *Suppose $f\colon (S^2, A) \to (S^2, B)$ is an admissable cover. The following are equivalent.*

1. *No element of $\mathcal{S}_B$ is in the kernel of the linear transformation λ_f.*
2. *No nontrivial composition of positive Dehn twists in the elements of a multicurve in $S^2 \setminus B$ is in the kernel of the virtual homomorphism ϕ_f.*
3. *The pullback map σ_f takes $\partial\mathcal{T}_B$ into $\partial\mathcal{T}_A$.*
4. *The pullback correspondence $X \circ Y^{-1}\colon \mathcal{M}_B \rightrightarrows \mathcal{M}_A$ is proper.*

Theorem 12.8 *Suppose $f\colon (S^2, A) \to (S^2, B)$ is an admissable cover. The following are equivalent.*

1. *The pullback relation $\mathcal{S}_B \cup \{o\} \overset{f}{\leftarrow} \mathcal{S}_A \cup \{o\}$ is constant.*
2. *The linear transformation $\lambda_f\colon \mathbb{R}[\mathcal{S}_B] \to \mathbb{R}[\mathcal{S}_A]$ is constant.*
3. *The virtual homomorphism $\phi_f\colon G_B \dashrightarrow G_A$ is constant.*
4. *The pullback map $\sigma_f\colon \mathcal{T}_B \to \mathcal{T}_A$ is constant.*
5. *The pullback correspondence $X \circ Y^{-1}\colon \mathcal{M}_B \rightrightarrows \mathcal{M}_A$ is constant.*

In the dynamic setting of a rational map $f\colon (S^2, P) \to (S^2, P)$ with hyperbolic orbifold, they prove that if $X \circ Y^{-1}$ has an invariant, nonempty, compact subset then ϕ_f is contracting and $\overset{f}{\leftarrow}$ has a finite global attractor.

12.2.6 Examples

As we saw previously, in the Buff-Epstein-Koch-Pilgrim paper [4] they give examples of three different behaviors of the pullback map σ_f near a fixed point. One of these examples, $f(z) = \frac{3z^2}{z^3+1}$, is also extensively studied by Lodge in the papers [19, Section 4] and [20, Section 5].

For the quadratic polynomials (the rabbit, $f(z) = z^2 + i$, and the polynomial f_c where c has preperiod 2 and period 1) considered by Bartholdi and Nekrashevych in [1], one can get information on the pullback map σ_f because of knowing the associated g-map. For more information see their paper [1, Sections 5.1, 6.3, 7.2], Koch's thesis [14, Section 11], or the excellent discussion in Hubbard's book [12, Appendix C7]. In their paper [13], Gregory Kelsey and Lodge enumerate all of the combinatorial equivalence classes of non-Euclidean quadratic Thurston maps with at most four postcritical points. They show that every such map with exactly four postcritical points is a twist (by an element of the mapping class group) of a rational map with at most three postcritical points. They compute the wreath recursions for the g-maps corresponding to these rational maps with at most three postcritical points, and use this to enumerate the combinatorial equivalence classes.

The NET maps website [24] has detailed information about the pullback map for many examples. As mentioned previously, a nearly Euclidean Thurston (or NET) map f is a Thurston map such that there are exactly four postcritical points and every critical point has degree 2. Since there are exactly four postcritical points, any nonempty multicurve in $S^2 \setminus P_f$ consists of a single curve, and any essential curve can be parametrized by its slope, an element of $\overline{\mathbb{Q}} = \mathbb{Q} \cup \{\infty\}$. If we consider only essential preimages as long as there is an essential preimage, the pullback relation on curves can be viewed as a *slope function* $\mu_f \colon \overline{\mathbb{Q}} \to \overline{\mathbb{Q}} \cup \{o\}$.

In their paper [6, Theorem 2.1] giving the basic theory of NET maps, Cannon, Floyd, Parry, and Pilgrim show that any NET map is a composition of a Euclidean Thurston map and a push homeomorphism. The Euclidean Thurston map is induced from a map $x \mapsto Ax + b$ of the plane to itself, where the entries of A and b are integers. It can be described by the columns of A (λ_1 and λ_2) and by the translation vector b, which we can assume is in the set $\{0, \lambda_1, \lambda_2, \lambda_1 + \lambda_2\}$. The push homeomorphism can be described up to isotopy by four integral points in the parallelogram in the plane spanned by $2\lambda_1$ and λ_2. The slope function can be computed from this combinatorial input. If $p/q \in \overline{\mathbb{Q}}$ and $\mu_f(p/q) \notin \{p/q, o\}$, then by the half-space theorem [6, Theorem 6.7] there is an open interval in $\mathbb{R} \cup \{\infty\}$ containing p/q which does not contain the slope of an annular obstruction.

Parry created his program NETmap [25] to input combinatorial data for a NET map and then compute the slope function on all curves with slope p/q satisfying $|p|, |q| \leq N$, where N is a bound supplied by the user. The program has been very successful in determining whether or not a given NET map is combinatorially equivalent to a rational map, and Floyd, Parry, and Pilgrim prove in [10, Corollary 1.2] that this rationality question is decidable for NET maps. In conjunction with the work on NET maps stemming from an AIM SQuaRE group consisting of Floyd, Kelsey, Koch, Lodge, Parry, Pilgrim, and Saenz, the program was expanded greatly by Parry. Among other things, the program gives detailed information on the subgroups of liftable elements for the mapping class group, for the pure mapping class group, and for the extended mapping class group. Using the information it produces for the subgroup of liftable elements for the extended mapping class group, it gives a guess for a fundamental region for the pullback map.

Here is an example which is discussed in detail in the documentation section of [24]. Let $\lambda_1 = (2, 0)$, let $\lambda_2 = (2, 2)$, and let $b = (0, 0)$. Let A be the 2×2 matrix with columns λ_1 and λ_2. Let Λ be the lattice in $\mathbb{R}^2$ with basis $\{\lambda_1, \lambda_2\}$, and let Γ be the group of isometries of $\mathbb{R}^2$ of the form $x \mapsto 2\lambda \pm x$ for some $\lambda \in \Lambda$. Let $S = \mathbb{R}^2/\Gamma$, and let $\pi \colon \mathbb{R}^2 \to S$ be the quotient map. Then S is a 2-sphere, and the map $x \mapsto Ax$ on $\mathbb{R}^2$ descends to the quotient space to a Euclidean Thurston map $g \colon S \to S$. Figure 12.1, which is one of 14 output files from NETmap for this example, shows a fundamental domain F for the action of Γ on $\mathbb{R}^2$. Let $h \colon S^2 \to S^2$ be the push homeomorphism that is supported in a neighborhood of the images of the gray arcs, and such that $h(\pi(0, 0)) = \pi(3, 1)$, $h(\pi(2, 0)) = \pi(1, 0)$, and $h(\pi(2, 2)) = \pi(1, 1)$. Our Thurston map is $f = h \circ g$.

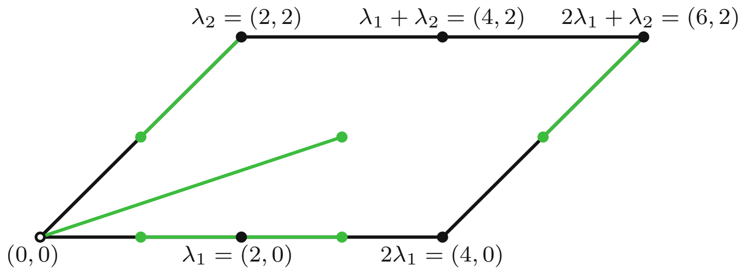


Fig. 12.1 A presentation diagram for the NET map f

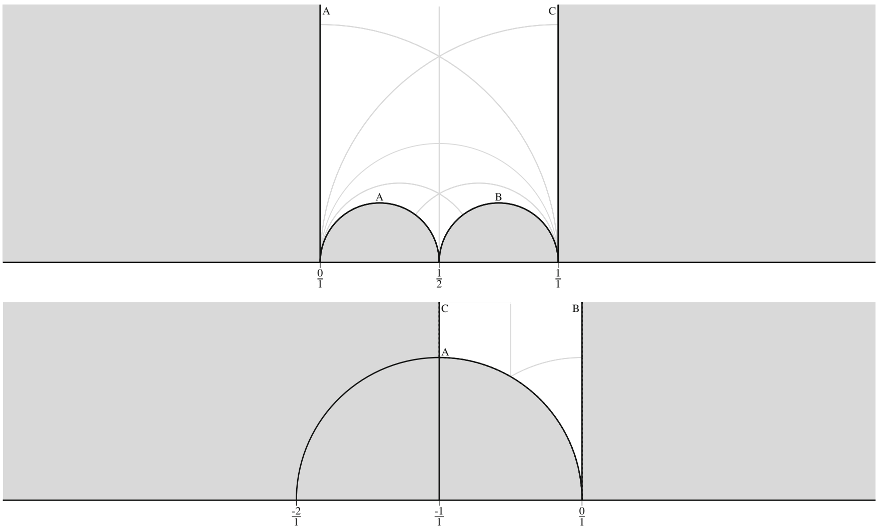


Fig. 12.2 The pullback map σ_f

Figure 12.2, which is another output file from NETmap for this example, gives information about the pullback map σ_f. The top of the figure is part of the domain of σ_f and the bottom of the figure is part of the codomain. The top shows a (white) quadrilateral bounded by four geodesics. Each of these geodesics is a reflection arc for the pullback map σ_k of an orientation-reversing, liftable homeomorphism k. Each of these reflection arcs maps under σ_f into the associated reflection arc of the lifted homeomorphism. The arcs are labeled A, B, and C so that σ_f preserves labels. Viewing the quadrilateral as a conformal triangle with vertices at $1/2$, 1, and ∞, there is a unique conformal homeomorphism taking it to the (white) triangle at the bottom of the figure and taking $1/2$ to 0, 1 to ∞, and ∞ to $-1+i$. Using a little more information from the other output files, one can show that σ_f is the analytic extension of this map to the upper-half plane by means of the reflection principle. See the NET map documentation at [24] for more details about this example. For a detailed description of another example, see the Floyd-Kelsey-Koch-Lodge-Parry-Pilgrim-Saenz paper [9, Section 6].

The example above is just one of an extensive collection of examples described on the NET maps website. For each of the 10, 626 possible dynamic portraits of a NET map of degree at most 40, it has input and output files for a NET map with that dynamic portait. And for each of the 46, 265 Hurwitz classes of NET maps of degree at most 30, there are input and output files for a NET map in that Hurwitz class.

12.2.7 Eigenvalues of the Pullback Map

Suppose f is a postcritically finite rational map, and suppose for convenience that the orbifold $\mathcal{O}_f$ is hyperbolic. Then σ_f has a unique fixed point τ_f and $\sigma_f^{\circ 2}$ is a weak contraction. What can we say about the eigenvalues of $D\sigma_f(\tau_f)$? If we fix an upper bound D on the degree, what can we say about the set of eigenvalues of the derivatives $D\sigma_f(\tau_f)$ for f a postcritically finite rational map with degree D?

Buff, Epstein, and Koch consider these and related questions in their paper [5]. To compute the eigenvalues of $D\sigma_f$, they study the coderivative of σ_f. Let $\mathcal{Q}(\widehat{\mathbb{C}})$ be the space of meromorphic quadratic differentials on $\widehat{\mathbb{C}}$ with simple poles, and let $\mathcal{Q}_f$ be the subspace of $\mathcal{Q}(\widehat{\mathbb{C}})$ of quadratic differentials whose poles are contained in P_f. There is a pushforward operator (called the *Thurston pushforward operator* in [5]) $f_*\colon \mathcal{Q}(\widehat{\mathbb{C}}) \to \mathcal{Q}(\widehat{\mathbb{C}})$. It restricts to a pushforward operator $f_*\colon \mathcal{Q}_f \to \mathcal{Q}_f$ and there is an induced operator $f_*\colon \mathcal{Q}(\widehat{\mathbb{C}})/\mathcal{Q}_f \to \mathcal{Q}(\widehat{\mathbb{C}})/\mathcal{Q}_f$. Let Σ_f be the set of eigenvalues of $f_*\colon \mathcal{Q}_f \to \mathcal{Q}_f$ and let Λ_f be the set of eigenvalues of $f_*\colon \mathcal{Q}(\widehat{\mathbb{C}})/\mathcal{Q}_f \to \mathcal{Q}(\widehat{\mathbb{C}})/\mathcal{Q}_f$. Then $\mathcal{Q}_f$ is isomorphic to the cotangent space of $\mathcal{T}_{P_f}$ at the point τ_f, and $f_*\colon \mathcal{Q}_f \to \mathcal{Q}_f$ is isomorphic to the coderivative of σ_f at τ_f.

Here are two of their theorems. The first is about the eigenvalues of a single map, and the second is about the eigenvalues of a family of unicritical maps.

Theorem 12.9 (Buff-Epstein-Koch) *Suppose $f\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is a postcritically finite rational map.*

i) The elements of Σ_f are algebraic numbers. If $\lambda \in \Sigma_f$ and λ is an algebraic integer, then either $\lambda = 0$ or f is a Lattès map and $\lambda \in \{\pm 1, \pm i, \frac{1}{2} \pm i\frac{\sqrt{3}}{2}, -\frac{1}{2} \pm i\frac{\sqrt{3}}{2}\}$.

ii) The elements of Λ_f are algebraic numbers. If $\lambda \in \Lambda_f$ is an algebraic integer, then $\lambda = 0$. Furthermore, $\Lambda_f = \{0\} \cup \{\lambda \in \widehat{\mathbb{C}} \setminus \{0\}$: there exist a positive integer m and a cycle of f of period m which has multiplier $1/\lambda^m$ and is not contained in $P_f\}$.

Given an integer $D \geq 2$, let $\Lambda(D) = \cup_f \Lambda_f$ and let $\Sigma(D) = \cup_f \Sigma(f)$, where in both cases the union is over all unicritical polynomials f of degree D whose finite critical point is periodic. Let $r_D = \frac{1}{2D}$ if D is even and let $r_D = \frac{1}{2D\cos(\pi/(2D))}$ if D is odd.

Theorem 12.10 (Buff-Epstein-Koch) *Let $D \geq 2$ be an integer.*

i) $\Sigma(D) \subset \{z \in \mathbb{C} : \frac{1}{4D} < |z| < 1\}$ *and* $\{z \in \mathbb{C} : r_D \leq |z| \leq 1\} \subset \overline{\Sigma(D)}$.
ii) $\Lambda(D) \subset \{z \in \mathbb{C} : \frac{1}{2D} < |z| < 1\}$.

Acknowledgements I thank Sarah Koch, Walter Parry, Kevin Pilgrim, and Edgar Saenz for helpful comments on an earlier version of this paper.

References

1. L. Bartholdi, V. Nekrashevych, Thurston equivalence of topological polynomials. Acta Math. **197**(1), 1–51 (2006)
2. B. Bielefeld, Y. Fisher, J. Hubbard, The classification of critically preperiodic polynomials as dynamical systems. J. Am. Math. Soc. **5**(4), 721–762 (1992)
3. X. Buff, G. Cui, L. Tan, Teichmüller spaces and holomorphic dynamics, in *Handbook of Teichmüller Theory, Vol. IV*, ed. by A. Papadopoulos, vol. 19. IRMA Lect. Math. Theor. Phys. (Eur. Math. Soc., Zürich, 2014), pp. 717–756
4. X. Buff, A. Epstein, S. Koch, K. Pilgrim, On Thurston's pullback map, in *Complex Dynamics* (A K Peters, Wellesley, MA, 2009), pp. 561–583
5. X. Buff, A.L. Epstein, S. Koch, Eigenvalues of the Thurston operator. J. Topol. **13**(3), 969–1002 (2020)
6. J.W. Cannon, W.J. Floyd, W.R. Parry, K.M. Pilgrim, Nearly Euclidean Thurston maps. Conform. Geom. Dyn. **16**, 209–255 (2012)
7. A. Douady, Systèmes dynamiques holomorphes, in *Bourbaki seminar, Vol. 1982/83*. Astérisque, vol. 105 (Soc. Math. France, Paris, 1983), pp. 39–63
8. A. Douady, J.H. Hubbard, A proof of Thurston's topological characterization of rational functions. Acta Math. **171**(2), 263–297 (1993)
9. W. Floyd, G. Kelsey, S. Koch, R. Lodge, W. Parry, K. M. Pilgrim, E. Saenz, Origami, affine maps, and complex dynamics. Arnold Math. J. **3**(3), 365–395 (2017)
10. W. Floyd, W. Parry, K.M. Pilgrim, Rationality is decidable for nearly Euclidean Thurston maps. Geom. Dedicata **213**, 487–512 (2021)
11. J.H. Hubbard, *Teichmüller Theory and Applications to Geometry, Topology, and Dynamics. Vol. 1*. Matrix Editions, Ithaca, NY, 2006. Teichmüller theory, With contributions by A. Douady, W. Dunbar, R. Roeder, S. Bonnot, D. Brown, A. Hatcher, C. Hruska and S. Mitra, With forewords by W. Thurston and C. Earle.
12. J.H. Hubbard, *Teichmüller Theory and Applications to Geometry, Topology, and Dynamics. Vol. 2*. Matrix Editions, Ithaca, NY, 2016. Surface homeomorphisms and rational functions
13. G. Kelsey, R. Lodge, Quadratic Thurston maps with few postcritical points. Geom. Dedicata **201**, 33–55 (2019)
14. S.C.H. Koch, La théorie de Teichmüller et ses applications aux endomorphismes de $\mathbb{P}^n$. PhD thesis, Université de Provence, 2007
15. S.C.H. Koch, A new link between Teichmüller theory and complex dynamics. PhD thesis, Cornell University, 2008
16. S. Koch, Teichmüller theory and critically finite endomorphisms. Adv. Math. **248**, 573–617 (2013)
17. S. Koch, K.M. Pilgrim, N. Selinger, Pullback invariants of Thurston maps. Trans. Am. Math. Soc. **368**(7), 4621–4655 (2016)

18. S.V.F. Levy, Critically finite rational map. PhD thesis, Princeton University, 1985
19. R. Lodge, Boundary values of the Thurston pullback map. PhD thesis, Indiana University, 2012
20. R. Lodge, Boundary values of the Thurston pullback map. Conform. Geom. Dyn. **17**:77–118 (2013)
21. J. Milnor, W. Thurston, On iterated maps of the interval, in *Dynamical Systems (College Park, MD, 1986–87)*, vol. 1342. Lecture Notes in Mathematics (Springer, Berlin, 1988), pp. 465–563
22. R.L. Moore, Concerning upper semi-continuous collections of continua. Trans. Am. Math. Soc. **27**(4), 416–428 (1925
23. D. Mumford, A remark on Mahler's compactness theorem. Proc. Am. Math. Soc. **28**, 289–294 (1971)
24. The NET map website, http://www.math.vt.edu/netmaps/
25. W. Parry, `NETmap`, software, available from http://www.math.vt.edu/netmaps/
26. K.M. Pilgrim, Canonical Thurston obstructions. Adv. Math. **158**(2), 154–168 (2001)
27. E.A. Saenz Maldonado, On nearly Euclidean Thurston maps. PhD thesis, Virginia Polytechnic Institute and State University, 2012
28. E.A. Saenz, On NET maps: examples and nonexistence results. Conform. Geom. Dyn. **23**, 147–163 (2019)
29. P. Scott, The geometries of 3-manifolds. Bull. London Math. Soc. **15**(5), 401–487 (1983)
30. N. Selinger, On Thurston's characterization theorem for branched covers. PhD thesis, Jacobs University, 2011
31. N. Selinger, Thurston's pullback map on the augmented Teichmüller space and applications. Invent. Math. **189**(1), 111–142 (2012)
32. M. Shishikura, On a theorem of M. Rees for matings of polynomials, in *The Mandelbrot Set, Theme and Variations*, vol. 274. London Math. Soc. Lecture Note Ser. (Cambridge Univ. Press, Cambridge, 2000), pp. 289–305
33. M. Shishikura, T. Lei, A family of cubic rational maps and matings of cubic polynomials. Exp. Math. **9**(1), 29–53 (2000)
34. K. Strebel, *Quadratic Differentials*, volume 5 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]* (Springer, Berlin, 1984)
35. L. Tan, Accouplements de polynômes complexes. PhD thesis, Université de Paris-Sud, 1987
36. L. Tan, Matings of quadratic polynomials. Ergodic Theory Dynam. Syst. **12**(3), 589–620 (1992)
37. L. Tan (ed.), *The Mandelbrot Set, Theme and Variations*, vol. 274. London Mathematical Society Lecture Note Series (Cambridge University Press, Cambridge, 2000)

Chapter 13
A Classification of Postcritically Finite Newton Maps

Russell Lodge, Yauhen Mikulich, and Dierk Schleicher

Abstract The dynamical classification of rational maps is a central concern of holomorphic dynamics. Much progress has been made, especially on the classification of polynomials and some approachable one-parameter families of rational maps; the goal of finding a classification of general rational maps is so far elusive. Newton maps (rational maps that arise when applying Newton's method to a polynomial) form a most natural family to be studied from the dynamical perspective. Using Thurston's characterization and rigidity theorem, a complete combinatorial classification of postcritically finite Newton maps is given in terms of a finite connected graph satisfying certain explicit conditions.

Keywords Newton map · Rational map · Parameter space · Renormalization · Hubbard tree · Combinatorial classification · Extended Newton graph · Thurston's theorem

2010 Mathematics Subject Classification Primary 30D05, 37F10, 37F20

R. Lodge
Department of Mathematics and Computer Science, Indiana State University, Terre Haute, IN, USA
e-mail: russell.lodge@indstate.edu

Y. Mikulich
Bexleyheath, UK

D. Schleicher (✉)
Aix-Marseille Université and CNRS, UMR 7373, Institut de Mathématiques de Marseille, Marseille, France
e-mail: dierk.schleicher@univ-amu.fr

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_13

13.1 Introduction

The past four decades have seen tremendous progress in the understanding of holomorphic dynamics. This is largely due to the fact that the complex structure provides enough rigidity, allowing many interesting questions to be reduced to tractable combinatorial problems.

To understand the dynamics of rational maps, an important first step is to understand the dynamics of postcritically finite maps, namely the maps where each critical point has finite forward orbit. Thurston's "Fundamental Theorem of Complex Dynamics" [5] is available in this setting, providing an important characterization and rigidity theorem for postcritically finite branched covers that arise from rational maps. Also, the postcritically finite maps are the structurally important ones, and conjecturally, the set of maps that are quasiconformally equivalent (in a neighborhood of the Julia set) to such maps are dense in parameter spaces [18, Conjecture 1.1].

Polynomials form an important and well-understood class of rational functions. In this case, the point at infinity is completely invariant, and is contained in a completely invariant Fatou component. This permits enough dynamical structure so that postcritically finite polynomials may be described in finite terms, e.g., using external angles at critical values or finite Hubbard trees. A complete classification of postcritically finite polynomials has been given [1, 22].

Classification results for families of rational functions are rare and mostly concern one-dimensional families. There is a recent classification of critically fixed rational maps [4, 12] and critically fixed anti-rational maps [10, 14] in arbitrary degree, but these classification results do not address higher period critical points. The noteworthy family that exceeds all these limitations is the family classified here: Newton maps.

Definition 13.1.1 (Newton Map) A rational function $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ is called a *Newton map* if there is some complex polynomial $p(z)$ so that $f(z) = z - \frac{p(z)}{p'(z)}$ for all $z \in \mathbb{C}$.

Denote such a Newton map by N_p. Newton maps of degree 1 and 2 are trivial and thus excluded from our entire discussion.

Note that N_p is precisely the function that is iterated when Newton's method is used to find the roots of the polynomial p. Each root of p is an attracting fixed point of the Newton map, and the point at infinity is a repelling fixed point. The algebraic number of roots of p is the degree of p, while the geometric number of roots of p (ignoring multiplicities) equals the degree of N_p. The space of degree d Newton maps considered up to affine conjugacy has $d-2$ degrees of freedom, given by the location of the d roots of p after affine conjugation. The space of degree d complex polynomials up to affine conjugacy has $d - 1$ degrees of freedom, given by the $d + 1$ coefficients after affine conjugation. Thus it is clear that Newton maps form a substantial subclass of rational maps, making the combinatorial classification all the more remarkable. A brief summary of "extended Newton graphs" is given below in the introduction, where it should be noted that arbitrary choices must be made in the construction of so-called "Newton rays" with the result that there is no unique extended Newton graph associated with a Newton map. The abstract graph definition and combinatorial equivalence are given in Definitions 13.4.5 and 13.5.13).

Main Theorem A (Classification of Postcritically Finite Newton Maps) *There is a natural bijection between the set of postcritically finite Newton maps (up to affine conjugacy) and the set of abstract extended Newton graphs (up to combinatorial equivalence) so that for every abstract extended Newton graph* (Σ, f)*, the associated postcritically finite Newton map has the property that any associated extended Newton graph is equivalent to* (Σ, f).

In the study of Newton maps, an important first theorem is the following characterization in terms of fixed point multipliers.

Proposition 13.1.2 (Head's Theorem [11]) *A rational map* f *of degree* $d \geq 3$ *is a Newton map if and only if for each fixed point* $\xi \in \mathbb{C}$*, there is an integer* $m \geq 1$ *so that* $f'(\xi) = (m-1)/m$.

This condition on multipliers forces ∞ to be a repelling fixed point by the holomorphic fixed point formula. In fact, for a postcritically finite Newton map all the finite fixed points must be superattracting (otherwise, the immediate basin contains a critical point that converges to the root), and this corresponds to the roots of p being simple.

There are a number of partial classification theorems for postcritically finite Newton maps. Tan Lei has given a classification of cubic Newton maps in terms of matings and captures (or alternatively in terms of abstract graphs [29]; see also earlier work by Head [11]). Luo produced a similar combinatorial classification for Newton maps of arbitrary degree subject to the condition that there is only a single non-fixed critical value, and this critical value is either periodic or eventually maps to a fixed critical point [16].

The classification of postcritically *fixed* Newton maps for arbitrary degree (those Newton maps whose critical points eventually land on fixed points) is given in [7] building on the work of [25]. The fundamental piece of combinatorial data is the *channel diagram* Δ which is constructed in [13]. This is a graph in the Riemann sphere whose vertices are given by the fixed points of the Newton map and whose edges are given by all accesses of the immediate basins of roots connecting the roots to ∞ (see the solid lines of Fig. 13.1). To capture the behavior of non-periodic critical points that eventually map to the channel diagram, it is natural to consider[1] the graph $N_p^{-n}(\Delta)$ for some integer n. However this graph is not necessarily connected (see Fig. 13.1 for an example), and so the *Newton graph of level* n associated with N_p is defined to be the component of $N_p^{-n}(\Delta)$ that contains Δ. It is shown in [7] that for any postcritically *finite* Newton map N_p, there is some

[1] We denote the n-th iterate of a dynamical system $f : X \to X$ by $f^n : X \to X$.

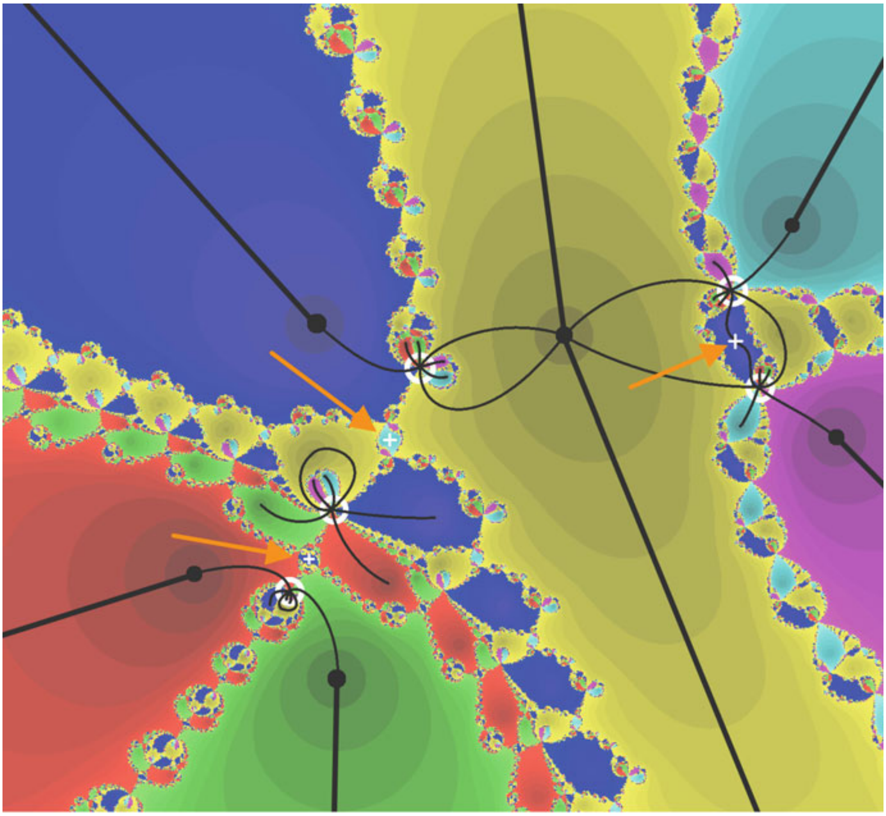


Fig. 13.1 Dynamical plane of a degree 6 Newton map N_p. The six roots are indicated by black dots. The five finite poles are indicated by white circles. The channel diagram Δ is drawn with thick black curves, and $N_p^{-1}(\Delta) \setminus \Delta$ is drawn with thin black curves. The Newton graph of level one Δ_1 is visible as the component of $N_p^{-1}(\Delta)$ that contains Δ. Note that $N_p^{-1}(\Delta) \setminus \Delta_1$ is nonempty and contains one connected component. There are three non-fixed simple critical points indicated by orange arrows and a white "+". The rightmost such critical point is mapped by N_p to the root in the blue basin, but more than one iterate is required to map the other two critical points to a root. There are no free critical points

level n so that the Newton graph of level n contains all critical points that eventually map to the channel diagram (this fact is non-trivial because preimage components of the channel diagram were discarded). For minimal n this component is called the *Newton graph* in the context of postcritically fixed maps, and the data consisting of this graph equipped with a graph map inherited from the dynamics of the Newton map is enough to classify postcritically *fixed* Newton maps.

We classify postcritically *finite* Newton maps, building on work of [19]. The chief difficulty in this generalized setting is the existence of critical points whose forward orbit does not contain a fixed point. We thus call a critical point *free* if it is not contained in the Newton graph Δ_n for any level n. In [15] a finite graph containing the postcritical set was constructed for a postcritically finite Newton map. The graph is composed of three types of pieces:

- the Newton graph (which contains the channel diagram) is used to capture the behavior of critical points that are eventually fixed.
- Hubbard trees are used to give combinatorial descriptions of renormalizations at periodic non-fixed postcritical points. See [8] for the construction of the renormalization. Preimages of the Hubbard trees are taken to capture the behavior of critical points whose orbits intersect the Hubbard trees (or equivalently the free critical points).
- Newton rays (single edges comprised of either an internal ray or a sequence of infinitely many preimages of channel diagram edges) are used to connect all Hubbard trees and their preimages to the Newton graph.

The construction of these three types of edges is given in [15] and not reproduced here, but an example is provided in Figs. 13.2 and 13.3.

The restriction of the Newton map to this "extended Newton graph" yields a graph self-map, and the resulting dynamical graphs are axiomatized (as abstract extended Newton graphs; see Definition 13.4.5).

Theorem 13.1.3 (Newton Maps to Graphs; [15, Theorem 1.2]) *For any extended Newton graph* $\Delta^*_{\mathcal{N}} \subset \widehat{\mathbb{C}}$ *associated with a postcritically finite Newton map* N_p*, the pair* $(\Delta^*_{\mathcal{N}}, N_p)$ *satisfies the axioms of an abstract extended Newton graph.*

It must be emphasized that arbitrary choices were made in the construction of the Newton rays, necessitating a rather subtle but natural combinatorial equivalence relation on our way to a classification.

Our first main result on the way to the classification is that every abstract extended Newton graph is realized by a unique Newton map (up to affine conjugacy); it is proven using Thurston's theorem. In the following theorem statement, $\overline{f}$ denotes the unique extension (up to Thurston equivalence) of the graph map f to a branched cover of the whole sphere, and the set of vertices of a graph Γ is denoted by Γ'.

Main Theorem B (Graphs to Newton Maps) *Let* (Σ, f) *be an abstract extended Newton graph (as in Definition 13.4.5). Then there is a postcritically finite Newton map* N_p*, unique up to affine conjugacy, with extended Newton graph* $\Delta^*_{\mathcal{N}}$ *so that the marked branched covers* $(\overline{f}, \Sigma')$ *and* $(N_p, (\Delta^*_{\mathcal{N}})')$ *are Thurston equivalent.*

Denote by **Newt** the set of postcritically finite Newton maps up to affine conjugacy, and by **NGraph** we denote the set of abstract extended Newton graphs under the graph equivalence of Definition 13.5.13. It follows from the statements of Theorem 13.1.3 and B that there are well-defined maps

$$\mathcal{F} : \textbf{Newt} \to \textbf{NGraph} \quad \text{and} \quad \mathcal{F}' : \textbf{NGraph} \to \textbf{Newt}$$

respectively. It will be shown that the mappings $\mathcal{F}$ and $\mathcal{F}'$ are bijective, and inverses of each other, yielding Main Theorem A.

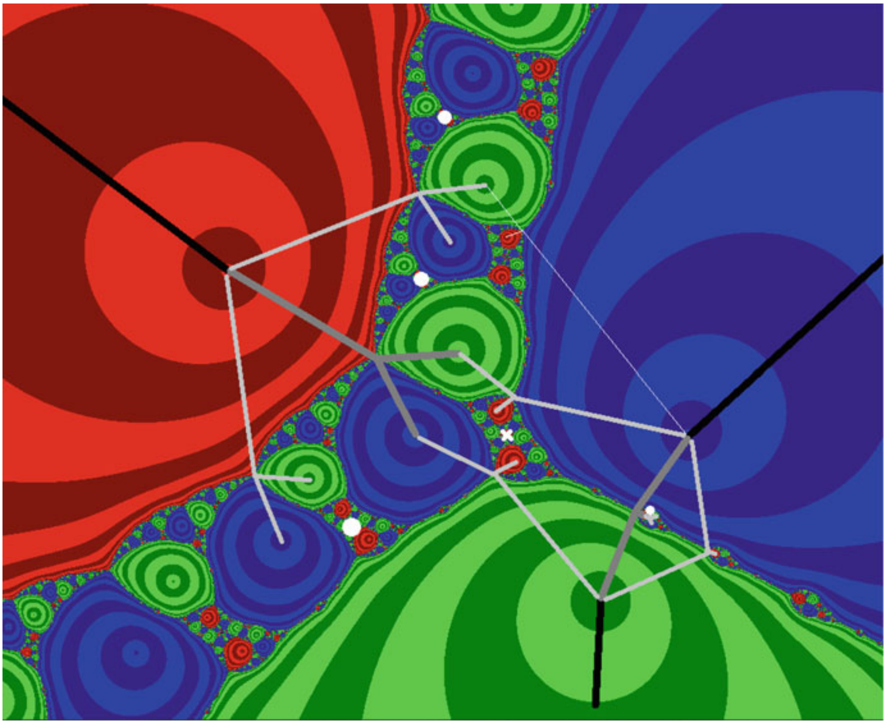


Fig. 13.2 The dynamical plane of a cubic Newton map N_p displaying part of the extended Newton graph. The centers of the biggest red, green, and blue basins are fixed critical points. The white "x" indicates a free critical point, and its orbit is indicated by white dots. It has period 5 and the corresponding polynomial-like map straightens to $z \mapsto z^2$ (which is visible in Fig. 13.3). Thick black edges indicate the channel diagram Δ, and successively lighter edges indicate the additional edges in Δ_1, Δ_2 and Δ_3 (due to the small scale many edges from Δ_3 are omitted, though enough are drawn to separate the filled Julia sets as required by the construction). The combinatorial invariant for N_p consists of Δ_3, the Hubbard trees of each of the five filled Julia sets, and five Newton rays connecting them to Δ_3. A Hubbard tree/Newton ray pair is exhibited in the zoom of Fig. 13.3. (Both images produced by Wolf Jung's Mandel software)

Remark 13.1.4 This paper not only provides a classification of the largest non-polynomial family of rational maps so far, it also lays foundations for classification and rigidity results in a substantially larger context. In particular, there is the following rigidity result: all Newton maps, postcritically finite or not, are *rigid* in the sense that any two such maps can be distinguished in purely combinatorial terms (plus conformal invariants such as multipliers of attracting cycles), except when they admit embedded polynomial-like dynamics that fails to be rigid [6] (see also [24] in the non-renormalizable case). In parallel, strong results about local connectivity for the Julia sets of Newton maps are developed. In particular, the boundary of every component of the basin of a root is locally connected; this was also shown independently in [30].

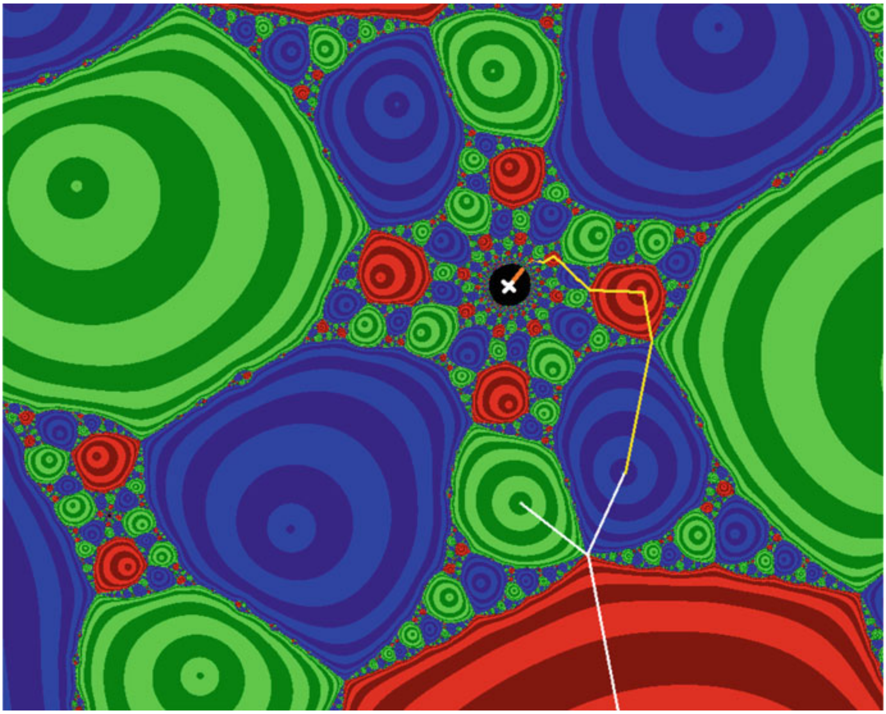


Fig. 13.3 Zoom of Fig. 13.2 at the free critical point. The filled Julia set is visible as a black disk, and its Hubbard tree is drawn in orange. The white edges are level 3 edges in the Newton graph Δ_3. The yellow edge is a period 5 Newton ray

The fundamental property of the dynamics that is underlying our work is that Fatou components have a common accessible boundary point at infinity, as well as the preimages of these Fatou components. A basic ingredient in more general classification and rigidity results builds on periodic Fatou components with common accessible boundary points, and for these our methods will be a key ingredient.

We also mention work of Mamayusupov [17] that establishes a bijection between the set of rational maps that arise as Newton maps of transcendental entire functions and the set of postcritically finite Newton maps of our study. Finally, we should also mention that Newton's method is much better at actually finding roots of complex polynomials than its reputation sometimes predicts; see for instance [23, 26–28].

Structure of This Paper Section 13.2 introduces Thurston's characterization and rigidity theorem for postcritically finite branched covers. This theorem asserts that a topological branched cover that has no obstructing multicurves is uniquely realized by a rational map (under a mild assumption that is irrelevant for our purposes), and that such multicurves are the only possible obstructions for existence. Since it is often very hard to show directly that a cover is unobstructed, we describe a theorem of Pilgrim and Tan that is very useful for this purpose, controlling the location of obstructions. Section 13.3 presents a result on how to extend certain kinds of graph maps to branched covers on the whole sphere.

Section 13.4 defines the abstract extended Newton graph, which will be shown to be a complete invariant for postcritically finite Newton maps. The equivalence on such graphs is defined in Sect. 13.5, and the connection between this combinatorial equivalence and Thurston equivalence is described.

Section 13.6 proves Theorem B by showing that abstract extended Newton graphs equipped with their graph self-maps extend to branched covers of the sphere that are unobstructed.

Section 13.7 proves Theorem A, establishing the combinatorial classification of postcritically finite Newton maps.

13.2 Thurston Theory on Branched Covers

We will be using Thurston's theorem to prove that the combinatorial model for postcritically finite Newton maps is realized by a rational map, and we present the requisite background in this section. As one observes from the statement of Thurston's theorem below, this amounts to showing that the combinatorial model has no obstructing multicurves. There are infinitely many multicurves in a sphere with four or more marked points, so a priori it is very hard to show obstructions do not exist. However, the "arcs intersecting obstructions" theorem of Pilgrim and Tan can in some cases drastically reduce the possible locations of obstructions.

Let $f : \mathbb{S}^2 \to \mathbb{S}^2$ be an orientation-preserving branched cover from the two-sphere to itself. Denote the set of critical points of f by C_f. Define the postcritical set P_f as follows:

$$P_f := \bigcup_{n \geq 1} f^n(C_f).$$

The map f is said to be *postcritically finite* if the set P_f is finite.

A *marked branched cover* is a pair (f, X), where $f : \mathbb{S}^2 \to \mathbb{S}^2$ is an orientation-preserving branched cover and X is a finite set containing P_f such that $f(X) \subset X$.

Definition 13.2.1 (Thurston Equivalence of Marked Branched Covers) Two marked branched covers (f, X) and (g, Y) are *Thurston equivalent* if there are two orientation-preserving homeomorphisms $\phi_1, \phi_2 : \mathbb{S}^2 \to \mathbb{S}^2$ such that

$$\phi_1 \circ f = g \circ \phi_2$$

and there exists an homotopy $\Phi : [0, 1] \times \mathbb{S}^2 \to \mathbb{S}^2$ with $\Phi(0, \cdot) = \phi_1$ and $\Phi(1, \cdot) = \phi_2$ such that $\Phi(t, \cdot)|_X$ is constant in $t \in [0, 1]$ with $\Phi(t, X) = Y$. If ϕ_1 and ϕ_2 are homotopic to the identity map, the marked branched covers are said to be *homotopic*.

It is perhaps more intuitive to rephrase Thurston equivalence as saying that f and g are given only up to homotopy rel the marked points, and this homotopy may be chosen so that f and g are topologically conjugate.

We say that a simple closed curve γ is a *simple closed curve in* $(\mathbb{S}^2, X)$ if $\gamma \subset \mathbb{S}^2 \setminus X$. Such a γ is *essential* if both components of the complement $\mathbb{S}^2 \setminus \gamma$ contain at least two points of X. Let γ_0, γ_1 be two simple closed curves in $(\mathbb{S}^2, X)$. We say that γ_0 and γ_1 are *isotopic relative to* X, written $\gamma_0 \simeq_X \gamma_1$, if there exists a continuous, one-parameter family of simple closed curves in $(\mathbb{S}^2, X)$ joining γ_0 and γ_1. We use $[\gamma]$ to denote the isotopy class of a simple closed curve γ. A *multicurve* is a collection of pairwise disjoint and non-isotopic essential simple closed curves in $(\mathbb{S}^2, X)$. A multicurve Π is said to be f-stable if for every $\gamma \in \Pi$, every essential connected component of $f^{-1}(\gamma)$ is isotopic relative to X to some element of Π.

Definition 13.2.2 (Thurston Linear Transform) For every f-stable multicurve Π we define the corresponding *Thurston linear transform* $f_\Pi : \mathbb{R}^\Pi \to \mathbb{R}^\Pi$ as follows:

$$f_\Pi(\gamma) = \sum_{\gamma' \subset f^{-1}(\gamma)} \frac{1}{\deg(f|_{\gamma'} : \gamma' \to \gamma)} [\gamma'],$$

where $[\gamma']$ denotes the element of Π isotopic to γ' if it exists. If there are no such elements, the sum is taken to be zero. Denote by λ_Π the largest eigenvalue of f_Π (by the Perron–Frobenius theorem, it exists and is non-negative real).

The Thurston linear transform is also known equivalently as the *Thurston matrix* or *multicurve matrix*.

Suppose that Π is a stable multicurve. A multicurve Π is called a *multicurve obstruction* (or *Thurston obstruction*) if $\lambda_\Pi \geq 1$. A real-valued $n \times n$ matrix A is called *irreducible* if for every entry (i, j), there exists an integer $k > 0$ such that $A^k_{i,j} > 0$. A multicurve Π is said to be *irreducible* if the matrix representing the linear transform f_Π is irreducible.

The statement of Thurston's theorem uses the notion of a hyperbolic orbifold. We omit the definition, referring the reader to [5] while observing that there are only a few well-understood cases where O_f is not hyperbolic, and that O_f is always hyperbolic if f has at least three fixed branched points. The latter is always the case for Newton maps of degree $d \geq 3$, so the restriction to hyperbolic orbifolds is of no concern to us.

Theorem 13.2.3 (Thurston's Theorem [3, 5]) *A marked branched cover (f, X) with hyperbolic orbifold is Thurston equivalent to a marked rational map if and only if (f, X) has no multicurve obstruction. Furthermore, if (f, X) is unobstructed, the marked rational map is unique up to Möbius conjugacy.*

We now present a theorem of Pilgrim and Tan [21] that will be used in Sect. 13.6 to show that certain marked branched covers arising from graph maps do not have obstructions and are therefore equivalent to rational maps by Thurston's theorem. First some notation will be introduced.

Assume that (f, X) is a marked branched cover of degree $d \geq 3$. An *arc* in $(\mathbb{S}^2, X)$ is a continuous map $\alpha : [0, 1] \to \mathbb{S}^2$ such that $\alpha(0)$ and $\alpha(1)$ are in X, the map α is injective on $(0, 1)$, and $X \cap \alpha((0, 1)) = \emptyset$. A set of pairwise non-isotopic arcs in $(\mathbb{S}^2, X)$ is called an *arc system*.

The following intersection number is used in the statement of Theorem 13.2.5; we use the symbol $\simeq$ to denote isotopy relative to X.

Definition 13.2.4 (Intersection Number) Let α and β each be an arc or a simple closed curve in $(\mathbb{S}^2, X)$. Their *intersection number* is

$$\alpha \cdot \beta := \min_{\alpha' \simeq \alpha,\, \beta' \simeq \beta} \#\{(\alpha' \cap \beta') \setminus X\} .$$

This intersection can be extended to arc systems and multicurves as follows: let A and B each be an arc system or a multicurve in $(\mathbb{S}^2, X)$. Then

$$A \cdot B := \min_{A' \simeq A,\, B' \simeq B} \#\{(A' \cap B') \setminus X\} .$$

For an arc system Λ, we introduce a linear map $f_\Lambda : \mathbb{R}^\Lambda \to \mathbb{R}^\Lambda$, which is a rough analogue of the Thurston linear map for multicurves. For $\lambda \in \Lambda$, let

$$f_\Lambda(\lambda) := \sum_{\lambda' \subset f^{-1}(\lambda)} [\lambda']_\Lambda ,$$

where $[\lambda']_\Lambda$ denotes the element of Λ homotopic to λ' rel X (the sum is taken to be zero if there are no such elements). It is said that Λ is *irreducible* if the matrix representing f_Λ is irreducible.

Denote by $\tilde{\Lambda}(f^n)$ the union of those components of $f^{-n}(\Lambda)$ that are isotopic to elements of Λ relative X, and define $\tilde{\Pi}(f^n)$ for a multicurve Π analogously. The following is a special case of a theorem from [21] that gives control on the location of irreducible multicurve obstructions by asserting that they may not intersect certain preimages of irreducible arc systems.

Theorem 13.2.5 (Arcs Intersecting Obstructions [21, Theorem 3.2]) *Let (f, X) be a marked branched cover, Π an irreducible multicurve obstruction, and Λ an irreducible arc system. Suppose furthermore that $\#(\Pi \cap \Lambda) = \Pi \cdot \Lambda$. Then exactly one of the following is true:*

(1) $\Pi \cdot \Lambda = 0$ *and* $\Pi \cdot f^{-n}(\Lambda) = 0$ *for all* $n \geq 1$.

(2) $\Pi \cdot \Lambda \neq 0$ *and for* $n \geq 1$*, each component of* Π *is isotopic to a unique component of* $\tilde{\Pi}(f^n)$*. The mapping* $f^n : \tilde{\Pi}(f^n) \to \Pi$ *is a homeomorphism and* $\tilde{\Pi}(f^n) \cap (f^{-n}(\Lambda) - \tilde{\Lambda}(f^n)) = \emptyset$*. More precisely, for each* $\gamma \in \Gamma$*, there is exactly one curve* $\gamma' \subset f^{-n}(\gamma)$ *such that* $\gamma' \cap \tilde{\Lambda}(f^n) \neq \emptyset$*. Moreover, the curve* γ' *is the unique component of* $f^{-n}(\gamma)$ *which is isotopic to an element of* Π.

13.3 Extending Maps on Finite Graphs

We present a sufficient condition under which a certain type of map of a graph in $\mathbb{S}^2$ has a unique continuous extension to the whole sphere up to equivalence. The following formulation follows [1, Chapter 5]. The Alexander Trick is foundational to such results and will be used elsewhere.

Lemma 13.3.1 (Alexander Trick) *Let $h : \mathbb{S}^1 \to \mathbb{S}^1$ be an orientation-preserving homeomorphism. Then there exists an orientation-preserving homeomorphism $\overline{h} : \overline{\mathbb{D}} \to \overline{\mathbb{D}}$ such that $\overline{h}|_{\mathbb{S}^1} = h$. The map $\overline{h}$ is unique up to isotopy relative $\mathbb{S}^1$.*

Definition 13.3.2 (Finite Graph) Let V be a finite set of distinct points in $\mathbb{S}^2$. Each element of V is called a *vertex*. An *edge* is a subset of $\mathbb{S}^2$ of the form $\lambda(I)$ where $I = [0, 1]$ and

- $\lambda : I \to \mathbb{S}^2$ is continuous and injective on $(0, 1)$, and
- $\lambda(x) \in V \iff x \in \{0, 1\}$.

Let E be a finite set of edges that (pairwise) intersect only at vertices. A *finite graph* (in $\mathbb{S}^2$) is a pair of the form (V, E).

We sometimes omit the reference to the ambient space $\mathbb{S}^2$ though it is always implicit.

Definition 13.3.3 (Subgraphs) Let $\Gamma_1 = (V_1, E_1)$ and $\Gamma_2 = (V_2, E_2)$ be finite graphs. We say that Γ_1 is a *subgraph* of Γ_2 (denoted $\Gamma_1 \subset \Gamma_2$) if $V_1 \subset V_2$ and $E_1 \subset E_2$.

Definition 13.3.4 (Graph Map) Let Γ_1, Γ_2 be connected finite graphs. A continuous map $f : \Gamma_1 \to \Gamma_2$ is called a *graph map* if it is injective on each edge of Γ_1, if $f(V_1) \subset V_2$ and $f^{-1}(V_2) \subset V_1$, and f is compatible with the embeddings of the graphs in $\mathbb{S}^2$.

The compatibility condition on f is a local condition at each vertex v, described as follows. If f is locally injective at v, then f is required to preserve the cyclic ordering at v. On the other hand, if f is not locally injective at v, then the cyclic ordering of the half-edges at v and $f(v)$ should be compatible with a local orientation-preserving cover of degree $\deg_v f$. We will use this definition only when the number of half-edges at v equals $\deg_v f$ times the number of edges at $f(v)$, and thus compatibility means that the half-edges at v have the same cyclic ordering as $f(v)$, repeated $\deg_v f$ times.

Definition 13.3.5 (Regular Extension) Let $f : \Gamma_1 \to \Gamma_2$ be a graph map. An orientation-preserving branched cover $\overline{f} : \mathbb{S}^2 \to \mathbb{S}^2$ is called a *regular extension* of f if $\overline{f}|_{\Gamma_1} = f$ and $\overline{f}$ is injective on each component of $\mathbb{S}^2 \setminus \Gamma_1$.

It follows that every regular extension $\overline{f}$ may have critical points only at the vertices of Γ_1, and the local degree of $\overline{f}$ at v coincides with $\deg_v(f)$.

Lemma 13.3.6 (Isotopic Graph Maps[1, Corollary 6.3]) *Let $f, g : \Gamma_1 \to \Gamma_2$ be two graph maps that coincide on the vertices of Γ_1 such that for each edge e in Γ_1 we have $f(e) = g(e)$ as a set. Suppose that f and g have regular extensions $\overline{f}, \overline{g} : \mathbb{S}^2 \to \mathbb{S}^2$. Then there exists a homeomorphism $\psi : \mathbb{S}^2 \to \mathbb{S}^2$, isotopic to the identity relative the vertices of Γ_1, such that $\overline{f} = \overline{g} \circ \psi$.*

We must establish some notation for the following proposition from [1]. Let $f : \Gamma_1 \to \Gamma_2$ be a graph map. For each vertex v of Γ_i with fixed $i \in \{1, 2\}$, choose a neighborhood $U_v \subset \mathbb{S}^2$ such that all edges of Γ_i that enter U_v are incident to v, the vertex v is the only vertex of Γ_i in U_v, and the neighborhoods U_v and U_w are disjoint for all vertices $v \neq w$ in Γ_i. We may assume without loss of generality that in local coordinates, U_v is a round disk of radius 1 that is centered at v and that the intersection of any edge of Γ_i with U_v is either empty or a radial line segment. Without loss of generality, we may assume that $f|_{U_v \cap \Gamma_1}$ is length-preserving for all vertices v in Γ_1.

We describe how to explicitly extend f to each U_v. For a vertex $v \in \Gamma_1$, let γ_1 and γ_2 be two adjacent edges ending there. In local coordinates, these are radial lines at angles Θ_1, Θ_2 where $0 < \Theta_2 - \Theta_1 \leq 2\pi$ (if v is an endpoint of Γ_1, then set $\Theta_1 = 0$, $\Theta_2 = 2\pi$). In the same way, choose arguments Θ'_1, Θ'_2 for the image edges in $U_{f(v)}$ and extend f to a map $\tilde{f}$ on $\Gamma_1 \cup \bigcup_v U_v$ defined by

$$\tilde{f}(\rho, \Theta) = \left(\rho, \frac{\Theta'_2 - \Theta'_1}{\Theta_2 - \Theta_1} \cdot (\Theta - \Theta_1) + \Theta'_1\right), \tag{13.1}$$

where (ρ, Θ) are polar coordinates in the sector bounded by the rays at angles Θ_1 and Θ_2. In particular, sectors are mapped onto sectors in an orientation-preserving way.

Proposition 13.3.7 ([1, Proposition 5.4]) *A graph map $f : \Gamma_1 \to \Gamma_2$ has a regular extension if and only if for every vertex $y \in \Gamma_2$ and every component U of $\mathbb{S}^2 \setminus \Gamma_1$, the extension $\tilde{f}$ is injective on*

$$\bigcup_{v \in f^{-1}(y)} U_v \cap U \ .$$

The fundamental combinatorial object in our classification of Newton maps is a finite graph Σ equipped with a self-map $f : \Sigma \to \Sigma$ (Definition 13.4.5). Strictly speaking, f is in general not a graph map since Newton ray edges contain finitely many preimages of vertices in the Newton graph that are not vertices in Σ (these vertices are purposely ignored on our way to producing a finite graph). This motivates the following weaker definition which is identical except that we no longer assume $f^{-1}(V_2) \subset V_1$.

Definition 13.3.8 (Weak Graph Map) A continuous map $f : \Gamma_1 \to \Gamma_2$ is called a *weak graph map* if it is injective on each edge of Γ_1, if $f(V_1) \subset V_2$, and f is compatible with the embeddings of the graphs in $\mathbb{S}^2$.

Remark 13.3.8 Given a weak graph map $f : \Gamma_1 \to \Gamma_2$, the combinatorics of the domain can be slightly altered to produce a graph map $\hat{f} : \hat{\Gamma}_1 \to \Gamma_2$ in the following natural way. We take the graph $\hat{\Gamma}_1$ to have vertices given by $f^{-1}(V_2)$, and edges given by the closures of complementary components of $\Gamma_1 \setminus f^{-1}(V_2)$. We simply take $\hat{f} = f$.

13.4 Abstract Extended Newton Graph

In [15], we extracted from every postcritically finite Newton map an extended Newton graph (Sect. 13.6.1), and we axiomatized these graphs in Sect. 13.7. In this section we review the definition of the abstract extended Newton graph which will be used in Sect. 13.7 of the present work to classify postcritically finite Newton maps. Abstract extended Newton graphs consist of three pieces: abstract Newton graphs, abstract extended Hubbard trees, and abstract Newton rays connecting the first two objects.

The definition of abstract extended Hubbard trees was given in Definition 4.4 of [15], and will not be repeated here. We simply note that it is the usual definition of degree d abstract Hubbard tree from [22], where the set of marked points includes all periodic points of periods up to some fixed length n (since postcritically finite Newton maps cannot have parabolic cycles, the number of periodic points of period i equals d^i). Such an abstract extended Hubbard tree is said to have *cycle type n*.

To define the abstract Newton graph, it is necessary to first define the abstract channel diagram.

Definition 13.4.1 An *abstract channel diagram* of degree $d \geq 3$ is a graph $\Delta \subset \mathbb{S}^2$ with vertices $v_\infty, v_1, \ldots, v_d$ and edges $e_1, \ldots, e_l$ that satisfies the following:

- $l \leq 2d - 2$;
- each edge joins v_∞ to some v_i for $i \in \{1, 2, \ldots, d\}$;
- each v_i is connected to v_∞ by at least one edge;
- if e_i and e_j both join v_∞ to v_k, then each connected component of $\mathbb{S}^2 \setminus \overline{e_i \cup e_j}$ contains at least one vertex of Δ.

The classification of postcritically fixed Newton maps was given in terms of a combinatorial object called the "abstract Newton graph" [7]. We define the term almost identically except that in the following definition Condition (3) is relaxed from equality to an inequality (this corresponds to the fact that postcritically finite maps may have critical points that are not eventually fixed).

Definition 13.4.2 (Abstract Newton Graph) Let Γ be a connected finite graph in $\mathbb{S}^2$ with vertex set $V(\Gamma)$ and $f : \Gamma \to \Gamma$ a graph map. The pair (Γ, f) is called an *abstract Newton graph of level* $\mathcal{N}_\Gamma$ if it satisfies the following conditions:

(1) There exists $d_\Gamma \geq 3$ and an abstract channel diagram $\Delta \subsetneq \Gamma$ of degree d_Γ such that f fixes each vertex and each edge of Δ (pointwise).

(2) If $v_\infty, v_1, \ldots, v_{d_\Gamma}$ are the vertices of Δ, then $v_i \in \overline{\Gamma \setminus \Delta}$ if and only if $i \neq \infty$. Moreover, there are exactly $\deg_{v_i}(f) - 1 \geq 1$ edges in Δ that connect v_i to v_∞ for $i \neq \infty$.
(3) $\sum_{x \in V(\Gamma)} \big(\deg_x f - 1\big) \leq 2d_\Gamma - 2$.
(4) $\mathcal{N}_\Gamma$ is the minimal integer so that $f^{\mathcal{N}_\Gamma - 1}(v) \in \Delta$ for all $v \in V(\Gamma)$ with $\deg_v f > 1$.
(5) $f^{\mathcal{N}_\Gamma}(\Gamma) \subset \Delta$.
(6) For every $v \in V(\Gamma)$ with $f^{\mathcal{N}_\Gamma - 1}(v) \in \Delta$, the number of adjacent edges in Γ equals $\deg_v f$ times the number of edges adjacent to $f(v)$.
(7) The graph $\overline{\Gamma \setminus \Delta}$ is connected.
(8) For every vertex $y \in V(\Gamma)$ and every component U of $\mathbb{S}^2 \setminus \Gamma$, the local extension $\tilde{f}$ from Eq. (13.1) is injective on $\bigcup_{v \in f^{-1}(y)} U_v \cap U$.

Next we define abstract Newton rays. Let Γ be a finite connected graph embedded in $\mathbb{S}^2$ and $f : \Gamma \to \Gamma$ a weak graph map so that after f is promoted to a graph map in the sense of Remark 13.3.8, it can be extended to a branched cover $\overline{f} : \mathbb{S}^2 \to \mathbb{S}^2$.

Definition 13.4.3 (Abstract Newton Ray) Let $\mathcal{R}$ be an arc in $\mathbb{S}^2$ whose endpoints are denoted $i(\mathcal{R})$ and $t(\mathcal{R})$. Then $\mathcal{R}$ is called an *abstract Newton ray with respect to* (Γ, f) if $\mathcal{R} \cap \Gamma = \{i(\mathcal{R})\}$ and $\overline{f}(\mathcal{R}) = \mathcal{R} \cup \mathcal{E}$, where $\mathcal{E}$ is a (possibly empty) subgraph of Γ. Such an abstract Newton ray is called a *periodic abstract Newton ray with respect to* (Γ, f) if moreover there is a minimal positive integer m so that $\overline{f}^m(\mathcal{R}) = \mathcal{R} \cup \mathcal{E}$, where $\mathcal{E}$ is a (possibly empty) subgraph of Γ. We say that the integer m is *the period* of $\mathcal{R}$, and that $\mathcal{R}$ *lands* at $t(\mathcal{R})$.

Definition 13.4.4 (Preperiodic Abstract Newton Ray) An abstract Newton ray $\mathcal{R}'$ is called a *preperiodic abstract Newton ray with respect to* (Γ, f) if the following hold:

- there is a minimal integer $l > 0$ such that $\overline{f}^l(\mathcal{R}') = \mathcal{R} \cup \mathcal{E}$, where $\mathcal{E}$ is a (possibly empty) subgraph of Γ and $\mathcal{R}$ is a periodic abstract Newton ray with respect to (Γ, f).
- $\mathcal{R}'$ is not a periodic abstract Newton ray with respect to (Γ, f).

We say that the integer l is *the preperiod* of $\mathcal{R}'$, and that $\mathcal{R}'$ *lands* at $t(\mathcal{R}')$.

Now we are ready to introduce the concept of an abstract extended Newton graph. Later we prove that this graph carries enough information to characterize postcritically finite Newton maps.

Definition 13.4.5 (Abstract Extended Newton Graph) Let $\Sigma \subset \mathbb{S}^2$ be a finite connected graph, and let $f : \Sigma \to \Sigma$ be a weak graph map. A pair (Σ, f) is called an *abstract extended Newton graph* if the following are satisfied:

(1) (Abstract Newton graph) There exists a positive integer $\mathcal{N}$ and an abstract Newton graph Γ at level $\mathcal{N}$ so that $\Gamma \subseteq \Sigma$. Furthermore $\mathcal{N}$ is minimal so that condition (4) holds.

(2) (Periodic Hubbard trees) There is a finite collection of (possibly degenerate) minimal abstract extended Hubbard trees $H_i \subset \Sigma$ which are disjoint from Γ, and for each H_i there is a minimal positive integer $m_i \geq 2$ called the *period of the tree* such that $f^{m_i}(H_i) = H_i$.
(3) (Preperiodic trees) There is a finite collection of possibly degenerate trees $H_i' \subset \Sigma$ of preperiod ℓ_i, i.e. there is a minimal positive integer ℓ_i so that $f^{\ell_i}(H_i')$ is a periodic Hubbard tree (H_i' is not necessarily a Hubbard tree). Furthermore for each i, the tree H_i' contains a critical or postcritical point.
(4) (Trees separated) Any two different periodic or pre-periodic Hubbard trees lie in different complementary components of Γ.
(5) (Periodic Newton rays) For every periodic abstract extended Hubbard tree H_i of period m_i, the graph Σ contains exactly one periodic abstract Newton ray $\mathcal{R}_i$ with respect to (Γ, f). The ray lands at a repelling fixed point $\omega_i \in H_i$ of f^{m_i} and has period m_i.
(6) (Preperiodic Newton rays) For every preperiodic tree H_i', there exists at least one preperiodic abstract Newton ray in Σ with respect to (Γ, f) connecting a vertex of H_i' to Γ.
(7) (Unique extendability) For every vertex $y \in V(\Sigma)$ and every component U of $\mathbb{S}^2 \setminus \Sigma$, the local extension $\tilde{f}$ from Eq. (13.1) is injective on $\bigcup_{v \in f^{-1}(y)} U_v \cap U$.
(8) (Topological admissibility) $\sum_{x \in V(\Sigma)} \left(\deg_x f - 1\right) = 2d_\Gamma - 2$, where d_Γ is the degree of the abstract channel diagram $\Delta \subset \Gamma$.
(9) (Edges and vertices) Every edge in Σ must be one of the following three types:

- Type N: An edge in the abstract Newton graph Γ of condition (1).
- Type H: An edge in a periodic or pre-periodic abstract Hubbard tree of condition (2) or (3).
- Type R: A periodic or pre-periodic abstract Newton ray with respect to (Γ, f) from condition (5) or (6).

As a consequence, every vertex of Σ is either a Hubbard tree vertex or a Newton graph vertex.

Remark 13.4.6 (Regular Extension) The purpose of condition (7) is that after f has been upgraded to a graph map following Remark 13.3.8, the hypothesis of Proposition 13.3.7 is met. Thus f has a regular extension $\overline{f}$ which is unique up to Thurston equivalence.

Remark 13.4.7 (Implied Auxiliary Edges) Suppose that H_i is a Hubbard tree (or Hubbard tree preimage) in some complementary component U_i of Γ with connecting Newton ray R_i. If H_i contains a critical point, the existence of a regular graph map extension from Condition (7) implies that Σ must have at least one pre-periodic Newton ray edge distinct from R_i connecting H_i to Γ. All such pre-periodic rays must map to $f(R_i)$ under f (ignoring the parts in Γ as usual), and each such ray is called an *auxiliary edge corresponding to R_i*.

Remark 13.4.8 (Consistency in [15]) It is well-known that a polynomial may have a fixed point that is the landing point of a non-fixed periodic cycle of rays. On the other hand, each polynomial has at least one fixed point that is the landing point of a fixed ray (in the quadratic case, this is the so-called β fixed point). There was some effort in [15] to allow more generally that a periodic Newton ray land at a fixed point in a Hubbard tree through a higher period access. Unfortunately this was not done consistently, and so there are some matters that we would like to clarify. First, Condition (5) of Definition 7.3 in [15] might seem to permit accesses of higher period r_i, but the uniqueness of Newton rays in a given complementary component of Γ (found in the same condition) immediately implies that $r_i = 1$. Thus, despite the superficial difference, Condition (5) in Definition 7.3 of [15] and Definition 13.4.5 above are actually equivalent. The only remaining issue is that in the proof of Theorem 6.2 in [15], one should always choose the fixed point of the first return map on the Hubbard tree to be the landing point of a fixed external ray under straightening. This ensures that the graph satisfies Condition (5).

13.5 Equivalence of Abstract Extended Newton Graphs

When the extended Newton graph was constructed for a postcritically finite Newton map in [15], the Newton graph and Hubbard tree edges were constructed intrinsically, but the construction of the Newton rays involved many choices. The endpoints and accesses of the Newton rays were chosen arbitrarily, and in the case of non-degenerate Hubbard trees, there are a countably infinite number of homotopy classes of arcs by which the tree could be connected to the Newton graph (corresponding to the fact that removing the Hubbard tree from the complementary component of the Newton graph produces a topological annulus).

Let (Σ_1, f_1) and (Σ_2, f_2) be two abstract extended Newton graphs. In this section, we define an equivalence relation for abstract extended Newton graphs so that we can tell from the combinatorics of (Σ_1, f_1) and (Σ_2, f_2) whether or not their extensions to branched covers are Thurston equivalent. In fact, the main difficulty in determining equivalence of these graphs comes from establishing the equivalence of extensions across the topological annuli just mentioned. This motivates the following notation: for an abstract Newton graph Σ, denote by Σ^- the resulting graph when all edges of type R are removed; only the type N and H edges remain. We keep the endpoints of the removed edges as vertices of Σ^-.

The combinatorial equivalence given below in Definition 13.5.13 must somehow encode the Thurston class of graph map extensions to complementary components of Σ_1^- and Σ_2^- that contain non-degenerate Hubbard trees (for other types of components it is already clear how to proceed because they are either topological disks or once-punctured disks). This is our primary focus from now until the definition is given.

Definition 13.5.1 (Newton Ray Grand Orbit) The (forward) *orbit* of a Newton ray R in an abstract extended Newton graph Σ is the set of all Newton rays R' in Σ so that $f^k(R)$ contains R' for some k. The *grand orbit* of a Newton ray R in Σ is the union of all Newton rays R' in Σ whose orbit contains an edge in the orbit of R.

Remark 13.5.2 (Some Simplifying Assumptions) To simplify notation, we assume in Sects. 13.5.1 and 13.5.2 that Σ_1 and Σ_2 are combinatorially and dynamically equal apart from their Newton rays. Specifically this means that the identity map on $\mathbb{S}^2$ induces a graph isomorphism between the Newton graphs of Σ_1 and Σ_2 (from now on denoted Γ), as well as the Hubbard trees. We also assume that $f_1 = f_2$ on all vertices of Σ_1 and Σ_2 (the restriction to vertices of either graph map will be denoted f).

In Sect. 13.5.1 we describe how to alter the endpoints and accesses of Newton ray grand orbits in Σ_1 so that they coincide with those of Σ_2 without changing the homotopy class of the graph map extension. Once this is done, a method is given in Sect. 13.5.2 to determine whether the rays yield equivalent extensions across complementary components of Σ_1^- and Σ_2^- that contain non-degenerate Hubbard trees; accordingly, an equivalence is placed on ray grand orbits. Finally the combinatorial equivalence of abstract extended Newton graphs is given in Sect. 13.5.3 in terms of the equivalence on ray grand orbits.

13.5.1 Making Ray Endpoints and Accesses Coincide

We present an initial alteration to the Newton ray grand orbits of Σ_1 and Σ_2 so that their endpoints and accesses to the Newton graph and Hubbard tree vertices coincide (it is possible and not infrequent for a repelling fixed point of a polynomial to be the landing point of multiple external rays, and we simply wish to fix a preferred external ray, corresponding to an access to the fixed point in the complement of the filled Julia set). These alterations are done so as to not change the homotopy classes of the graph map extensions, so after the alterations the different graph maps are easier to compare because they only differ in the homotopy classes of rays connecting Hubbard trees to the Newton graph with corresponding endpoints, and so can be distinguished using an integer condition. Even though the operation is performed on ray grand orbits, it is somewhat non-dynamical in nature because a periodic Newton ray may be replaced with a Newton ray that is not periodic.

Lemma 13.5.3 *Suppose that $\Sigma_1^- = \Sigma_2^-$ and that $f_1|_{\Sigma_1^-} = f_2|_{\Sigma_2^-}$. Let H_m be a Hubbard tree of period $m \geq 2$ in both Σ_1 and Σ_2, where H_m is contained in the complementary component U_m of Γ. Let $R_{1,m} \subset \Sigma_1$ be a Newton ray landing at $\omega_1 \in H_m$ and $R_{2,m} \subset \Sigma_2$ a Newton ray landing at $\omega_2 \in H_m$. Then there is a Newton ray $R'_{2,m}$ landing at H_m whose grand orbit under the extension $\overline{f_2}$ has the same endpoints and accesses as the ray grand orbit of $R_{1,m}$ (while the homotopy class of $R'_{2,m}$ will usually be different from that of $R_{q,m}$).*

Proof Let H_{m+1} be the Hubbard tree that contains $f(H_m)$, and let V_m be the complementary component of $f(\Gamma)$ containing H_{m+1}. Note that upon restricting the domain,

$$\overline{f}_2 : U_m \setminus \overline{f}_2^{-1}(H_{m+1}) \to V_m \setminus H_{m+1}$$

defines a covering map between two topological annuli. Under this map, the image of $R_{1,m}$ is a curve in V_m that is not necessarily simple, but using annular coordinates this image curve is homotopic in $V_m \setminus H_{m+1}$ rel endpoints to a simple curve $\rho'_{2,m+1}$. Let $R'_{2,m}$ be the lift of $\rho'_{2,m+1}$ under $\overline{f}_2$ that is homotopic to $R_{1,m}$, having the same accesses and endpoints (such a lift exists by the homotopy lifting property rel the vertex set). Clearly $\overline{f}_2(R'_{2,m})$ is a simple arc in $V_m \setminus H_{m+1}$. We have thus shown that $R'_{2,m}$ is a ray that has the correct endpoints and accesses.

Next we produce the ray grand orbit of $R'_{2,m}$ by a lifting procedure. Let H_{m-1} be a Hubbard tree so that $f_2(H_{m-1})$ contains H_m, and let $R_{1,m-1}$ be the ray in the grand orbit of $R_{1,m}$ that lands at H_{m-1}. Define U_{m-1} to be the complementary component of Γ containing H_{m-1}, and let $V_{m-1} := \overline{f}_2(U_{m-1})$. Let z_m be the endpoint of $f_2(R_{1,m-1})$ that is in ∂V_{m-1}. Note that the portion of the path $f_2(R_{1,m-1})$ that is contained in Γ defines a path connecting z_m to the endpoint of $R'_{2,m}$ that lies in the Newton graph. The concatenation of this path with $R'_{2,m}$ yields a path in $\overline{V_{m-1}}$ that connects H_m to z_m. There is a lift of this concatenation that has the same endpoints and accesses as $R_{1,m-1}$, and we denote this lift by $R'_{2,m-1}$.

Continuing this lifting procedure inductively produces the desired ray grand orbit. □

Corollary 13.5.4 *Suppose that* $\Sigma_1^- = \Sigma_2^-$ *and that* $f_1|_{\Sigma_1^-} = f_2|_{\Sigma_2^-}$. *There is a graph* Σ'_2 *that is the domain of a continuous map* f'_2 *so that*

- Σ_2^- *is a subgraph of* Σ'_2 *and* $f_2|_{\Sigma_2^-} = f'_2|_{\Sigma_2^-}$,
- *the only edges of* Σ'_2 *that are not in* Σ_2^- *are Newton ray edges,*
- *the endpoints and accesses of the Newton rays in* Σ_1 *coincide with those of the rays in* Σ'_2, *and*
- f'_2 *is the restriction of* $\overline{f}_2$ *to* Σ'_2.

Proof Apply the preceding lemma to each Newton graph ray orbit in Σ_1. □

Remark 13.5.5 Having changed the Newton rays, a corresponding change is made to the auxiliary edges. The auxiliary edges associated with $R'_{2,i}$ are taken to be all the Newton rays in $\overline{f}_2^{-1}(\overline{f}_2(R_{2,i}))$ that are contained in U_i. It is easily seen that Σ'_2 in the Corollary satisfies all of the properties of an abstract extended Newton graph except that the Newton rays are not necessarily periodic or preperiodic. The loss of periodicity is actually not significant since we only need to understand certain topological properties of the extension described next.

13.5.2 Equivalence on Newton Ray Grand Orbits

We now wish to compare the extensions of graph maps over complementary components of Σ_1^- and Σ_2^- containing nondegenerate Hubbard trees in their closures. The other complementary components may only be disks or once-punctured disks; these are not discussed here because there is a unique extension over such components up to isotopy given by the Alexander trick [9, Chapter 2.2].

Restricting attention to the complementary components of Γ that contain the grand orbit of a single Hubbard tree, we will show in Lemma 13.5.8 that whether or not two extensions are equivalent (in the sense of Definition 13.5.6) can be determined solely in terms of numerical properties of the Newton rays. We can then define a combinatorial equivalence on Newton ray grand orbits so that two ray grand orbits are Thurston equivalent if and only if the extensions to the complementary components of Γ intersecting the grand orbit are equivalent (see Lemma 13.5.11).

Let H_1 be a nondegenerate Hubbard tree in Σ_1 (and Σ_2) of preperiod $r \geq 0$ and period $m \geq 2$, and let $H_i = f^{i-1}(H_1)$ for $1 \leq i \leq r+m$. Let U_i be the complementary component of Γ that contains H_i and let $\mathscr{U} = \cup_i U_i$. Each $\overline{f_1}(U_i)$ is homeomorphic to a disk, and is a complementary component of $f(\Gamma)$ that contains both H_{i+1} and U_{i+1} as a proper subset. So strictly speaking, the restriction of $\overline{f_1}$ and $\overline{f_2}$ to $\mathscr{U}$ do not define a dynamical system on $\mathscr{U}$. To remedy this, fix a homeomorphism $\overline{f_1}(\mathscr{U}) \to \mathscr{U}$ that restricts to the identity on each Hubbard tree. Postcomposing each of $\overline{f_1}$ and $\overline{f_2}$ by this homeomorphism produces two maps which we denote $\overline{f_1}, \overline{f_2} : \mathscr{U} \to \mathscr{U}$ by a slight abuse of notation.

Definition 13.5.6 (Thurston Equivalent Graph Extensions Over $\mathscr{U}$) We say that two extensions $\overline{f_1}, \overline{f_2} : \mathscr{U} \to \mathscr{U}$ of the graph maps f_1, f_2 are *Thurston equivalent over $\mathscr{U}$* if there are homeomorphisms $\phi_1, \phi_2 : \overline{\mathscr{U}} \to \overline{\mathscr{U}}$ so that:

$$\phi_1 \circ \overline{f_1} = \overline{f_2} \circ \phi_2$$

and there exists a homotopy $\Phi : [0, 1] \times \overline{\mathscr{U}} \to \overline{\mathscr{U}}$ with $\Phi(0, \cdot) = \phi_1$, $\Phi(1, \cdot) = \phi_2$ so that for all $t \in [0, 1]$, we have that $\Phi(t, \cdot)|_{\partial\mathscr{U}} \subset \partial\mathscr{U}$ and $\Phi(t, \cdot)$ restricts to the identity on graph vertices for all t. If ϕ_1 and ϕ_2 are homotopic to the identity in the sense just mentioned, we say that the extensions are homotopic over $\mathscr{U}$.

Each Hubbard tree H_i is contained in a complementary component U_i of Γ so that $U_i \setminus H_i$ is a non-degenerate topological annulus. Let T_i denote the right-hand Dehn twist about this annulus. The T_i operate on pairwise disjoint annuli because no two H_i lie in the same complementary component of the Newton graph. Thus any two such twists commute.

Let $R_{1,i}, R_{2,i}$ for $1 \leq i \leq r+m$ denote the Newton ray edges connecting H_i to Γ in Σ_1, Σ_2 respectively. By Corollary 13.5.4, we may assume $R_{1,i}$ and $R_{2,i}$ have the same endpoints and accesses. The equality symbol is used for arcs to indicate that they are homotopic rel endpoints on $\mathscr{U}$. Note that for all i, there is a unique $\ell_i \in \mathbb{Z}$ and $\ell_i' \in \mathbb{Z}$ so that $T_i^{\ell_i}(R_{1,i}) = R_{2,i}$ and $T_{i+1}^{\ell_i'}(\overline{f_1}(R_{1,i})) = \overline{f_2}(R_{2,i})$.

Remark 13.5.7 It is very likely that the numerical condition in the following lemma can be simplified. For example, in Proposition 2.2 of [2], the Thurston class of a polynomial mating postcomposed by a Dehn twist about the equator is expressed in terms of the same mating with one of the polynomials rotated. A similar sort of behavior is expected when applying a Dehn twist about a filled Julia set for a Newton map. Though less conceptual, the following lemma can be proven quickly.

Lemma 13.5.8 (Numerics of Equivalent Extensions) *The extensions* $\overline{f_1}$ *and* $\overline{f_2}$ *over* $\mathscr{U}$ *of the graph maps* f_1 *and* f_2 *are Thurston equivalent if and only if there are integers* $n_1, \ldots, n_{r+m-1}$ *that satisfy the following system of linear equations:*

$$d_i(n_i - \ell_i) + \ell_i' = n_{i+1} \tag{13.2}$$

where $1 \leq i \leq r+m-1$ *and* $n_r = n_{r+m}$.

Proof Suppose that the extensions $\overline{f_1}$ and $\overline{f_2}$ are Thurston equivalent. Then there are $n_i \in \mathbb{Z}$ so that up to branched cover homotopy,

$$S \circ \overline{f_1} = \overline{f_2} \circ S \tag{13.3}$$

where $S = T_1^{n_1} \circ \ldots \circ T_{m+r-1}^{n_{m+r-1}}$.

Fix i as in the statement of the lemma. All of the Dehn twists $T_1, \ldots, T_{m+r-1}$ fix $\overline{f_1}(R_{1,i})$ except possibly T_{i+1}, and thus the expression on the left side of Eq. (13.3) acts on the ray $R_{1,i}$ as follows:

$$S \circ \overline{f_1}(R_{1,i}) = T_{i+1}^{n_{i+1}} \circ \overline{f_1}(R_{1,i}) \tag{13.4}$$

and the right side acts on $R_{1,i}$ as follows:

$$\begin{aligned}
\overline{f_2} \circ S(R_{1,i}) &= \overline{f_2}(T_i^{n_i}(R_{1,i})) \\
&= \overline{f_2}(T_i^{n_i - \ell_i}(R_{2,i})) \\
&= T_{i+1}^{d_i(n_i - \ell_i)}(\overline{f_2}(R_{2,i})) \\
&= T_{i+1}^{d_i(n_i - \ell_i) + \ell_i'}(\overline{f_1}(R_{1,i})).
\end{aligned}$$

Equating the expression in the previous line and the right side of Eq. (13.4) we obtain Eq. (13.2).

If on the other hand Eq. (13.2) has integer solutions, it follows that $S \circ \overline{f_1}$ and $\overline{f_2} \circ S$ are homotopic over $\mathscr{U}$ using a close analog of Lemma 13.3.6. Thus the extensions $\overline{f_1}$ and $\overline{f_2}$ are Thurston equivalent over $\mathscr{U}$. □

Remark 13.5.9 Since the grand orbit of a Hubbard tree can be written as the union of the forward orbit of finitely many Hubbard trees, Definition 13.5.6 extends to the case of complementary components of the Newton graph containing the grand orbit of a Hubbard tree. Similar numerics as in the previous lemma hold for this slightly more general case.

Definition 13.5.10 (Newton Ray Grand Orbit Equivalence) We say that the grand orbit of the Newton ray $R_{1,i} \subset \Sigma_1$ landing at H_i is equivalent to the grand orbit of the ray $R_{2,i} \subset \Sigma_2$ landing at H_i if after applying Corollary 13.5.4 to guarantee that $R_{1,i}$ and $R_{2,i}$ have the same endpoints and accesses, the numerical condition of Lemma 13.5.8 is satisfied.

Lemma 13.5.11 *Let (Σ_1, f_1) and (Σ_2, f_2) be two abstract extended Newton graphs with $\Sigma_1^- = \Sigma_2^-$ and $f_1 = f_2$. Then if the corresponding Newton ray orbits are equivalent under Definition 13.5.10, the extensions of f_1 and f_2 to the sphere are Thurston equivalent as marked covers.*

Proof Applying Lemma 13.5.4 to $f_2 : \Sigma_2 \to \Sigma_2$, we may assume that the Newton ray grand orbit of f_2 has the same endpoints and accesses as Σ_1. Let $\mathscr{U}$ be the union of the complementary components of Σ_2^- that contain a non-degenerate Hubbard tree in the closure. Then define the homeomorphisms $\phi_1, \phi_2 : \mathbb{S}^2 \to \mathbb{S}^2$ to be the identity on $\mathbb{S}^2 \setminus \mathscr{U}$ and on the components of $\mathscr{U}$ (which are necessarily annuli), the maps are defined to be the self-homeomorphism of $\mathscr{U}$ from Lemma 13.5.8. Then Lemma 13.5.8 and Lemma 13.3.6 imply that f_1 and f_2 are Thurston equivalent as marked covers. □

13.5.3 *Equivalence on Abstract Extended Newton Graphs*

We now define the combinatorial equivalence relation on abstract extended Newton graphs that is used in the classification theorem (Theorem A) and prove an important result connecting this equivalence with Thurston equivalence. Note that the simplifying assumptions of Remark 13.5.2 are in effect for Sect. 13.5.3.

Lemma 13.5.12 (Extension Across Newton Rays) *Let (Σ_1, f_1) and (Σ_2, f_2) be abstract extended Newton graphs. Let $\phi_1^-, \phi_2^- : \Sigma_1^- \to \Sigma_2^-$ be graph homeomorphisms that preserve the cyclic order of edges at all the vertices of Σ_1^-, Σ_2^-, and satisfy the equation $\phi_1^- \circ f_1 = f_2 \circ \phi_2^-$ on Σ_1^-. Then if the accesses/endpoints of each Newton ray in Σ_1 correspond under ϕ_1^- and ϕ_2^- to the accesses/endpoints of a Newton ray in Σ_2, then ϕ_1^-, ϕ_2^- extend to graph homeomorphisms $\phi_1, \phi_2 : \Sigma_1 \to \Sigma_2$ that preserve the cyclic order of edges at each vertex, with $\phi_1 \circ f_1 = f_2 \circ \phi_2$ on Σ_1.*

Proof For a given Newton ray R in Σ, the graph maps ϕ_1 and ϕ_2 are already defined at the endpoints $i(R)$ and $t(R)$ of R. Thus the image of the single edge R under ϕ_1 must be taken to be the unique Newton ray in Σ_2 connecting $\phi_1(t(R))$ and $\phi_1(i(R))$, and likewise for ϕ_2. The extension of the conjugacy across R is accomplished by pullback. □

Definition 13.5.13 (Equivalence Relation for Abstract Extended Newton Graphs) Let (Σ_1, f_1) and (Σ_2, f_2) be abstract extended Newton graphs. We say that (Σ_1, f_1) and (Σ_2, f_2) are *equivalent* if and only if

- there exist two homeomorphisms $\phi_1^-, \phi_2^- : \Sigma_1^- \to \Sigma_2^-$ that are graph maps and preserve the cyclic order of edges at all the vertices of Σ_1^-, Σ_2^-,
- the equation $\phi_1^- \circ f_1 = f_2 \circ \phi_2^-$ holds on Σ_1^-, and
- upon applying Corollary 13.5.4 and Lemma 13.5.12 to produce the extensions $\phi_1, \phi_2 : \Sigma_1 \to \Sigma_2$, each ray grand orbit of f_2 is equivalent (see Definition 13.5.10) to a ray grand orbit of $\phi_1 \circ f_1 \circ \phi_2^{-1}$ and vice versa.

We now show that two abstract extended Newton graphs are combinatorially equivalent if and only if their extensions are Thurston equivalent.

Theorem 13.5.14 (Combinatorial Formulation of Thurston Equivalence) *Let* (Σ_1, f_1) *and* (Σ_2, f_2) *be abstract extended Newton graphs with graph map extensions* $\overline{f_1}, \overline{f_2} : \mathbb{S}^2 \to \mathbb{S}^2$ *respectively. Then* (Σ_1, f_1) *and* (Σ_2, f_2) *are equivalent in the sense of Definition 13.5.13 if and only if* $(\overline{f_1}, \Sigma_1')$ *and* $(\overline{f_2}, \Sigma_2')$ *are Thurston equivalent as marked branched covers.*

Proof First assume that the two abstract extended Newton graphs are equivalent. By definition of extended Newton graph equivalence, there are graph homeomorphisms $\phi_1, \phi_2 : \Sigma_1 \to \Sigma_2$ that satisfy the conditions in Definition 13.5.13. Since the complementary components of Σ_1 and Σ_2 are all disks, ϕ_1, ϕ_2 can be extended to global homeomorphisms $\bar{\phi_1}, \bar{\phi_2}$. Since the Newton ray grand orbits are equivalent in the sense of Definition 13.5.10, Lemma 13.5.11 implies that there must be homeomorphisms S_1 and S_2 of the sphere which are both products of Dehn twists about the non-degenerate Hubbard trees as in Lemma 13.5.8 so that

$$S_1 \circ \bar{\phi}_1 \circ f_1 = f_2 \circ \bar{\phi}_2 \circ S_2 \tag{13.5}$$

where S_1 is homotopic to S_2 relative to the vertices of Σ_1. The maps on both sides of Eq. (13.5) are both regular extensions of a graph map (see Proposition 13.3.7) and they also satisfy the hypotheses of Lemma 13.3.6. Thus f_1 and f_2 are Thurston equivalent as marked branched covers.

Now suppose that $(\overline{f}_1, \Sigma_1')$ and $(\overline{f}_2, \Sigma_2')$ are Thurston equivalent as marked branched covers. Take $g_0, g_1 : (\mathbb{S}^2, \Sigma_1') \to (\mathbb{S}^2, \Sigma_2')$ to be the maps from the definition of Thurston equivalence where $g_0 \circ \overline{f_1} = \overline{f_2} \circ g_1$. Let e be an edge of Σ_1^- with endpoints ∂e. Then $g_1(e)$ connects the two points in $g_1(\partial e)$. Moreover, g_1 preserves the cyclic order at each vertex of Σ_1^-, because it is an orientation-preserving homeomorphism of $\mathbb{S}^2$. Let $g' : (\mathbb{S}^2, \Sigma_2') \to (\mathbb{S}^2, \Sigma_2')$ be a homeomorphism that for all edges e not Newton rays maps each $g_1(e)$ to an edge of Σ_2 that connects the two points in $g_1(\partial e)$.

Then $g' \circ g_1$ realizes an equivalence between the two abstract extended Newton graphs (let $\phi_0 = \phi_1 = g' \circ g_1$ in Definition 13.5.13), except that the Newton rays must still be shown to be equivalent. Apply Corollary 13.5.4 so that all Newton rays have corresponding endpoints under ϕ_0, ϕ_1. Then since $\overline{f_1}$ and $\overline{f_2}$ are Thurston equivalent as branched covers, Lemma 13.5.8 implies the rays are equivalent. Thus (Σ_1, f_1) and (Σ_2, f_2) are combinatorially equivalent. □

13.6 Newton Maps from Abstract Extended Newton Graphs

We now prove that all abstract extended Newton graphs are realized by Newton maps.

Proof of Theorem B It suffices to show that the marked branched cover $(\overline{f}, \Sigma')$ is unobstructed, where Σ' denotes the set of vertices of Σ. The conclusion will follow by Head's theorem, where the holomorphic fixed point theorem is used to argue that the point at infinity is repelling [20].

Suppose to the contrary that Π is a multicurve obstruction for $(\overline{f}, \Sigma')$, and without loss of generality assume Π is irreducible. Recall from Condition (1) of Definition 13.4.5 that Σ contains an abstract Newton graph Γ which in turn contains an abstract channel diagram Δ. The following lemma restricts where obstructions may exist, using Theorem 13.2.5.

Lemma 13.6.1 *If Π is a multicurve obstruction for $(\overline{f}, \Sigma')$, then*

$$\Pi \cdot (\Gamma \setminus \Delta) = 0.$$

Proof Suppose first that there exists an edge λ in Δ so that $\lambda \cdot \Pi \neq 0$. Since $\{\lambda\}$ itself forms an irreducible arc system, the second case of Theorem 13.2.5 implies that Π intersects no other preimage of λ except for λ itself.

If on the other hand, $\lambda \cdot \Pi = 0$, the first case of Theorem 13.2.5 implies that no preimage of λ intersects Π. Since every edge in Γ is a lift of an edge in Δ, the conclusion follows. □

The proof of the theorem is now completed by showing that whether or not Π has intersection with Δ in minimal position, a contradiction results.

13.6.1 Contradiction for the Case $\Pi \cdot \Delta \neq 0$

Let γ_1 be any curve in Π so that $\gamma_1 \cdot \Delta \neq 0$. Recall from Definition 13.4.2 that $\overline{\Gamma \setminus \Delta}$ is connected. It is a consequence of Lemma 13.6.1 that $\overline{\Gamma \setminus \Delta}$ does not intersect Π; thus $\overline{\Gamma \setminus \Delta}$ must be a subset of one of the complementary components of γ_1. Denote the complementary component of γ_1 that does not contain $\overline{\Gamma \setminus \Delta}$ by $D(\gamma_1)$. None of the vertices of Γ except possibly v_∞ lie in $D(\gamma_1)$. However, there must be at least two vertices of Σ in $D(\gamma_1)$ for otherwise γ_1 would not be essential. The only

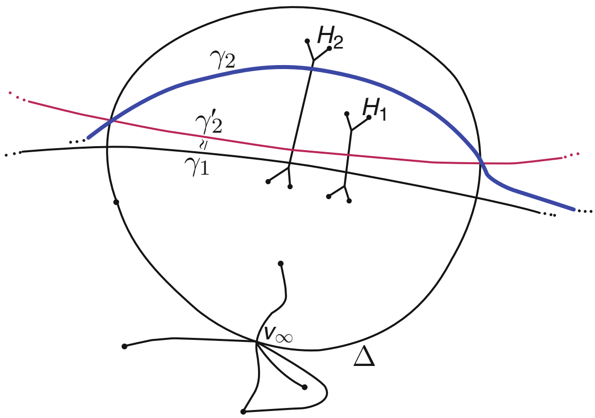


Fig. 13.4 Illustration of the first case $\Pi \cdot \Delta \neq 0$

vertices of Σ which could possibly be in $D(\gamma_1)$ are v_∞ and Hubbard tree vertices. Due to the connectedness of Σ, at least one of the following must hold: $\gamma_1 \cdot H_1 \neq 0$ for some Hubbard tree H_1 or $\gamma_1 \cdot R_1 \neq 0$ for some Newton ray R_1. We only prove the Hubbard tree case, noting that the Newton ray case is identical.

First suppose $\gamma_1 \cdot H_1 \neq 0$ (see Fig. 13.4). Let $\gamma_2 \in \Pi$ be some curve whose preimage under $\overline{f}$ has a component γ_2' which is homotopic to γ_1 rel vertices (γ_2 exists by irreducibility). Clearly γ_2' must intersect H_1 and Δ. Let λ_1 be some component of $\gamma_2' \cap (S^2 \setminus \Delta)$ that intersects H_1. Recall that H_1 must either be non periodic or have period at least two, so in either case $H_2 := \overline{f}(H_1)$ does not intersect H_1. Also H_1 and H_2 must be in the same complementary component of Δ because λ_1 connects H_1 to Δ (without passing through any other edges of Γ due to Lemma 13.6.1), and $\overline{f}$ is an orientation-preserving map that fixes each edge of Δ.

Now we show that H_1 and H_2 can be connected by some path that avoids Γ except at its endpoints. Since all critical points of $\overline{f}$ are contained in Σ (by Condition (8) of Definition 13.4.5) the Riemann–Hurwitz formula implies that $\overline{f}^{-1}(\Delta) \subset \Gamma$. The edges of Δ are invariant under $\overline{f}$, so the endpoints of $\lambda_2 := \overline{f}(\lambda_1)$ are in the same edges as the endpoints of λ_1. Since $\overline{f}^{-1}(\Delta) \subset \Gamma$, we have that λ_2 intersects Δ only at its endpoints. Starting at an intersection of H_1 and λ_1, traverse λ_1 until right before the intersection with the edge of Δ. Traverse a path in a small neighborhood of this edge until λ_2 is reached without intersecting any edges of Γ. Traverse λ_2 until H_2 is reached. This completes the construction of a path $\lambda_{1,2}$ from H_1 to H_2 that does not intersect Δ. Moreover, Lemma 13.6.1 implies that λ_1 and λ_2 do not intersect $\Gamma \setminus \Delta$ avoiding Γ and so $\lambda_{1,2}$ does not intersect Γ. This contradicts the assumption that H_1 and H_2 are separated by the Newton graph (Condition (4) of Definition 13.4.5).

13.6.2 *Contradiction for the Case* $\mathbf{\Pi \cdot \Delta = 0}$

Using Lemma 13.6.1 we see that $\Pi \cdot \Gamma = 0$. Recall the assumption that every complementary component of Γ contains at most one abstract extended Hubbard tree (Condition (4) of Definition 13.4.5).

Suppose that U is such a complementary component containing some $\gamma \in \Pi$. The only postcritical points that could possibly be contained in U are vertices of Hubbard trees, so U contains one Hubbard tree or one Hubbard tree preimage. Since Π is irreducible, the Hubbard tree must in fact be periodic and since γ is essential the Hubbard tree is non-degenerate. Thus U contains exactly one non-degenerate periodic abstract Hubbard tree H of some period m. Define $F := \overline{f}^m$, and note that Π is also a multicurve obstruction for F. Extract an irreducible multicurve obstruction for F from Π, which we again denote by Π, and assume that U still contains some component of Π.

We show that the two Thurston linear maps F_Π and $(F|_U)_\Pi$ are equal. In fact, we show $\Pi \subset U$. Suppose that W is a complementary component of Γ different from U, and $\gamma' \subset W$ for some $\gamma' \in \Pi$. By the irreducibility of Π, there is some $n > 0$ and a component γ'' of $F^{-n}(\gamma')$ that is homotopic to γ rel vertices. Note that $\gamma'' \subset U$ and that its complementary component that is a subset of U contains some vertices of Σ which must in fact be vertices of H. Since γ'' is homotopic to a subset of each arbitrarily small neighborhood of H, we obtain a contradiction since $F^n(\gamma'') \subset W$ but $F^n(H) = H \subset U$. Thus the two Thurston linear maps F_Π and $(F|_U)_\Pi$ are equal.

This contradicts the realizability (or unobstructedness) of the abstract Hubbard tree H [22, Theorem II.4.7], and thus no such obstruction Π exists, completing the proof.

13.7 Proof of the Classification Theorem

Theorem 1.2 of [15] asserts that every postcritically finite Newton map has an extended Newton graph that satisfies the axioms of Definition 13.4.5, and we have shown in Sect. 13.6 that every abstract extended Newton graph extends to an unobstructed branched cover, and is therefore realized by a Newton map. We now check that these two assignments are well-defined on equivalence relations and are inverses of each other, giving an explicit classification of postcritically finite Newton maps in terms of combinatorics.

Recall that **Newt** is the set of postcritically finite Newton maps up to affine conjugacy, and that **NGraph** is the set of abstract extended Newton graphs up to Thurston equivalence (Definition 13.5.13). Equivalence classes in both cases are denoted by square brackets. Our first goal is to show that the assignments made in Theorems 13.1.3 and B are well-defined on the level of equivalence classes, namely, they induce mappings $\mathcal{F}$: **Newt** $\to$ **NGraph** and $\mathcal{F}'$: **NGraph** $\to$ **Newt**.

We now argue that $\mathcal{F}$ is well-defined. The construction from [15] of the extended Newton graph for a fixed Newton map involved no choices in the construction of type H and N edges, and possibly many choices in the construction of type R edges. Let $(\Delta^*_{\mathcal{N},1}, N_p)$ and $(\Delta^*_{\mathcal{N},2}, N_p)$ be two extended Newton graphs constructed for N_p. Proposition 6.4 in [15] asserts that $\Delta^-_{\mathcal{N},1} = \Delta^-_{\mathcal{N},2}$ and $N_p|_{\Delta^-_{\mathcal{N},1}} = N_p|_{\Delta^-_{\mathcal{N},2}}$ (recall that $\Delta^-_{\mathcal{N},1}$ denotes the graph $\Delta_{\mathcal{N},1}$ with all Newton ray edges removed). We thus only need to show that the Newton ray grand orbits are equivalent. The branched cover $(N_p, (\Delta^*_{\mathcal{N},1})')$ is identical as a branched cover to $(N_p, (\Delta^*_{\mathcal{N},2})')$ and they are both extensions of graph maps $N_p|_{\Delta^*_{\mathcal{N},1}}$ and $N_p|_{\Delta^*_{\mathcal{N},2}}$ respectively. Theorem 13.5.14 then implies equivalence for corresponding ray grand orbits.

Well-definedness of $\mathcal{F}'$ is immediate from the fact that equivalent graphs have Thurston equivalent extensions (Theorem 13.5.14) which correspond to affine conjugate Newton maps by Thurston rigidity (Theorem 13.2.3).

Proof of Theorem A We first show injectivity of $\mathcal{F} : \textbf{Newt} \to \textbf{NGraph}$. Let N_{p_1} and N_{p_2} be two postcritically finite Newton maps that have equivalent extended Newton graphs $\Delta^*_{\mathcal{N},1}$ and $\Delta^*_{\mathcal{N},2}$. Theorem 13.1.3 asserts that each of these graphs satisfies the axioms of an abstract extended Newton graph, and since both graphs are equivalent, the marked branched covers $(N_{p_1}, (\Delta^*_{\mathcal{N},1})')$ and $(N_{p_2}, (\Delta^*_{\mathcal{N},2})')$ are equivalent by Theorem 13.5.14. We may then conclude that N_{p_1} and N_{p_2} are affine conjugate using Thurston rigidity.

Next we show injectivity of $\mathcal{F}' : \textbf{NGraph} \to \textbf{Newt}$. Suppose that a postcritically finite Newton map N_p realizes two abstract extended Newton graphs (Σ_1, f_1) and (Σ_2, f_2). By minimality of the extended Hubbard trees and the Newton graph, we know that $\Sigma_1' = \Sigma_2'$. Then the marked branched covers (N_p, Σ_1') and (N_p, Σ_2') are Thurston equivalent. By Theorem 13.5.14 we conclude that (Σ_1, f_1) and (Σ_2, f_2) are combinatorially equivalent.

Finally we prove that $\mathcal{F}$ and $\mathcal{F}'$ are bijective and inverses of each other. Let $(\Sigma, f) \in \textbf{NGraph}$ be an abstract extended Newton graph. It follows from Theorem B that (Σ, f) is realized by a postcritically finite Newton map N_p. Thus

$$\mathcal{F}'([(\Sigma, f)]) = [N_p].$$

Denote by $\Delta^*_{\mathcal{N}}$ an extended Newton graph associated with N_p which is guaranteed by Theorem 13.1.3 so that

$$\mathcal{F}([N_p]) = \left[\left(\Delta^*_{\mathcal{N}}, N_p\right)\right].$$

The injectivity statement just proved implies that under the equivalence of Definition 13.5.13,

$$[(\Sigma, f)] = \left[\left(\Delta^*_{\mathcal{N}}, N_p\right)\right].$$

Thus $\mathcal{F} \circ \mathcal{F}'$ is the identity, and consequently the mapping $\mathcal{F}$: **Newt** → **NGraph** is bijective and $\mathcal{F}' \circ \mathcal{F}$ is the identity. □

This completes the combinatorial classification of postcritically finite Newton maps.

Acknowledgements This research was partially supported by the Deutsche Forschungsgemeinschaft (DFG), as well as the Advanced Grant 695621 HOLOGRAM of the European Research Council (ERC), which is gratefully acknowledged. The authors are most grateful to the anonymous referee for very helpful comments that have led to marked improvements.

References

1. B. Bielefeld, Y. Fisher, J. Hubbard, The classification of critically preperiodic polynomials as dynamical systems. J. Am. Math. Soc. **5**, 721–762 (1992)
2. X. Buff, A. Epstein, S. Koch, Twisted matings and equipotential gluings. Ann. Fac. Sci. Toulouse Math. **6**(21), 995–1031 (2012)
3. X. Buff, G. Cui, L. Tan, Teichmüller spaces and holomorphic dynamics, in *Handbook of Teichmüller Theory*, ed. by A. Papadopoulos, vol. IV (European Mathematical Society, 2014), pp. 717–756
4. K. Cordwell, S. Gilbertson, N. Nuechterlein, K. Pilgrim, S. Pinella, On the classification of critically fixed rational maps. Conform. Geom. Dyn. **19**, 51–94 (2015)
5. A. Douady, J. Hubbard, A proof of Thurston's topological characterization of rational functions. Acta Math. **171**, 263–297 (1993)
6. K. Drach, D. Schleicher, Rigidity of Newton dynamics (2018). arXiv:1812.11919
7. K. Drach, Y. Mikulich, J. Rückert, D. Schleicher, A combinatorial classification of postcritically fixed Newton maps. Ergodic Theory Dyn. Syst. **39**(11), 2983–3014 (2019)
8. K. Drach, R. Lodge, D. Schleicher, M. Sowinski, Puzzles and the Fatou-Shishikura injection for rational Newton maps. Trans. Am. Math. Soc. **374**, 2753–2784 (2021)
9. B. Farb, D. Margalit, *A Primer on Mapping Class Groups*. Princeton Mathematical Series, vol. 49 (Princeton University Press, Princeton, 2011)
10. L. Geyer, Classification of critically fixed anti-rational maps (2020). arXiv:2006.10788
11. J. Head, The combinatorics of Newton's method for cubic polynomials. Ph.D. Thesis, Cornell University (1987)
12. M. Hlushchanka, Tischler graphs of critically fixed rational maps and their applications (2019). arXiv:1904.04759v1
13. J. Hubbard, D. Schleicher, S. Sutherland, How to find all roots of complex polynomials by Newton's method. Invent. Math. **146**, 1–33 (2001)
14. R. Lodge, Y. Luo, S. Mukherjee, Circle packings, kissing reflection groups and critically fixed anti-rational maps. Forum Math. Sigma (2020). arXiv:2007.03558
15. R. Lodge, Y. Mikulich, D. Schleicher, Combinatorial properties of Newton maps. Indiana Univ. Math. J. **70**(5), 1833–1867 (2021)
16. J. Luo, Combinatorics and holomorphic dynamics: captures, matings, Newton's method. Ph.D. Thesis, Cornell University (1995)
17. K. Mamayusupov, A characterization of postcritically minimal Newton maps of complex exponential functions. Ergodic Theory Dyn. Syst. **39**(10), 2855–2880 (2019)
18. C. McMullen, *Complex Dynamics and Renormalization*. Annals of Mathematics Studies, vol. 135 (Princeton University Press, Princeton, 1994)
19. Y. Mikulich, A classification of postcritically finite Newton maps. Ph.D. Thesis, Jacobs University Bremen (2011)

20. J. Milnor, *Dynamics in One Complex Variable*. Annals of Mathematics Studies, vol. 160 (Princeton University Press, Princeton, 2006)
21. K. Pilgrim, L. Tan, Combining rational maps and controlling obstructions. Ergodic Theory Dyn. Syst. **18**, 221–245 (1998)
22. A. Poirier, Hubbard trees. Fundam. Math. **208**(3), 193–248 (2010)
23. M. Randig, D. Schleicher, R. Stoll, Newton's method in practice II: the iterated refinement Newton method and near-optimal complexity for finding all roots of some polynomials of very large degrees (2017). arXiv:1703.05847
24. P. Roesch, Y. Yin, J. Zeng, Rigidity of non-renormalizable Newton maps (2018). arXiv:1811.09978
25. J. Rückert, Newton's method as a dynamical system. Ph.D. Thesis, International University Bremen (2006)
26. D. Schleicher, On the efficient global dynamics of Newton's method for complex polynomials (2011). arXiv:1108.5773
27. D. Schleicher, R. Stoll, Newton's method in practice: finding all roots of polynomials of degree one million efficiently. Theor. Comput. Sci. **681**, 146–166 (2017)
28. S. Shemyakov, R. Chernov, D. Rumiantsau, D. Schleicher, S. Schmitt, A. Shemyakov, Finding polynomial roots by dynamical systems — a case study. Discrete Contin. Dyn. Syst. **40**(12), 6945–6965 (2020)
29. L. Tan, Branched coverings and cubic Newton maps. Fundam. Math. **154**, 207–259 (1997)
30. X. Wang, Y. Yin, J. Zeng, Dynamics of Newton maps (2018). arXiv:1805.11478

Chapter 14
The Development of the Theory of Automatic Groups

Sarah Rees

Abstract We describe the development of the theory of automatic groups. We begin with a historical introduction, define the concepts of automatic, biautomatic and combable groups, derive basic properties, then explain how hyperbolic groups and the groups of compact 3-manifolds based on six of Thurston's eight geometries can be proved automatic. We describe software developed in Warwick to compute automatic structures, as well as the development of practical algorithms that use those structures. We explain how actions of groups on spaces displaying various notions of negative curvature can be used to prove automaticity or biautomaticity, and show how these results have been used to derive these properties for groups in some infinite families (braid groups, mapping class groups, families of Artin groups, and Coxeter groups). Throughout the text we flag up open problems as well as problems that remained open for some time but have now been resolved.

Keywords Automatic group · Hyperbolic group · Finite state automaton · Combing · 3-manifold group · Decision problem · Word problem · Conjugacy problem · Artin group · Coxeter group · Mapping class group

AMS Subject Classifications 20F10, 20F36, 20F55, 20F65, 20F67, 57M60; Secondary Classification: 03D10, 68Q04

14.1 Introduction

This chapter describes the development of the theory of automatic groups. It aims to explain the definition, and put that into mathematical and historical context, to detail what is known, give brief accounts of some of the big problems in the subject that have already been solved, and describe those problems that remain open.

S. Rees (✉)
Newcastle, Tyne and Wear, UK
e-mail: Sarah.Rees@newcastle.ac.uk

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_14

Thurston is credited with the definition of automatic groups, and is one of six authors of one of the primary early references of the subject [29]; but some of the foundations were laid in particular in work of Gromov on hyperbolic groups [35], Cannon on properties of the fundamental groups of compact hyperbolic manifolds [18], Gilman on groups with rational cross-sections [34]. The standard reference is certainly the book [29], but that is supplemented by some powerful results in [7, 32, 33], while Farb's article [30] gives a useful and readable overview of early development of the subject.

The definition of an automatic group was originally designed to identify properties of a group that were observed in the fundamental groups of compact hyperbolic 3-manifolds, and which facilitated computation with those groups. Such groups are finitely generated. When a group is automatic, its associated automatic structure allows the elements of the group to be represented as strings belonging to a particularly well structured set of strings, for which certain computations can be easily performed using finite state automata, as we shall see below.

Within this introductory section, we shall give some historical background, then define the notation and terminology that we shall need in the remainder of this chapter. Section 14.2 contains the definition of an automatic group, identifies the basic properties, and describes the most natural examples, and non-examples. Section 14.3 describes computation with automatic groups, how automatic structures may be computed, how they may, and have been, used. Section 14.4 describes how automaticity or biautomaticity of a group may be deduced from the geometry of a space on which the group has a good action. Section 14.5 describes the derivation of results proving automaticity or biautomaticity of groups in some well known families of group, which often used techniques or results described in Sect. 14.4. Finally Sect. 14.6 describes some problems that remain open.

14.1.1 Historical Background

Alongside Thurston, it is natural to indentify Cannon, Epstein and Holt as the key figures in the early development of automatic groups. Much of the information in this section comes from discussion with these three people [19], or can be found in the preface of the standard reference [29].

Cannon's article [19] identifies the International Congress of Mathematicians in Helsinki in 1978 as a location at which key ideas that influenced the development of the concept of an automatic group were discussed.

In his plenary address, Thurston discussed the construction of geometric structures on a 3-manifold M, and the tesselation of its universal cover $\tilde{M}$ by a structure dual to the Cayley graph of $\pi_1(M)$. Thurston's geometrisation conjecture [70], subsequently proved by Perelman, claimed that every closed 3-manifold was geometrisable, that is, admitted a canonical decomposition into pieces each admitting one of eight types of geometric structure.

In his article [19], Cannon attributes to Thurston at that conference the conjecture that the growth series of a group G acting discretely, cocompactly and isometrically on a finite dimensional hyperbolic space $\mathbb{H}_n$ should be a rational function. Cannon proved that conjecture in [18], where he identified features of $\mathbb{H}_n$ within the Cayley graph $\mathsf{Cay}(G, X)$ for G with respect to a finite generating set X. In particular, he proved that $\mathsf{Cay}(G, X)$ admits finitely many types of "cones" on geodesics, and deduced from this the rationality of the growth function of G. Cannon also proved that the word and conjugacy problems for G could be solved using analogues of Dehn's algorithms for those in hyperbolic surface groups . Gromov's 1987 article [35] defined a combinatorial notion of hyperbolicity for a graph, and hence for a group (via its Cayley graph), and generalised Cannon's results to groups satisfying this definition of hyperbolicity. There is a substantial body of material studying (Gromov) hyperbolic groups, in particular [2].

Thurston realised that the finiteness of the set of cone types in one of Cannon's groups of hyperbolic isometries allowed the construction of a finite state automaton recognising the set of geodesic words within the group; rationality of the growth function is an immediate consequence of that set of words being the language of a finite state automaton. "Fellow travelling" properties of quasi-geodesic paths in $\mathbb{H}_n$ that had been recognised by Cannon allowed the construction of further automata that recognised right multiplication in the goup by a generator.

Now Thurston defined the concept of an automatic group. He called a group with finite generating set X *automatic* if it possessed a representative set of words L over X, such that one finite state automaton recognised the words in L, and other automata recognised pairs of words in L related in the group under right multiplication by the generators in X. Very early on, according to Holt, groups of this type were known as *regular groups*. But this terminology conflicted with other uses of the term regular, and so was soon changed.

Initially, in particular in [7, 29], the study of the family of automatic groups was largely driven by the desire to find within it the groups of the geometrisable 3-manifolds, and hence to harness computational techniques that were provided by the association of automatic groups with regular languages. Epstein realised very early on that any automatic group must be finitely presented, while Thurston deduced that any such group had quadratic Dehn function and hence word problem soluble in quadratic time. Epstein and Holt in Warwick worked, together with the author of this chapter, to develop practical procedures to (attempt to) build automatic structures for finitely presented groups, and to compute within the groups using those structures.

14.1.2 Mathematical Background and Notation

All the groups that we consider will be finitely generated. If X is a finite generating set for a group G, then we write $G = \langle X \rangle$. In that case every element of G can be represented as a product (or string) of elements of X and their inverses. We denote by X^{-1} the set of symbols x^{-1} for which $x \in X$, and then by $X^{\pm}$ the disjoint union of X and X^{-1}; every non-identity element of G can now be described as a string of elements of $X^{\pm}$. The identity element, which we denote by 1, can be described as a product of length 0.

Given a finite set A, we define a *string* w over A to be a sequence $a_1a_2\cdots a_n$ with $a_i \in A$, and call n the *length* of w, denoted by $|w|$; we may alternatively use the term *word* over A rather than string. A subsequence $a_ia_{i+1}\cdots a_j$ of w is called a *substring* or *subword*. We write $w(i)$ for the *prefix* $a_1\cdots a_i$ of w. We call the string or word of length 0 over A the *empty string* or *empty word* and denote that by ε. As is standard, we denote by A^+ the set of all strings over A of finite length > 0 and by A^* the union $A^+ \cup \{\varepsilon\}$. Given an ordering of the elements of A, we define the shortlex ordering on A^* as follows: for words $u = x_1\dots x_r$ and $v = y_1\dots y_s$, we define $v <_{\text{slex}} u$ if $|v| < |u|$, or if $|u| = |v|$ and for some i, $y_1 = x_1, \dots, y_{i-1} = x_{i-1}$ but $y_i < x_i$.

When X is a generating set for a group G, and $w \in (X^{\pm})^*$, it is often convenient to abuse notation and use w to indicate not only that string over $X^{\pm}$ but also the group element that the string represents; if $w, v \in X^{\pm}$, we write $w = v$ to denote that w, v are identical as strings, and $w =_G v$ to denote that w, v represent the same group element. If $g \in G$, we denote by $|g|$ the length of the shortest word over $X^{\pm}$ that represents g. Suppose that $\mathsf{Cay} = \mathsf{Cay}(G, X)$ is the *Cayley graph* of G over X, that is the graph with vertex set G and, for each $g \in G$, $x \in X$, directed edges labelled x and x^{-1} connecting the ordered pairs of vertices (g, gx) and (gx, g). Then for each $g \in G$, a path labelled by w joins the vertex g of Cay to the vertex gw; we shall represent that path as ${}_gw$.

When G is finitely generated by X, we define a *language* for G over X to be a subset of $(X^{\pm})^*$ that contains at least one representative of each element of G, that is, that maps onto G under the map assigning each product over X to the element it represents.

For the free group F_n on a set X of n generators $x_1, \dots, x_n$, a language is provided by the set of all *freely reduced* words of length ≥ 0 over $X^{\pm}$, that is, the set of all words within which no subword $x_ix_i^{-1}$ or $x_i^{-1}x_i$ appears. For the free abelian group $\mathbb{Z}^n$ on the same set of n generators, a language is provided by the set of all words of the form $x_{i_1}^{r_1}x_{i_2}^{r_2}\cdots x_{i_k}^{r_k}$, with $k \geq 0$, $i_1 < i_2 < \cdots < i_k$ and $r_i \in \mathbb{Z}\setminus\{0\}$. In each of these two examples the language provides a unique representative for each group element.

Each of the two languages just described is an example of a *regular language*, that is, it is the set $L(M)$ of strings accepted by a *finite state automaton* (fsa) M with alphabet $\{x_1^{\pm1}, x_2^{\pm1} \dots, x_n^{\pm1}\}$. Finite state automata provide standard models of bounded memory computation and are defined and studied in [44]. It is common to represent a finite state automaton M with alphabet A as a finite directed graph, with each directed edge labelled by one or more elements of A, one vertex identified as the *start*, and a subset of the vertices selected as *accepting*. A word w is then accepted by M if it labels at least one directed path from the start to an accepting vertex; if there is no such path, or if the end point (*target*) of every such path is a non-accepting vertex then w is not accepted. It is standard to call the vertices of M its

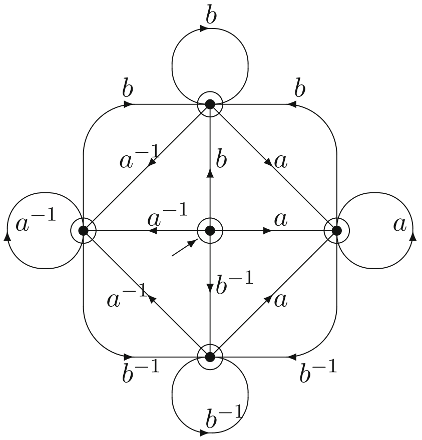


Fig. 14.1 fsa M_1, M_2 giving the languages described in the text for F^2 and $\mathbb{Z}^2$

states, the directed edges its *transitions* and the set of accepted words its *language* $L(M)$. In the cases where $n = 2$, the languages described above for the free and free abelian groups over $\{a, b\}$ are accepted by the two finite state automata shown in Fig. 14.1; in each diagram, following convention, the start state is indicated by an arrow, and the accepting states are ringed. In each of the two examples, each of the five states shown in the diagram is accepting, but a further *failure state* is not shown, which constitutes a sixth state; the failure state is non-accepting, any transitions not shown in the diagram are assumed to be to that failure state, and all transitions from the failure state are to the failure state.

14.2 Automatic Groups

14.2.1 Definition of an Automatic Group

Now suppose that G is a group with finite generating set X. For $k \in \mathbb{N}$, words w, v over $X^{\pm}$ are said to *k-fellow travel* in G if for each $i \leq \max\{|w|, |v|\}$ the distance between the vertices $w(i)$ and $v(i)$ of $\mathsf{Cay} = \mathsf{Cay}(G, X)$ (using the graph metric) is at most k. Equivalently, we say that the paths ${}_1w$ and ${}_1v$ of Cay k-fellow travel. A group G with finite generating set X is defined to be *automatic* over X if

A1 there is a language L for G over X that is regular,
A2 there is an integer k such that, for each $y \in X \cup \{1\}$, and for any $w, v \in L$ with $wy =_G v$, the paths ${}_1w$, ${}_1v$ k-fellow travel in Cay.

We call L the *language*, the fsa accepting L the *word acceptor* and k the *fellow traveller constant* of an *automatic structure* for G.

The fsa M_1 illustrated in Fig. 14.1 is the word acceptor of an automatic structure with fellow traveller constant 1 for F_2 over $\{a, b\}$; each element of the group has a unique representative in the language, and given two words $w, v \in L(M_1)$ and $y \in \{a^{\pm 1}, b^{\pm}\}$ with $wy =_{F_2} v$, one of the words is a maximal prefix of the other, and so the words 1-fellow travel in G.

Similarly, the fsa M_2 of Fig. 14.1 is the word acceptor of an automatic structure with fellow traveller constant 2 for $\mathbb{Z}^2$ over $\{a, b\}$. Again each element of the group has a unique representative in the language, and given two words $w, v \in L(M_2)$ and $y \in \{a^{\pm 1}, b^{\pm 1}\}$ with $wy =_{\mathbb{Z}^2} v$, corresponding vertices on the paths ${}_1w$ and ${}_1v$ in $\mathsf{Cay}(\mathbb{Z}^2, \{a, b\})$ are joined in the graph by a path of length 1 or 2. The language $L(M_2)$ is the set of all SHORTLEX minimal geodesic representatives of group elements; we call this a *shortlex automatic structure* for $\mathbb{Z}^2$. Note that we can define a similar shortlex automatic structure for $\mathbb{Z}^n$.

In the definition of automaticity given in [29] the condition A2 given above is replaced by the following condition:

A2' For each $y \in X \cup \{\{1\}\}$, the set of pairs (w, v) for which $w, v \in L$ and $wy =_G v$ is a regular language when viewed as a set of strings over the alphabet of pairs $\{(a, b) : a, b \in X^{\pm} \cup \{\$\}\}$; the character \$ is a *padding symbol* used to deal with the situation where $|w| \neq |v|$, in which case the shorter of the two words is padded with \$s at its end.

The automata recognising the regular languages just described are known as the *multiplier automata* of the automatic structure, usually denoted by M_y, for each choice of y.

In the presence of A1 the conditions A2 and A2' are equivalent. This is a consequence of the fact that the k-fellow travelling of a pair of words w, v can be tracked by an automaton whose state set $\mathcal{D}$ corresponds to a set of words of length at most k; a pair of words (w, v) is accepted by that automaton so long as all the products $w'^{-1}v'$ associated with prefixes $w' := w(i), v' := v(i)$ of w, v are represented by words in $\mathcal{D}$. We call such an automaton a *word difference machine*, and the associated set $\mathcal{D}$ its corresponding set of *word differences*.

Where G is automatic over its finite generating set X, with automatic structure L, k, then G is said to be *biautomatic* (and (L, k) to be a *biautomatic structure* for G) if the additional condition A3 is satisfied:

A3 for each $y \in X$, and for any $w, v \in L$ with $yw =_G v$, the paths ${}_yw, {}_1v$ k-fellow travel in $\mathsf{Cay}(G, X)$.

This further fellow traveller condition can be expressed in terms of fsa that recognise left multiplication, usually denoted by ${}_yM$, for $y \in X$. It is an open question whether all automatic groups are biautomatic.

The concept of automaticity can be generalised to one of *asynchronous automaticity* by replacing the fellow traveller condition by an *asynchronous fellow travel condition*; for two words w, v to asynchronously fellow travel within a group G it is the distance between vertices $w(j_i)$ and $v(k_i)$ that must be bounded, where, for some $m \geq \max(|w|, |v|)$, the sequences $(j_0, j_1, \ldots, j_m)$ and $(k_0, \ldots, k_m)$ are both increasing sequences of integers, with $j_0 = k_0 = 0$, $j_m = |w|$, $k_m = |v|$, and for $0 \leq l < m$, $j_{l+1} - j_l$ and $k_{l+1} - k_l$ are in $\{0, 1\}$. Asynchronous automaticity is certainly a more general concept than automaticity, and it is satisfied by examples such as the Baumslag–Solitar groups which are certainly not automatic.

It is fairly standard to call a language L for a group G that satisfies the condition A2 (but not necessarily A1) a *combing* for G, and a language that satisfies both A2 and A3 a *bicombing* for G; however some authors use these terms differently, e.g. impose additional (geometric) conditions on L. Again, the fellow travelling condition can be replaced by an asynchronous one, in order to define asynchronous combings and bicombings. The basic properties of combable groups are studied in [12], where it is proved that non-automatic combable groups exist (answering a question posed in [29]), as well as combable groups that are not bicombable.

Given an automatic (or biautomatic) structure (L, k) for a group G, it is straightforward (using well known properties of regular languages, such as the "Pumping lemma" [44]) to modify the structure and achieve a new automatic structure with particular properties. For instance we can achieve a structure in which every element of G has a unique representative (a structure *with uniqueness*) a *prefix closed* structure in which the language contains every prefix of every one of its elements, a *quasigeodesic* structure in which every element is represented by a (λ, ϵ)-quasigeodesic, We note that a word w representing an element g of a group G is called a (λ, ϵ)-*quasigeodesic* if every subword w' of w has length at most $\lambda|g'| + \epsilon$, where g' is the element represented by w'. Note that it is not clear that all combinations of properties can be achieved within the language of a single automatic structure. In particular it is an open question [29] whether, given an automatic structure for a group G, an automatic structure can be derived for G that is both prefix closed and has uniqueness.

Note that the definitions of automaticity and biautomaticity are independent of choice of generating set; that is if G has an automatic structure over a finite generating set X, then it has one over any other finite generating set Y.

14.2.2 Basic Properties of Automatic Groups

Some properties of automatic groups can be deduced very easily from basic properties of regular languages, which imply certain constraints on their Cayley graphs. In particular any automatic group is finitely presented with soluble word problem, and quadratic Dehn function, while any biautomatic group has soluble conjugacy problem. We recall that the word problem is soluble in G if an algorithm exists that can decide whether or not any input word represents the identity, and the conjugacy problem is soluble if an algorithm exists that can decide whether or not two input words represent elements that are conjugate within the group; it is an open question whether the conjugacy problem is soluble for automatic groups. It also is an open question whether the isomorphism problem is soluble for automatic groups, that is, whether an algorithm that was given as input automatic structures for a pair of groups G, H could decide whether or not G and H were isomorphic. It is conjectured in [29] that this problem is insoluble. Note that it is soluble for hyperbolic groups [24, 67].

In order to explain these statements in more detail, we use the language of van Kampen diagrams. Informally (essentially, following [51]), given a group G with presentation $\langle X \mid R \rangle$ and a word w over X that represents the identity of G, we define a van Kampen diagram Δ_w for w to be a finite, connected, directed, planar graph, with a selected *basepoint*, whose directed edges are labelled by elements of X, in such a way that the boundary of every face of the graph (known as a *cell*) is labelled (from some starting point, in some orientation) by a word from R, while the boundary of the graph is labelled (from the basepoint) by w. As a directed, edge labelled graph, Δ_w maps (not necessarily injectively) into the Cayley graph $\mathsf{Cay}(G, X)$. The *area of the diagram* $\mathsf{Area}(\Delta_w)$ is defined to be the number of cells it contains; of course its value is dependent on the set R, and would change if R were changed.

We define the *area of the word* w to be the minimum of the areas of all van Kampen diagrams that represent w. And we define the *Dehn function* (or *isoperimetric function*) for G, $f\colon \mathbb{N} \to \mathbb{N}$, to be the function for which $f(n)$ is the maximum area of all words w of length n over $X^{\pm}$ that represent the identity of G. Although the precise form of the Dehn function depends on the chosen presentation for G, it can be shown that two Dehn functions corresponding to different presentatives are related by a natural notion of equivalence, and in particular if one is polynomially bounded, then both are, by polynomials of the same degree.

Proposition 14.2.1 *Every automatic group is finitely presented, with a quadratic upper bound on the Dehn function, and hence soluble word problem.*

We sketch the proof, which is that of [43, Theorem 5.2.13].

Proof We suppose that L, k are the language and fellow traveller constant of an automatic structure over a generating set X; we may assume that L consists of quasigeodesics. Suppose that $w = a_1 \cdots a_n$ is a word of length n representing the identity. Now we define words $w_0, \ldots, w_n$ as follows. We define $w_0 = w_n$ to be a representative in L of 1, and for each $i = 1, \ldots, n-1$ we choose w_i to be a representative in L of the prefix of w of length i; since L is quasigeodesic, we can choose w_i of length at most $|w_0| + Ci$, for some constant C of the automatic structure. We start with a disk within the plane whose boundary is labelled by w, and divide it into cells to form a van Kampen diagram Δ_w with boundary w as follows. First, a loop labelled by w_0 connects the basepoint to itself, while for each i a path labelled w_i connects the basepoint to the point on the boundary distance i along w, and none of these paths cross each other. Then, since the paths ${}_1w_{i-1}$ and ${}_1w_i$ in $\mathsf{Cay}(G, X)$ fellow travel at distance at most k, we can construct paths of length at most k that connect corresponding vertices on the paths within the disk labelled by those two words, and hence divide the region between the two paths into cells each of length at most $2k + 2$. In this way we divide the interior of the diagram into a

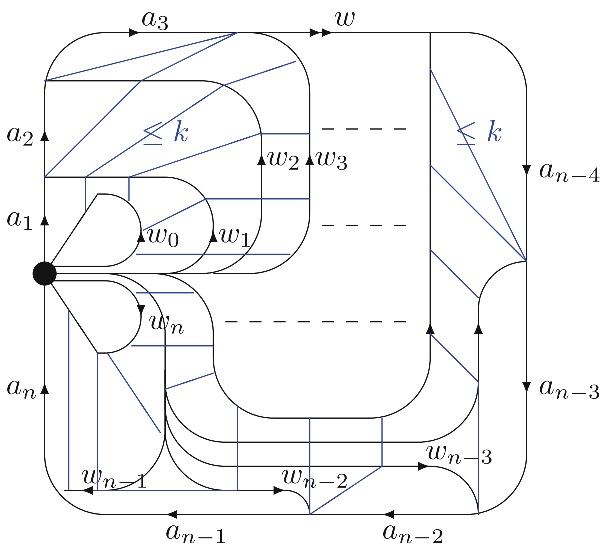


Fig. 14.2 Van Kampen diagram for a representative of the identity in an automatic group

number of cells labelled by words of length at most $2k + 2$. together with two cells labelled by the word $w_0 = w_n$, as illustrated in Fig. 14.2.

Using the bounds on $|w_i|$, we see that the total number of cells is bounded by a quadratic function of n. We now define R to be the set of all words of length up to $2k+2$ that represent the identity, together with the word w_0. Then $\langle X \mid R\rangle$ is a finite presentation for G, and, relative to R, Δ_w has quadratic area. □

A similar argument proves an exponential upper bound on the Dehn function for any asynchronously automatic group; it is an open question [29] whether a polynomial time solution to the word problem must exist.

The most straightforward way to prove a group non-automatic is probably to show that it has a Dehn function that is above quadratic. This argument proves easily the non-automaticity of the Baumslag–Solitar groups $\langle a, b \mid ba^pb^{-1} = a^q$ for which $p, q > 0$ and $p \neq q$, since they have exponential Dehn function; in fact they provide examples of non-automatic groups that are asynchronously automatic.

But there are many groups with quadratic Dehn functions that are known by other methods not to be automatic.

The non-automaticity of the groups $SL_n(\mathbb{Z})$ for $n \geq 3$ is proved in [29]. The group $SL_2(\mathbb{Z})$ is well known to be virtually free, and hence hyperbolic with linear Dehn function. The group $SL_3(\mathbb{Z})$ has exponential Dehn function and so is certainly non-automatic. However, the existence of a quadratic Dehn function for $SL_n(\mathbb{Z})$ with $n \geq 5$ was proved in [75] in 2013 (and had been conjectured by Thurston, in fact for $n \geq 4$). In order to prove non-automaticity of the group for all $n \geq 3$, Epstein and Thurston derived higher dimensional isoperimetric inequalities that would have to hold in any combable group of isometries acting properly discontinuously with compact quotient on a k-connected Riemannian manifold [29, Theorem 10.3.5]. The non-automaticity of $\mathrm{SL}_n(\mathbb{Z})$ now follows by the construction of a proper discontinous cocompact action on a suitable contractible manifold, and the demonstration that a higher dimensional isoperimetric inequality fails; hence $\mathrm{SL}_n(\mathbb{Z})$ is proved to be non-combable and so non-automatic.

Van Kampen diagrams can also be used to prove solubility of the conjugacy problem in any biautomatic group, by demonstrating the existence of a conjugator of bounded length. The proof below, valid for any bicombable group, is taken from [68]; an earlier result of [33] constructs an automaton out of the biautomatic structure to solve the problem.

Proposition 14.2.2 *Given a biautomatic group G, any two words u, v representing conjugate elements are conjugate by an element of length at most $a^{|u|+|v|}$, for some constant a (depending only on the biautomatic structure). Hence any biautomatic group has soluble conjugacy problem.*

Proof We choose a biautomatic structure (L, k) over a finite generating set X, and suppose that the words u, v over $X^{\pm}$ represent conjugate elements of G. Let $N := |X^{\pm}|^{k(|u|+|v|)}$. We find a conjugator of length at most N, and so $a = |X^{\pm}|^k$.

For suppose that an element $g \in G$ conjugates u to v, that is that $gu =_G vg$, and that $w, w' \in L$ represent the elements g and ug, respectively. We consider the paths ${}_1w$, ${}_1w'$ and ${}_uw$ within the Cayley graph $\mathsf{Cay}(G, X)$, and see that the biautomaticity of G ensures that ${}_1w$ and ${}_1w'$ fellow travel at distance at most $|u|k$, and that ${}_1w'$ and ${}_uw$ fellow travel at distance at most $|v|k$. We deduce that we can construct a van Kampen diagram with boundary labelled by $wuw^{-1}v^{-1}$ in which chords of length at most $(|u| + |v|)k$ join boundary vertices in corresponding positions on the two boundary subwords labelled by w, as shown on the left hand side of Fig. 14.3. Where $|w| = n$, let $d_1, d_2, \ldots, d_{n-1}$ be the words that label those chords.

Now if $n > |X^{\pm}|^{(|u|+|v|)k}$, then for some i, j we have $d_i = d_j$. In that case, where $\hat{w}$ is the word formed from w by deleting its middle section of length $j - i$, from its $(i + 1)$-th to its j-th letter, we can form the van Kampen diagram with boundary word $\hat{w}u\hat{w}^{-1}v^{-1}$ shown on the right hand side of Fig. 14.3 by deleting the central part of the diagram we already constructed for $wuw^{-1}v^{-1}$. □

Various combinations of automatic groups are known to be automatic [7, 29]: these include free products, direct products, certain amalgamated products and HNN extensions of automatic groups, as well as subgroups of finite index in automatic groups, groups with automatic groups as subgroups of finite index, quotients of automatic groups by finite normal subgroups. Some, but not all, of these closure properties also hold for biautomatic groups. It is an open question whether direct factors of automatic groups must be automatic (but the analogous result is proved for biautomatic groups [58]). It is also open [29] whether a group with a biautomatic group as a subgroup of finite index must be biautomatic.

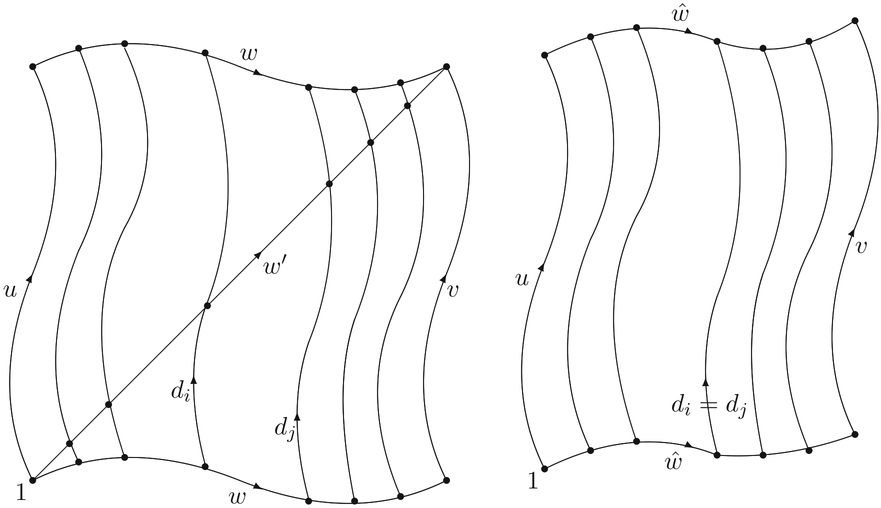


Fig. 14.3 Finding a conjugator of bounded length

14.2.3 Basic Examples and Non-examples

14.2.3.1 Virtually Abelian Groups, Soluble Groups

We already described shortlex automatic structures for the free abelian group $\mathbb{Z}^n$. In fact $\mathbb{Z}^n$ is also biautomatic, but with a different (less straightforward) language, and indeed so is every virtually abelian group. However it was already proved in [29] that an automatic nilpotent group must be virtually abelian; the proof uses the fact that a regular language with polynomial growth cannot satisfy a (synchronous) fellow traveller property. It was conjectured by Thurston that the same result must hold for an automatic soluble group. That conjecture remains open, but it was proved for automatic polycylic groups in [37], using an embedding of a finite index subgroup of a polycyclic group of exponential growth as a lattice in an appropriate Lie group, where [29, Theorem 10.3.5] about higher dimensional isoperimetric functions could be applied, which had previously been used to prove the non-automaticity of $\mathrm{SL}_n(\mathbb{Z})$ for $n \geq 3$. Much more recently it was proved in [66] that biautomatic soluble groups must be virtually abelian.

14.2.3.2 Hyperbolic Groups

Maybe the most natural examples of non-abelian automatic groups are provided by the large family of *word hyperbolic* groups, which contains all finitely generated free groups as well as the fundamental groups of all compact hyperbolic manifolds.

A group G with finite generating set X is said to be *word hyperbolic* if its Cayley graph $\mathsf{Cay}(G, X)$ is a δ-hyperbolic metric space, for some $\delta \geq 0$; a geodesic metric space $(\mathcal{X}, d)$ is δ-hyperbolic if for any triangle in $\mathcal{X}$ with geodesic sides $\gamma_1, \gamma_2, \gamma_3$ and for any vertex p on the side γ_1 there is a vertex q on the union $\gamma_2 \cup \gamma_3$ of the other two sides for which $d(p, q) < \delta$ (we say that triangles in $\mathcal{X}$ are *δ-slim*). The property of being word hyperbolic is independent of the choice of a finite generating set for G, although the value of δ is not. The fundamental groups of compact hyperbolic manifolds give examples, as do finitely generated free groups (which are 0-hyperbolic with respect to free generating sets).

We note that there are many equivalent definitions of hyperbolicity for metric spaces (and hence for finitely generated groups), which are explained in [2]. In particular there is a characterisation in terms of *thin* rather than *slim triangles* (and a linear relationship between the associated parameters "δ").

It is proved in [29] that a word hyperbolic group G is automatic over any generating set X, with an automatic structure whose language consists of all geodesic words over the selected generating set. The regularity of that set of geodesic words is equivalent to the fact that the Cayley graph $\mathsf{Cay} = \mathsf{Cay}(G, X)$ contains finitely many cone types. For $g \in G$, represented by a geodesic word w, we define the *cone* $C(g)$ (or $C(w)$) on the vertex g of Cay to be the set of (geodesic) paths γ within Cay starting at g for which the concatenation $\eta\gamma$ of a geodesic path η from 1 to g with γ is also geodesic. The *cone type* $[C(g)]$ or $[C(w)]$ of the cone is defined to be the set of words that label the paths within it. Now for $y \in X \cup X^{-1}$, if wy is also geodesic then for any word v,

$$v \in [C(wy)] \iff yv \in [C(w)].$$

It follows that we can recognise the set of geodesic words over $X^{\pm}$ with an fsa whose states correspond to the cone types, with a transition from $[C(w)]$ to $[C(wy)]$ on y whenever wy is geodesic, but otherwise to a single *failure state* (i.e. a non-accepting sink state). We can illustrate this construction in the free abelian group $\mathbb{Z}^2$ with generating set $\{a, b\}$, where there are nine cone types $[C(w)]$, defined by the nine geodesic words $\varepsilon, a, b, a^{-1}, b^{-1}, ab, ab^{-1}, a^{-1}b, a^{-1}b^{-1}$, and consisting of the nine possible sets of geodesic words in which each generator appears either only with positive exponent, or only with negative exponent, or not at all. The fsa is illustrated in Fig. 14.4. This automaton is *not* part of an automatic structure for $\mathbb{Z}^2$; it cannot be since, for example, the vertices distance i from the origin on the geodesic words $a^i b^i$ and $b^i a^i$ are distance $2i$ apart within the Cayley graph, and hence this language does not satisfy a fellow travelling property.

Given the finiteness of the set of cone types in a word hyperbolic group, biautomaticity of any word hyperbolic group now follows once it is observed that the fellow travelling of two geodesic words with common (or adjacent) start and end vertices can be derived from the slimness of triangles. In fact, it is proved by Papasoglu [64] that this fellow traveller condition characterises word hyperbolic groups, and hence so does the existence of a (bi)automatic structure that consists of all geodesic words. A procedure to test for hyperbolicity that is based on this

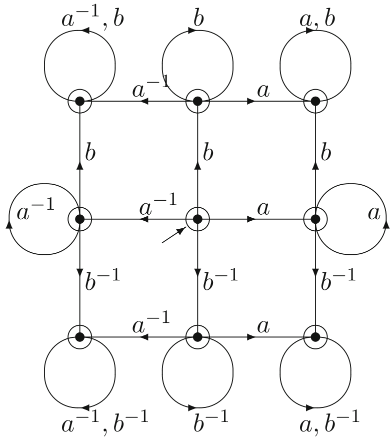


Fig. 14.4 fsa recognising geodesics in $\mathbb{Z}^2$

result is described in [73]. Starting with a shortlex automatic structure (L, k) for a group G over X, the procedure attempts to construct an automatic structure $(\widehat{L}, \widehat{k})$ with $\widehat{L} \supset L$ and $\widehat{k} \geq k$, and such that $\widehat{L}$ contains all geodesic words over $X^{\pm}$. It will terminate with such a structure precisely when G is hyperbolic. An improved procedure, based on the same result was developed by Holt and Epstein [26] and implemented in KBMAG.

The fundamental groups of finite volume hyperbolic manifolds (geometrically finite hyperbolic groups) were proved biautomatic by Epstein [29], with a further biautomatic structure subsequently described by Lang [52].

Geometrically finite hyperbolic groups were the motivating examples for Bowditch's definition [9] of a group hyperbolic relative to a collection of subgroups; a geometrically finite hyperbolic group is hyperbolic relative to a collection of abelian groups. The major part of the definition of relative hyperbolicity is the requirement that the Cayley graph of a group hyperbolic relative to a collection $\mathcal{H}$ of subgroups becomes hyperbolic after the contraction of edges within left cosets of subgroups in $\mathcal{H}$. However weaker and stronger versions of the definition exist depending on whether or not a condition of *bounded coset penetration* is required to hold. Under the stronger definition (studied in [63]) it is proved, in particular in [5], that groups hyperbolic relative to shortlex biautomatic subgroups are themselves shortlex biautomatic. The shortlex biautomaticity of geometrically finite hyperbolic groups is a consequence of this result.

A further generalisation of hyperbolic groups is provided by semihyperbolic groups, which were introduced by Bridson and Alonso in [3]; the class contains all biautomatic groups (hence all hyperbolic groups) and all CAT(0) groups (see Sect. 14.4). A group G with finite generating set X is defined to be *weakly semihyperbolic* if $\mathsf{Cay}(G, X)$ admits a bounded quasi-geodesic bicombing (with a unique combing path $s_{g_1,g_2}(t)$ identified between any pair g_1, g_2 of vertices of the graph), and *semihyperbolic* if it has such a bicombing that is equivariant under the action of G (so that $g.s_{g_1,g_2}(g) = s_{gg_1,gg_2}(t)$). This class of groups satisfies many closure properties, and all groups within it are finitely presented, with soluble word and conjugacy problems.

14.2.3.3 Fundamental Groups of Compact 3-Manifolds

It is proved in [29] that the fundamental groups of compact 3-manifolds based on six of Thurston's eight model geometries for compact 3-manifolds [71] admit automatic structures. But it is also proved that the fundamental groups of closed manifolds based on the Nil and Sol geometries (which are non-abelian, nilpotent and soluble, respectively) cannot even be asynchronously automatic [10, 29].

However, using combination theorems for automatic groups, it can be proved (as in [29, Theorem 12.4.7], but our wording is slightly different) that an orientable, connected, compact 3-manifold with incompressible toral boundary whose prime factors have JSJ decompositions containing only hyperbolic pieces has automatic fundamental group. It was proved in [13, Theorem B] that the fundamental group of a manifold as above in which manifolds based on Nil and Sol are allowed within the JSJ decomposition, while not automatic, still admits an asynchronous combing based on an indexed language [1].

14.3 Computing with Automatic Groups

14.3.1 Building Automatic Structures

The original motivation for the definition of automatic groups was computational, and so it was important from the beginning of the subject to be able to construct automatic structures, that is, given a presentation for a group G, to have a mechanism for building the word acceptor and multiplier automata of an associated automatic structure. Software to build these automata was developed at the University of Warwick, and the procedure used is described in [28]. The original programs were subsequently rewritten by Holt, and released within his KBMAG package [40], now available within both GAP and Magma computational systems [55, 69].

The basic procedure is the same in both versions (the ideas are due to Holt) and we describe it briefly now, but refer the reader to [28] or [41] for more details.

A presentation for a group G over a finite generating set X is input, together with an ordering of the set $X^{\pm}$. The procedure attempts to prove G to be shortlex automatic over X (with the given ordering) by first constructing a set of automata consisting of W and M_y for $y \in X^{\pm} \cup \{\varepsilon\}$, and then attempting to verify that those automata are indeed the automata of a shortlex automatic structure. If verification tests fail, some looping is possible within the procedure, and indeed that looping could continue indefinitely (or at least until the computer runs out of resources).

If all verification tests pass, then the procedure will have verified the shortlex automaticity of G by construction and checking of a shortlex automatic structure.

So the procedure may succeed in proving shortlex automaticity of G. But if it fails, it has certainly not proved that G is not automatic, or even that G is not shortlex automatic, but rather it suggests that G is unlikely to be shortlex automatic over the given generating set X, with the given ordering of the elements of X. We note that the question of automaticity for a finitely presented group is undecidable in general; this follows from the undecidability of questions such as triviality for a group. We note too that it is an open question [29] whether every automatic group must be shortlex automatic with respect to some ordered generating set.

The first step of the procedure to prove shortlex automaticity is the construction of a rewrite system $\mathcal{R}$ from the group presentation that is compatible with the shortlex order. By definition, $\mathcal{R}$ is a set of substitution rules $\rho : u \to v$, for $u, v \in (X^{\pm})^*$, and with $v <_{\text{slex}} u$; in order that $\mathcal{R}$ encodes the presentation we require that every relator from the group presentation is a cyclic conjugate of the product uv^{-1} or its inverse for at least one such rule.

The next step is to run the Knuth–Bendix procedure for a while on $\mathcal{R}$. The Knuth–Bendix procedure (described in [43]) is a general procedure that, given as input a rewrite system $\mathcal{R}$ for strings compatible with a partial order, modifies it by adding rules that are consequences of existing rules and deleting rules that have become redundant, in order to produce a new rewrite system. The procedure attempts to build a finite *complete* system, for which any input word w can be rewritten after a finite number of steps to a unique irreducible word w' (where irreducible means that w' cannot be rewritten further). However with this goal the procedure may never terminate; all that is guaranteed is that after bounded time the modified system must contain enough rules to reduce any word up to some bounded length to an irreducible.

In fact the procedure to construct a shortlex automatic structure for G does not need the Knuth–Bendix procedure to terminate on the input rewrite system $\mathcal{R}$. Instead, while the Knuth–Bendix procedure is running it accumulates the set $\mathcal{D}$ of *word differences* $u(i)^{-1}v(i)$ and their inverses (reduced according to the current modification of $\mathcal{R}$) that correspond to prefixes of the rules $u \to v$ in the system. Where $u = u_1 \cdots u_m$, and $v = v_1 \cdots v_{m'}$, a transition is added from each word difference $u(i)^{-1}v(i)$ to $u(i+1)^{-1}v(i+1)$, creating a *word difference machine* that can recognise fellow travelling with respect to $\mathcal{D}$.

The Knuth–Bendix procedure is paused when it seems that the set $\mathcal{D}$ and the associated automaton have stabilised. And then a candidate word acceptor WA is constructed, designed to reject a word u if a string v exists with $v <_{\text{slex}} u$ for which (u, v) fellow travels according to $\mathcal{D}$ while also the word difference $u^{-1}v$ reduces, according to the current rewrite system, to the empty word.

Similarly, multiplier automata are constructed for each $y \in X^{\pm} \cup \{\varepsilon\}$, using a direct product construction on automata to recognise pairs of words u, v for which $u, v \in L(W)$, (u, v) fellow travels according to $\mathcal{D}$, while also the word difference $u^{-1}v$ reduces, according to the current rewrite system, to y.

Now a series of elementary tests is applied to the candidate automata. If some of these tests fail, then $\mathcal{D}$ has been proved to be inadequate, and the Knuth–Bendix procedure is restarted. If and when those tests are passed, further tests known as *axiom checking* are applied, and a positive result for these tests proves the automata to provide a shortlex automatic structure for G. If the axiom checks fail then the procedure is abandoned.

14.3.2 Calculation Using the Automatic Structure

Once an automatic structure has been constructed for a group G, much can be computed using the automata of that structure. Various of these functions are available within the KBMAG package [40].

It is straightforward to enumerate the language of a finite state automaton. Hence we can enumerate a set of representative words for an automatic group, with unique representation if necessary (recall that once an automatic structure has been derived, a structure with unique representation can be derived from that).

For any regular language L the generating function $\sum_{n=0}^{\infty} s_L(n)x^n$, where $s_L(n)$ denotes the number of words of length n in L, is a rational function, and can be computed from an automaton recognising L. Hence the growth series of an automatic group is computable, given a geodesic automatic structure.

Reduction of an input word to the "normal form" defined by the language L of the automatic structure for G can be performed using a combination of the word acceptor and multiplier automata, or alternatively using the word difference machine.

Finiteness of otherwise of an automatic group is immediately recognisable from a word acceptor for an automatic structure; the language is infinite precisely when the automaton admits loops. In this way, the Heineken group $G = \langle x, y, z \mid [x, [x, y]] = z,\ [y, [y, z]] = x,\ [z, [z, x]] = y \rangle$ was proved infinite, by Holt using KBMAG; computation with the automatic structure subsequently revealed the group to be hyperbolic. Previously that group had been proposed as a possible example of a finite group with a balanced presentation. Similarly, a second proof of the infiniteness of the Fibonacci group $F(2, 9)$ was provided by the construction of an automatic structure for it [39].

Tests for hyperbolicity [26, 73] that make use of automatic structures for G together with Papasoglu's characterisation of hyperbolic groups have already been described in Sect. 14.2.3.2. The second of those is implemented in KBMAG, as is an algorithm [26] estimating the thinness constant (related to, but not equal to, the slimness constant) for geodesic triangles in the Cayley graph of a word hyperbolic group.

Quadratic and linear time solutions to the conjugacy problem in a hyperbolic group are described in [15] and [17, 27]. A practical cubic time solution that restricts to infinite order elements is due to Marshall [56], using some ideas from Swenson, and has been implemented in the GAP system.

14.4 Group Actions and Negative Curvature

One of the basic principles of geometric group theory is generally referred to as the the Milnor–Švarc lemma:

If a group G has a "nice" (properly discontinuous and cocompact) discrete, isometric action on a metric space $\mathcal{X}$ then its Cayley graph is quasi-isometric to $\mathcal{X}$. In particular a group with such an action on a δ-hyperbolic space is word hyperbolic.

A variety of results derive automaticity or biautomaticity of a group from its "nice" actions on spaces in which some kind of non-positive curvature can be found.

Theorem 14.4.1 (Gersten and Short [32, 33]) *A group acting discretely and fixed point freely on a piecewise Euclidean 2-complex of type $A_1 \times A_1$, A_2, B_2 or G_2 (corresponding to tesselations of the Euclidean plane by squares, equilateral triangles, or triangles with angles $(\pi/2, \pi/4, \pi/4)$ or $(\pi/2, \pi/3, \pi/6)$) is biautomatic.*

As a consequence of the above results, and within the same two articles, Gersten and Short deduce that groups satisfying any of the small cancellation conditions C(7) or else T(p) and T(q) with $(p, q) \in \{(3, 7), (4, 5), (5, 4)\}$ (defined in [54]) are hyperbolic, and hence in particular biautomatic, and then that groups satisfying the small cancellation conditions C(6), or C(4) and T(4), or C(3) and T(6) are biautomatic.

A geodesic metric space $\mathcal{X}$ is defined to be CAT(0) if for any geodesic triangle in the space, and for any two points p, q on the sides of that triangle, the distance between p and q in $\mathcal{X}$ is no more than the distance between the points in corresponding positions on the sides of a geodesic triangle with the same side lengths in the Euclidean plane, as illustrated in Fig. 14.5. A complete CAT(0) space is often called a *Hadamard space*. A group is called CAT(0) if it acts properly and cocompactly on a CAT(0) space.

The CAT(-1) condition is defined similarly with respect to the hyperbolic plane; any CAT(-1) space is δ-hyperbolic, for some δ, and hence CAT(-1) groups are word hyperbolic.

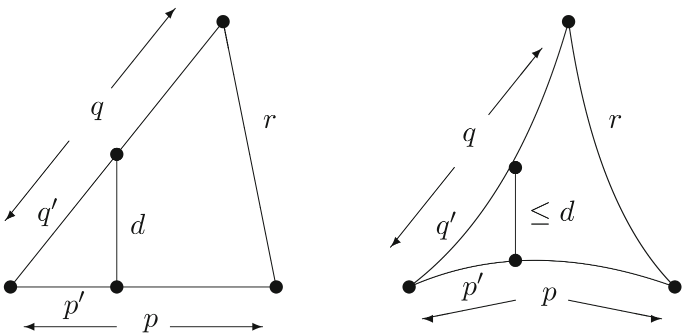


Fig. 14.5 Comparable triangles in Euclidean and CAT(0) spaces

A (not necessarily geodesic) metric space (X, d) is said to have non-positive curvature (or curvature ≤ 0) if every point of X is contained in a CAT(0) neighbourhood. By the Cartan-Hadamard theorem [14] the universal cover of a complete connected space of non-positive curvature is CAT(0).

Niblo and Reeves studied in particular groups acting on CAT(0) cube complexes:

Theorem 14.4.2 (Niblo and Reeves [61]) *A group acting faithfully, properly discontinuously and cocompactly on a simply connected and non-positively curved cube complex is biautomatic.*

A cube complex is defined to be a metric polyhedral complex in which each cell is isometric to the Euclidean cube with side lengths 1, where the gluing maps are isometries. Such a complex is non-positively curved provided that it contains at most one edge joining any two vertices, and no triangles of edges, and (by a result of Gromov [35]) is CAT(0) if non-positively curved and simply connected.

Actions of Coxeter groups on CAT(0) cube complexes are constructed in [62], but are not necessarily cocompact. However in some cases it follows from those or related constructions that the Coxeter groups are biautomatic (see Sect. 14.5.2).

There are many open problems relating to CAT(0) groups (see for example [31]). The question of whether every CAT(0) group must be biautomatic was recently resolved in the negative by Leary and Minasyan [53], who constructed an example of a 3-dimensional CAT(0) group which could admit no biautomatic subgroup of finite index. It is still unknown whether non-automatic CAT(0) groups can exist.

However a restricted class of CAT(0) groups is provided by groups that act geometrically on CAT(0) spaces with isolated flats. A k-flat in a CAT(0) space is an isometrically embedded copy of Euclidean space $\mathbb{R}^k$. This family contains a number of interesting examples, including geometrically finite Kleinian groups, the fundamental groups of various compact manifolds, and limit groups, arising from the solutions of equations over free groups. Groups of this type are studied in [46], where more details (of definition and examples) can be found. Theorem 1.2.2 of that article establishes a number of properties of such groups, including their biautomaticity.

A form of non-positive curvature in simplicial complexes is defined in [50]: a flag simplicial complex $\mathcal{X}$ is called k-systolic if connected, simply connected and locally k-large (no minimal ℓ-cycle with $3 < \ell < k$ in the link of a vertex). A group is called k-systolic if it acts simplicially, properly discontinuously and cocompactly on a k-systolic simplical complex, and is called systolic if 6-systolic.

Theorem 14.4.3 (Januszkiewicz and Swiatkowski [50]) *7-systolic groups are hyperbolic, 6-systolic groups are biautomatic.*

This result is used to prove biautomaticity of a large class of Artin groups [47], as detailed in Sect. 14.5.1.

A *Helly graph* is a graph in which every family of pairwise intersecting balls has a non-empty intersection. A group is called Helly if it acts properly and cocompactly by graph automorphisms on a Helly graph; word hyperbolic groups, CAT(0) cubical groups and C(4)-T(4) small cancellation groups are all examples. It is proved in [22] that all Helly groups are biautomatic. This result is used to prove biautomaticity of another large class of Artin groups [48], as detailed in Sect. 14.5.1.

14.5 Some Automatic and Biautomatic Families

Over a period of more than 30 years, automatic and biautomatic structures were found for various families of groups, including braid groups, many Artin groups, mapping class groups, and Coxeter groups. But some questions remain open for these families.

14.5.1 Braid Groups, Artin Groups and Mapping Class Groups

Automatic structures for the braid group B_n on n strands and also for the (closely related) mapping class group of the $(n+1)$-punctured sphere were constructed by Thurston and are described in [29]; one of the structures described for the braid groups is symmetric, proving the braid groups to be biautomatic. The automaticity (but not necessarily biautomaticity) of the mapping class group of the $n+1$-punctured sphere then follows from the fact that it contains the quotient of the braid group B_n by its centre as a subgroup of index $n+1$.

The braid group on $n+1$ strands is isomorphic to the Artin group of finite type A_n. We recall that an Artin group is a group defined by a presentation of the form

$$\langle x_1, x_2, \cdots, x_n \mid \overbrace{x_i x_j x_i \cdots}^{\mathsf{m_{ij}}} = \overbrace{x_j x_i x_j \cdots}^{\mathsf{m_{ij}}},\ i \neq j \in \{1, 2, \ldots, n\} \rangle,$$

relating to a symmetric, integer *Coxeter matrix* $(\mathsf{m_{ij}})$, or equivalently a *Coxeter diagram* Γ on n vertices, whose edge $\{i, j\}$ is labelled $\mathsf{m_{ij}}$, and is naturally associated with a Coxeter group by adding relations $x_i^2 = 1$ for each i. The Artin group has finite type if the associated Coxeter group is finite (and hence Γ is a disjoint union of diagrams from the well–known list of spherical Coxeter diagrams).

In [23], Charney used results of Deligne to extend Thurston's construction for the braid groups to all finite type Artin groups. Charney's construction provided biautomatic structures for all finite type Artin groups; these biautomatic structures were geodesic over the "Garside" generating sets, but not over the standard generators x_i. Biautomatic structures for all Garside groups (of which finite type Artin groups are examples) were described by Dehornoy [25].

For Artin groups of FC type (free products of finite type groups with amalgamation over parabolic subgroups, for which the complete subgraphs of the labelled graph formed by deleting all ∞-labelled edges from Γ are all of finite type), *asynchronously automatic* structures were constructed in [4], and used to define quadratic time solutions to the word problem; we recall that an exponential (rather then quadratic) time solution is guaranteed by asynchronous automaticity. Right-angled Artin groups (those for which all the parameters $\mathsf{m_{ij}}$ are within the set $\{2, \infty\}$, which form a subset of FC type) were then proved automatic in [38, 72]. Very recently [48] Artin groups of FC type have been proved to be Helly, and hence biautomatic.

Mosher's paper [57] answered a major open question raised by Thurston's proof of the automaticity of the mapping class group of the punctured sphere. Using quite different techniques from Thurston, Mosher proved automaticity of the mapping class group of any surface of finite type, that is, the group of (orientation preserving) homeomorphisms modulo isotopy of any surface obtained from a compact surface by removing at most finitely many points. In the case of a surface with at least one puncture the automatic structure is explicitly defined (and could be constructed), in terms of a complex whose vertices are *ideal triangulations* on S (triangulations with vertex set the puncture set) and whose edges are elementary moves between ideal triangulations. The more general case can be reduced to the case of a punctured surface using a short exact sequence. The question of whether the mapping class group was in fact biautomatic was finally solved by Hamenstaedt's construction of a biautomatic structure in 2009 [36].

An Artin group is defined to have *large type* if all the associated parameters $\mathsf{m_{ij}}$ are at least 3, *extra large type* if all $\mathsf{m_{ij}}$ are at least 4. For large and especially extra large groups small cancellation techniques associated with negatively curved geometry were developed in [6]. All extra large Artin groups were proved biautomatic in [65], using those small cancellation techniques; the language is a set of geodesics over the standard generating set. All those groups and many others of large type were found by Brady and McCammond [11] to act appropriately on piecewise Euclidean non-positively curved 2-complexes of types A_2 or B_2, and hence, by results of [32, 33] to be biautomatic (but in this case the biautomatic structure is defined over a non-standard generating set).

All Artin groups of large type were proved to be shortlex automatic over their standard generating sets in [42]. A rewrite system was described, which rewrote any word to shortlex geodesic form using sequences of moves on 2-generator substrings. The result extended beyond large type to *sufficiently large* type, where some parameters $\mathsf{m_{ij}}$ might take the value 2 (provided that for any triple i, j, k, if $\mathsf{m_{ij}} = 2$, then either $\mathsf{m_{ik}} = \mathsf{m_{jk}} = 2$ or at least one of $\mathsf{m_{ik}}$ and $\mathsf{m_{jk}}$ is infinite). Biautomatic structures for all large type Artin groups (and in fact for the slightly large class of *almost large* groups) were proved to exist in [47], where all those groups were proved to have appropriate actions on *systolic* complexes. An Artin group is called almost large if for any triple i, j, k it is only possible to have $\mathsf{m_{ij}} = 2$ if one of $\mathsf{m_{ik}}$ or $\mathsf{m_{jk}}$ is infinite, and for any 4-set i, j, k, l at most 2 of $\mathsf{m_{ij}}$, $\mathsf{m_{jk}}$, $\mathsf{m_{kl}}$, $\mathsf{m_{il}}$ can be equal to 2 unless one of the four parameters is infinite.

14.5.2 Coxeter Groups

The proof in [16] of shortlex automaticity of any Coxeter group relative to its standard generating set provided a result that had long been conjectured. We recall that a Coxeter group W is described by a presentation

$$\langle x_1, \ldots, x_n \mid x_i^2 = 1,\ (x_i x_j)^{\mathsf{m_{ij}}} = 1,\ i \neq j \in \{1, \ldots, n\}\rangle,$$

relative to a Coxeter matrix $(\mathsf{m_{ij}})$ and associated Coxeter diagram Γ; the set $X = \{x_1, \ldots, x_n\}$ is its *standard generating set*.

The proof of the theorem constructs an automatic structure for W using properties of its associated *root system*, which arises from the natural isomorphism between W and a reflection group $\overline{W}$ as we now describe; more details can be found in [49]. The group $\overline{W}$ is generated by a set of *reflections* $r_1, \ldots, r_n$ of $\mathbb{R}^n$ defined by $r_i(v) := v - 2\langle v, e_i\rangle e_i$, for $v \in \mathbb{R}^n$, where $e_i : i = 1, \ldots, n$ is a basis for $\mathbb{R}^n$ and $\langle , \rangle$ is the symmetric, bilinear form on $\mathbb{R}^n$ defined by $\langle e_i, e_j\rangle = -\cos(\pi/\mathsf{m_{ij}})$. The isomorphism from W to $\overline{W}$ maps x_i to r_i, and induces an action of W on $\mathbb{R}^n$. The roots of W are defined to be the elements of the set $\Phi = W\{e_1, \ldots, e_n\}$, which decomposes as a disjoint union $\Phi^+ \cup \Phi^-$ of *positive roots* (vectors $\sum \lambda_i e_i$ with all $\lambda_i \geq 0$) and their negatives.

Brink and Howlett's proof of regularity of the set of shortlex geodesic words in W is derived from their proof in [16] of the finiteness of the set of positive roots for W that *dominate* any given positive root; a positive root α is said to dominate a second positive root β if whenever $w(\alpha)$ is negative, for $w \in W$, then so is $w(\beta)$. We define $\widetilde{\Delta}_W$ to be the set of positive roots that dominate no others. Then a word acceptor WA for a shortlex automatic structure for W can be built whose accepting states are all subsets of $\widetilde{\Delta}_W$ [16, Proposition 3.3].

The transitions in WA are determined by the following observation from [16, Lemma 3.1]. When $w = x_{i_1} \cdots x_{i_l}$ is a shortlex geodesic word representing an element of W then, for $x_i \in X$, the word $w' = wx_i$ is non-geodesic if and only if there exists $j \in \{1, \ldots, l\}$ for which $e_i = x_{i_l} \cdots x_{i_{j+1}}(e_{i_j})$. In the case where $w' = wx_i$ is geodesic, that fails to be shortlex minimal if and only if there exists $j \in \{1, \ldots, l\}$ and a generator $x_k \prec x_{i_j}$ for which $e_i = x_{i_l} \cdots x_{i_j}(e_k)$. In that case the word $x_{i_1} \cdots x_{i_{j-1}} x_k x_{i_j} \cdots x_{i_l}$ is shortlex minimal. Based on these two facts, transition on a generator x_i from (the state corresponding to) a subset S of $\widetilde{\Delta}_W$ is to a failure state F if $e_i \in S$. But for $e_i \notin S$, transition is to the intersection with $\widetilde{\Delta}_W$ of the set

$$S'' = \{x_i(\alpha) \mid \alpha \in S\} \cup \{e_i\} \cup \{x_i(e_k) \mid x_k \prec x_i\}.$$

A similar construction to the above, described in [45], proves regularity of the set of all geodesic words in G over S.

The question of whether all Coxeter groups are not just automatic but actually biautomatic remains open. Work of Niblo and Reeves [62] shows that any finitely generated Coxeter group G acts properly discontinuously by isometries on a locally finite, finite dimensional CAT(0) cube complex; their construction is based on the root system Φ associated with G, and an extension of the dominance relation of [16] from Φ^+ to Φ. When the action of G on the cube complex is cocompact, then biautomaticity follows, using [61]. Cocompact actions are proved in [62] to exist whenever G is right-angled or word hyperbolic (by [59] word hyperbolicity of G is recognisable from the diagram Γ). It is also observed in [62] that, by [74], cocompact actions are guaranteed whenever G contains only finitely many conjugacy classes of subgroups isomorphic to rank 3 parabolic subgroups $\langle x_i, x_j, x_k \rangle$ (associated with rank 3 subdiagrams Γ_{ijk} of Γ) for which $\mathsf{m_{ij}}, \mathsf{m_{il}}, \mathsf{m_{jk}}$ are all finite; [21] used this result to derive biautomaticity of G provided that Γ contains no affine subdiagram of rank 3 or more. Subsequently, Caprace [20] proved biautomaticity of all relatively hyperbolic Coxeter groups using results from [46].

The dimension of a Coxeter group is defined to be the dimension of its *Davis complex*, equivalently the maximal rank of any of its spherical parabolic subgroups. It follows that a Coxeter group is 2-dimensional if none of the rank 3 subdiagrams Γ_{ijk} is spherical, equivalently if for all i, j, k, $\frac{1}{\mathsf{m_{ij}}} + \frac{1}{\mathsf{m_{il}}} + \frac{1}{\mathsf{m_{jk}}} \leq 1$. The biautomaticity of all 2-dimensional Coxeter groups is proved in [60]. The construction of a geodesic language generalises ideas from [62], and the result generalises an earlier result proving biautomaticity for certain 2-dimensional groups that used the results of [62].

14.6 Open Problems

More than 30 years after the subject started there continue to be many open problems involving automatic groups. Some of these problems date from the beginning of the subject, and are listed in [29]. Some but not all of these have been mentioned within this chapter. In particular, it remains open whether automatic groups exist that are not biautomatic (see Sect. 14.2) also whether automatic groups exist that do not have soluble conjugacy problem (see Sect. 14.2.2) , whether all soluble automatic groups must be virtually abelian. The most recent progress on this last question was made by the proof of Romankov [66], that a soluble biautomatic group must be virtually abelian (see Sect. 14.2.3.1).

It is still unknown whether a non-biautomatic Coxeter group can exist (Sect. 14.5.2), or a non-automatic Artin group (Sect. 14.5.1).

There are many open problems relating to group actions, in particular, whether a CAT(0) group must be automatic. The very recent construction in [53] of a 3-dimensional CAT(0) group that cannot be biautomatic (Sect. 14.4) represents a major advance on this problem; it does not resolve the question of automaticity. The question of whether biautomaticity or automaticity are implied for a 2-dimensional, piecewise Euclidean CAT(0) group remains open (but we note the recent contribution to this problem of the main result of [60], see Sect. 14.5.2). The 2-dimensional problem is number 43 on a list of open problems within geometric group theory that was published ten years ago in [31], and motivated a body of research, and rapid solution of some of the problems. However, some of the problems listed in this useful and extensive list, or in the earlier list [8], remain open.

References

1. A.V. Aho, Indexed grammars – an extension of context-free grammars. Assoc. Comput. Mach. **15**, 647–671 (1968)
2. J. Alonso, T. Brady, D. Cooper, V. Ferlini, M. Lustig, M. Mihalik, M. Shapiro, H. Short, Notes on word-hyperbolic groups, in *Proceedings of the Conference on Group Theory from a Geometrical Viewpoint*, ed. by E. Ghys, A. Haefliger , A. Verjovsky. held in I.C.T.P., Trieste, March 1990 (World Scientific, Singapore, 1991)
3. J. Alonso, M.R. Bridson, Semihyperbolic groups. Proc. London Math. Soc. **70**(1), 56–114 (1995)
4. J.A. Altobelli, The word problem for Artin groups of FC type. J. Pure Appl. Algebra **129**(1), 1–22 (1998)
5. Y. Antolin, L. Ciobanu, Finite generating sets of relatively hyperbolic groups and applications to geodesic languages. Int. J. Algebra Comput. **11**(04), 467–487 (2001)
6. K.I. Appel, P.E. Schupp, Artin groups and infinite Coxeter groups. Invent. Math. **72**(2), 201–220 (1983)
7. G. Baumslag, S.M. Gersten, M. Shapiro, H.B. Short, Automatic groups and amalgams. J. Pure Appl. Algebra **76**(3), 229–316 (1991)
8. M. Bestvina, Questions in geometric group theory, http://www.math.utah.edu/~bestvina/eprints/questions-updated.pdf
9. B.H. Bowditch, Relatively hyperbolic groups. Int. J. Algebra Comput. **22**(3), 1250016, 66pp. (2012)
10. N. Brady, Sol geometry groups are not asynchronously automatic. Proc. London Math. Soc. **83**(1), 93–119 (2001)
11. T. Brady, J. McCammond, Three-generator Artin groups of large type are biautomatic. J. Pure Appl. Algebra **151**(1), 1–9 (2000)
12. M.R. Bridson, Combings of groups and the grammar of reparameterization. Comment. Math. Helv. **78**(4), 752–771 (2003)
13. M.R. Bridson, R.H. Gilman, Formal language theory and the geometry of 3-manifolds. Comment. Math. Helv. **71**(4), 525–555 (1996)
14. M.R. Bridson, A. Haefliger, Metric spaces of non-positive curvature, in *Grundlehren der mathematischen Wissenschaften* [Fundamental Principles of Mathematical Sciences], vol. 319 (Springer, Berlin, 1999) xxii+643pp.
15. M.R. Bridson, J. Howie, Conjugacy of finite subsets in hyperbolic groups. Internat. J. Algebra Comput. **15**(4), 725–756 (2005)
16. B. Brink, R.B. Howlett, A finiteness property and an automatic structure for Coxeter groups. Math. Ann. **296**(1), 179–190 (1993)
17. D.J. Buckley, D.F. Holt, The conjugacy problem in hyperbolic groups for finite lists of group elements. Internat. J. Algebra Comput. **23**(5), 1127–1150 (2013)
18. J.W. Cannon, The combinatorial structure of cocompact discrete hyperbolic groups. Geom. Dedicata **16**, 123–148 (1984)
19. J.W. Cannon, *Max Dehn and the word problem, Max Dehn: Polymathis mathematician*, ed. by J. Lorenat, J. Mccleary, D. Rowe, M. Senechal, AMS, to appear
20. P.-E. Caprace, Buildings with isolated subspaces and relatively hyperbolic Coxeter groups. Innov. Incidence Geom. **10**, 15–31 (2009)

21. P.-E. Caprace, B. Mühlherr, Reflection triangles in Coxeter groups and biautomaticity. J. Group Theory **8**(4), 467–489 (2005)
22. J. Chalopin, V. Chepo, A. Genevois, H. Hirai, D. Osajda, Helly groups (2021). arXiv:2002.06895
23. R. Charney, Artin groups of finite type are biautomatic. Math. Ann. **292**(4), 671–683 (1992)
24. F. Dahmani, V. Guirardel, The isomorphism problem for all hyperbolic groups, Geom. Funct. Anal. **21**(2), 223–300 (2011)
25. P. Dehornoy, Groupes de Garside. Ann. Sci. Ecole Norm. Sup. **35**(2), 267–306 (2002)
26. D.B.A. Epstein, D.F. Holt, Computation in word hyperbolic groups. Int. J. Algebra Comput. **11**(04), 467–487 (2001)
27. D.B.A. Epstein, D.F. Holt, The linearity of the conjugacy problem in word-hyperbolic groups. Int. J. Algebra Comput. **16**(2), 287–305 (2006)
28. D.B.A. Epstein, D.F. Holt, S.E. Rees, The use of Knuth–Bendix methods to solve the word problem in automatic groups. J. Symb. Comput. **12**, 397–414 (1991)
29. D.B.A. Epstein, J.W. Cannon, D.F. Holt, S. Levy, M.S. Patterson, W.P. Thurston, *Word Processing in Groups* (Jones and Bartlett, Burlington, 1992)
30. B. Farb, Automatic groups: a guided tour. Enseign. Math. **38**(3–4), 291–313 (1992)
31. B. Farb, C. Hruska, A. Thomas, Problems on automorphism groups of nonpositively curved polyhedral complexes and their lattices. *Geometry, Rigidity, and Group Actions*. Chicago Lectures in Mathematics (University of Chicago Press, Chicago, 2011), pp. 515–560
32. S.M. Gersten, H.B. Short, Small cancellation theory and automatic groups. Invent. Math. **102**(2), 305–334 (1990)
33. S.M. Gersten, H.B. Short, Small cancellation theory and automatic groups. II. Invent. Math. **105**(3), 641–662 (1991)
34. R.H. Gilman, Groups with a rational cross-section, in *Combinatorial Group Theory and Topology* (Alta, Utah, 1984). Annals of Mathematics Studies, vol. 111, (Princeton University of Press, Princeton, 1987), pp. 175–183
35. M. Gromov, Hyperbolic groups, in *Essays in Group Theory*. Mathematical Sciences Research Institute Publications, vol. 8, (Springer, New York, 1987), pp. 75–263
36. U. Hamenstaedt, Geometry of the mapping class group II: A biautomatic structure (2009). arXiv:0912.0137
37. A. Harkins, Combing lattices of soluble Lie groups, PhD thesis, University of Newcastle, 2001
38. S. Hermiller, J. Meier, Algorithms and geometry for graph products of groups. J. Alg. **171**, 230–257 (1995)
39. D.F. Holt, An alternative proof that the Fibonacci group F(2,9) is infinite. Exp. Math. **4**(2), 97–100 (1995)
40. D.F. Holt, *KBMAG—Knuth–Bendix in Monoids and Automatic Groups*, software package (1995). Available from http://homepages.warwick.ac.uk/~mareg/download/kbmag2/
41. D.F. Holt, The Warwick automatic groups software, in *Geometric and Computational Perspectives on Infinite Groups* (Minneapolis, MN and New Brunswick, NJ, 1994). DIMACS: Series in Discrete Mathematics and Theoretical Computer Science, vol. 25, (American Mathematical Society, Providence, 1996), pp. 69–82
42. D.F. Holt, S. Rees, Shortlex automaticity and geodesic regularity in Artin groups. Groups Complex. Cryptol. **5**, 1–23 (2013)
43. D.F. Holt, S. Rees, C.E. Röver, *Groups, Languages and Automata*. LMS Student Texts, vol. 88 (Cambridge University Press, London, 2017)
44. J.E. Hopcroft, J.D. Ullman, *Introduction to Automata Theory, Languages and Computation* (Addison-Wesley, Boston, 1979)
45. R.B. Howlett, Miscellaneous facts about Coxeter groups, Lectures given at the ANU Group Actions Worshop, October 1993, Research Report 93–38 (1993)
46. G.C. Hruska, B. Kleiner, Hadamard spaces with isolated flats. Geom. Top. **9**, 1501–1538 (2005)
47. J. Huang, D. Osajda, Large-type Artin groups are systolic. Proc. Lond. Math. Soc. **120**(1), 95–123 (2020)
48. J. Huang, D. Osajda, Helly meets Garside and Artin (2019). arXiv:1904.09060

49. J.E. Humphreys, *Reflection Groups and Coxeter Groups* Cambridge Studies in Advanced Mathematics, vol. 29 (CUP, Cambridge, 1990)
50. T. Januszkiewicz, J. Świątkowski, Simplicial nonpositive curvature. Publ. Math. Inst. Hautes Etudes Sci. **104**, 1–85 (2006)
51. D.L. Johnson, *Presentations of Groups*. LMS Student Texts, vol. 15 (CUP, Cambridge, 1990)
52. U. Lang, Quasigeodesics outside horoballs. Geom. Dedicata **63**(2), 205–215 (1996)
53. I. Leary, A. Minasyan, Commensurating HNNextensions: nonpositive curvature and biautomaticity (2021). arXiv:1907.03515
54. R.C. Lyndon, P.E. Schupp, *Combinatorial Group Theory*. Reprint of the 1977 edition. Classics in Mathematics (Springer, Berlin, 2001)
55. Magma Computational Algebra System version 2.26–8. Available at http://magma.maths.usyd.edu.au/magma/
56. J. Marshall, Computational problems in hyperbolic groups. Int. J. Algebra Comput. **15**(1), 1–13 (2005)
57. L. Mosher, Mapping class groups are automatic. Ann. Math. **142**(2), 303–384 (1995)
58. L. Mosher, Central quotients of biautomatic groups. Comment. Math. Helv. **72**(1), 16–29 (1997)
59. G. Moussong, Hyperbolic Coxeter groups, PhD thesis, Ohio State University, 1988
60. Z. Munro, D. Osajda, P. Przytycki, 2-dimensional Coxeter groups are biautomatic (2020). arXiv:2006.07947
61. G.A. Niblo, L.D. Reeves, The geometry of cube complexes and the complexity of their fundamental groups. Topology **37**, 621–633 (1998)
62. G.A. Niblo, L.D. Reeves, Coxeter groups act on CAT(0) cube complexes. J. Group Theory **6**(3), 399–413 (2003)
63. D. Osin, *Relatively Hyperbolic Groups: Intrinsic Geometry, Algebraic Properties, and Algorithmic Problems*. Memoirs of the American Mathematical Society, vol. 179, No. 843 (American Mathematical Society, Providence, 2006)
64. P. Papasoglu, Strongly geodesically automatic groups are hyperbolic. Invent. Math. **121**(2), 323–334 (1995)
65. D. Peifer, Artin groups of extra-large type are biautomatic. J. Pure Appl. Algebra **110**(1), 15–56 (1996)
66. V. Romankov, Polycyclic, metabelian or soluble of type $(FP)_\infty$ groups with Boolean algebra of rational sets and biautomatic soluble groups are virtually abelian. Glasg. Math. J. **60**(1), 209–218 (2018)
67. Z. Sela, The isomorphism problem for hyperbolic groups. I. Ann. Math. **141**(2), 217–283 (1995)
68. H.B. Short, Groups and combings. https://www.i2m.univ-amu.fr/~short/Papers/bicomball.pdf
69. The GAP Group. *GAP — Groups, Algorithms, and Programming, Version 4.8.10* (2018). http://www.gap-system.org
70. W.P. Thurston, Three-dimensional manifolds, Kleinian groups and hyperbolic geometry. Bull. Amer. Math. Soc. (NS) **6**(3), 357–381 (1982)
71. W.P. Thurston, *Three-Dimensional Geometry and Topology* (Princeton University Press, Princeton 1997)
72. L.A. VanWyk, Graph groups are biautomatic. J. Pure Appl. Algebra **94**(3), 341–352 (1994)
73. P. Wakefield, Procedures for automatic groups, PhD thesis, University of Newcastle, 1998
74. B.T. Williams, Two topics in geometric group theory, PhD thesis, University of Southampton, 1999
75. R. Young, The Dehn function of SL(n; $\mathbb{Z}$). Ann. Math. **177**(3), 969–1027 (2013)

Chapter 15
Geometry and Combinatorics via Right-Angled Artin Groups

Thomas Koberda

Abstract We survey the relationship between the combinatorics and geometry of graphs and the algebraic structure of right-angled Artin groups. We concentrate on the defining graph of the right-angled Artin group and on the extension graph associated to the right-angled Artin group. Additionally, we discuss connections with geometric group theory and complexity theory.

Keywords Right-angled Artin group · Extension graph · Graph expanders · Hamiltonian graph · k-Colorability · Graph automorphism · Acylindrical group action · Quasi-isometry · Commensurability · Mapping class group · Curve graph

2020 Mathematics Subject Classification Primary: 20F36, 20F65, 05C50; Secondary: 05C45, 05C48, 05C60, 68Q15, 03D15

15.1 Introduction

In this paper, we survey the interplay between the algebraic structure of right-angled Artin groups, the combinatorics of graphs, and geometry. Throughout the paper, let Γ be a finite simplicial graph, and we write $V(\Gamma)$ and $E(\Gamma)$ for the set of vertices and edges of Γ, respectively. The *right-angled Artin group* on Γ, denoted by $A(\Gamma)$, is the group defined by

$$A(\Gamma) = \langle V(\Gamma) \mid [v, w] = 1 \text{ if and only if } \{v, w\} \in E(\Gamma)\rangle.$$

T. Koberda (✉)
Department of Mathematics, University of Virginia, Charlottesville, VA, USA
https://sites.google.com/view/koberdat

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9_15

15.1.1 Scope of This Survey

Right-angled Artin groups interpolate between free groups and abelian groups, and they exhibit a wide range of complex phenomena. Moreover, they are simple enough that their structure is relatively tractable, and hence one can come to understand these groups fairly well. They are prototypical examples of CAT(0) groups, and they serve as toy examples that mirror many important properties of and inform conjectures about more complicated groups, such as mapping class groups. Some well-known and difficult conjectures about mapping class groups, such as the characterization of convex cocompact subgroups, admit complete, tractable analogues in the case of right-angled Artin groups; see [88] for a detailed discussion.

In this article, we will concentrate on some specific aspects of right-angled Artin groups, which we will outline in the remainder of this section. For a survey of the general properties of right-angled Artin groups, the reader is directed to [36].

Some of the basic questions we will discuss are as follows.

Question 15.1.1 *What is the exact relationship between the group theoretic structure of the group $A(\Gamma)$ and the combinatorial structure of Γ?*

The reader will find that there are two answers to Question 15.1.1, the trivial one and the nontrivial one. The trivial one will be a consequence of Theorem 15.2.6 below, which shows that Γ is completely determined by the cohomology algebra of $A(\Gamma)$, and in fact by the degree one and two parts together with the cup product pairing. Thus, one can in principle recover Γ from $A(\Gamma)$, so that any combinatorial properties of Γ is automatically determined by the algebraic structure of $A(\Gamma)$. Conversely, the algebraic structure of $A(\Gamma)$ is, in a sense that is so general as to render it almost meaningless, "known" by the graph Γ.

There is a more interesting approach to Question 15.1.1 that seeks to find a dictionary between the combinatorics of Γ and the algebra of $A(\Gamma)$, by passing between specific graph-theoretic and group-theoretic properties that are analogous. This line of inquiry yields some otherwise nonobvious insights that have applications outside of geometric group theory, such as in cryptography and complexity theory. Some sample results we will discuss in the sequel are the following:

Theorem 15.1.2 ([51]) *Let Γ be a finite simplicial graph. Then Γ admits a nontrivial automorphism if and only if the outer automorphism group* $\mathrm{Out}(A(\Gamma))$ *contains a finite nonabelian group.*

Theorem 15.1.3 ([54]) *Let Γ be a finite simplicial graph with n vertices. Then Γ admits a k-coloring if and only if $A(\Gamma)$ surjects to a product*

$$F_{n_1} \times \cdots \times F_{n_k},$$

where F_{n_i} is a free group of rank n_i, and

$$\sum_{i=1}^{k} n_i = n.$$

Theorem 15.1.4 ([55]) *Let Γ be a finite simplicial graph. Then Γ admits a Hamiltonian cycle if and only if the cohomology algebra of $A(\Gamma)$ is a Hamiltonian vector space.*

In Theorem 15.1.4, *Hamiltonicity* of a vector space means that there is a bilinear form satisfying certain "connectivity" conditions. We direct the reader to Sect. 15.3.3 for precise definitions.

Whereas the graph Γ is evidently intimately related to the structure of $A(\Gamma)$, the graph Γ is not always ideally suited for the study of the internal structure of $A(\Gamma)$, since there is no natural interesting action of $A(\Gamma)$ on Γ. However, one can augment Γ in the "smallest way possible" in order to get a graph on which $A(\Gamma)$ acts. The key idea is to conflate a vertex v of Γ with an element of $A(\Gamma)$. One can then consider the set

$$V(\Gamma^e) = \{v^g \mid v \in V(\Gamma), g \in A(\Gamma)\} \subset A(\Gamma)$$

of all conjugates of vertices of Γ, where here we write $v^g = g^{-1}vg$. It is true though largely irrelevant that $V(\Gamma^e)$ is not canonically defined as a subset of $A(\Gamma)$, since automorphisms of $A(\Gamma)$ need not preserve the set of conjugates of given vertex generators of $A(\Gamma)$.

We build a graph Γ^e, called the *extension graph* of Γ (cf. [78]), by putting an edge $\{v^g, w^h\}$ between vertices in $V(\Gamma^e)$ whenever $[v^g, w^h] = 1$ in $A(\Gamma)$. The group $A(\Gamma)$ now acts in a canonical way on Γ^e, i.e. by conjugation.

Question 15.1.5 *What is the relationship between the structure of $A(\Gamma)$ and the structure of Γ^e? What is the geometry of the action of $A(\Gamma)$ on Γ^e?*

The graph Γ^e, though algebraically defined, is very closely related to Hagen's *contact graph* [63], which encodes the intersection pattern between hyperplanes in a natural CAT(0) cube complex on which $A(\Gamma)$ acts. This, together with an analogy between the extension graph and the curve graph associated to a hyperbolic surface of finite type, is an entry point into the theory of hierarchically hyperbolic spaces (HHSs) and hierarchically hyperbolic groups (HHGs) (see [15, 16], for instance). We will largely avoid discussing that aspect of the theory in this paper.

The extension graph carries a large amount of data about the subgroup structure of $A(\Gamma)$. A sample result we will discuss is the following:

Theorem 15.1.6 ([78]) *Suppose Γ has no triangles, and let Λ be an arbitrary finite simplicial graph. Then $A(\Lambda)$ occurs as a subgroup of $A(\Gamma)$ if and only if Λ occurs as a subgraph of Γ^e.*

The action of $A(\Gamma)$ on Γ^e by conjugation, though perhaps simple at first glance, serves to unify the group theory of $A(\Gamma)$, the geometry of Γ^e, and the intrinsic CAT(0) geometry of $A(\Gamma)$. We state the following result that we will discuss in some detail, and we will defer definitions of the terminology until then.

Theorem 15.1.7 ([79]) *The action of $A(\Gamma)$ on Γ^e is acylindrical. Moreover, the following are equivalent.*

(1) The element $g \in A(\Gamma)$ acts loxodromically on Γ^e.
(2) The element $g \in A(\Gamma)$ acts as a rank one isometry of the universal cover of the Salvetti complex of Γ.
(3) The element $g \in A(\Gamma)$ is not conjugate into a join subgroup of $A(\Gamma)$.

We will not give detailed proofs of most of the results in this survey. We will give proof sketches where it is feasible, and we will strive to give complete references. As already suggested above, we will omit large parts of the theory and neglect various viewpoints. The specific topics discussed herein undoubtedly reflect the idiosyncratic tastes of the author.

15.1.2 Notation and Terminology

Most of the notation and terminology used in this survey is standard or nearly standard. All graphs will be undirected and simplicial unless otherwise noted, so that in particular there are no double edges nor edges that start and end at a single vertex. The *complement* of a graph Γ is the complement of Γ in the complete graph on the vertices of Γ; that is, complete all the missing edges of Γ and then delete the edges that were present in Γ. Two vertices are therefore connected by an edge in the complement of Γ if and only if they are not connected by an edge in Γ.

A graph Γ is a *join* if its complement graph is disconnected. The join of graphs Γ_1 and Γ_2 is written $\Gamma_1 * \Gamma_2$, and every vertex of Γ_1 is adjacent to every vertex of Γ_2. The join of two graphs mimics the geometric join in topology: if A and B are topological spaces, then the join $A * B$ is the quotient of $A \times B \times I$ that collapses $A \times B \times \{0\}$ to A and $A \times B \times \{1\}$ to B. For us, a subgraph Λ of a graph Γ is always *full*, which is to say Λ contains all edges that are present in Γ. A *clique* is a complete graph, and a *k-clique* is a complete graph on k vertices. The set $V(\Gamma)$, viewed as a subset of $A(\Gamma)$, is called the set of *vertex generators* of $A(\Gamma)$. The *link* of a vertex $v \in V(\Gamma)$ is written $\mathrm{Lk}(v)$ and consists of the vertices that are adjacent to v. If $\emptyset \neq S \subset V(\Gamma)$ then

$$\mathrm{Lk}(S) = \bigcap_{s \in S} \mathrm{Lk}(s).$$

The *star* of v is given by $\mathrm{St}(v) = \mathrm{Lk}(v) \cup \{v\}$. The *degree* of a vertex v is given by $|\mathrm{Lk}(v)|$. A vertex v is *isolated* if $\mathrm{Lk}(v)$ is empty. A graph is *totally disconnected* if every vertex is isolated. A *path* in Γ is a tuple of vertices $\overline{p} = (v_1, \ldots, v_k)$ in $V(\Gamma)$ such that $\{v_i, v_{i+1}\} \in E(\Gamma)$ for all suitable indices. The parameter k is arbitrary, and the *length* of the path $\overline{p}$ is $k - 1$. A *cycle* or *circuit* is a path for which $v_1 = v_k$ and for which $v_i \neq v_{i+2}$ for all suitable indices. A graph is *connected* if for all pairs of vertices $v, w \in V(\Gamma)$, there is a path in Γ such that $v = v_1$ and $w = v_k$.

The *rank of a linear map* is the dimension of its image, and the *rank of a group* is the minimal number of generators of the group. The identity element of a group is denoted 1 with an exception in the case of additive abelian groups when it is written 0.

Let Γ be a (possibly infinite) graph. We build a graph Γ_k, called the *clique graph* of Γ as follows. We start with the vertices and edges of Γ. For every complete subgraph $K \subset V(\Gamma)$ with at least two vertices, we add a new vertex v_K to $V(\Gamma_k)$. If K_1 and K_2 are cliques such that $V(K_1) \cup V(K_2)$ also spans a complete subgraph of Γ, then we add an edge $\{v_{K_1}, v_{K_2}\}$ to $E(\Gamma_k)$. Finally, we add an edge between each vertex of the form v_K and the vertices making up K. The resulting graph is the clique graph. It is helpful to illustrate the clique graph with an example: if Γ is a graph without triangles then the only cliques with two or more vertices are the edges of Γ. In this case, the clique graph of Γ is just Γ with an extra vertex v_e for each edge $e \in E(\Gamma)$, and two edges connecting v_e to the two vertices of Γ spanning e. Thus, Γ_k is just a copy of Γ with a "fin" hanging off each edge.

15.1.3 A Remark About Generators

When we specify a right-angled Artin group, we will write $A(\Gamma)$. Since $A(\Gamma)$ as an abstract group determines Γ up to isomorphism, the specification of Γ (viewed as an abstract graph) does not constitute a choice of generators for $A(\Gamma)$. However, once we speak of particular generators of $A(\Gamma)$, we have implicitly chosen an identification of $V(\Gamma)$ with a set of generators for $A(\Gamma)$. The author has taken pains to avoid ambiguities that could cause confusion for the reader.

15.2 The Cohomology Ring of a Right-Angled Artin Group

A central role in the dictionary between algebra and combinatorics is played by the cohomology of a right-angled Artin group. Recall that the cohomology of a group G is defined to be the cohomology of a $K(G, 1)$, which is unique up to homotopy equivalence (see [64], for instance). A right-angled Artin group has a very easy to describe $K(G, 1)$, and a large number of natural retractions allows for an efficient calculation of the cohomology algebra. For the entirety of this section, R will denote a commutative ring with a unit, unless otherwise noted.

15.2.1 The Topology of the Salvetti Complex

We will write $\mathscr{S}(\Gamma)$ for the Salvetti complex of Γ, and we construct it as follows (cf. [36]). Let Γ be a graph with $n = |V(\Gamma)|$. We fix a bijection between $V(\Gamma)$

and $\{1, \ldots, n\}$. Consider now the unit cube $[0, 1]^n \subset \mathbb{R}^n$. We build a certain subset $S \subset [0, 1]^n$, cube by cube. For $1 \leq i \leq n$, we write J_i for the unit segment in the i^{th} coordinate direction, emanating from the origin. We include J_i in S for all i. Now, if $K \subset \{1, \ldots, n\}$ consists of a collection of vertices which span a complete subgraph of Γ, then we include the subcube of $[0, 1]^n$ spanned by $\{J_i\}_{i\in K}$ in S.

Once S has been constructed in this way, we set $\mathscr{S}(\Gamma)$ to be the image of S in $\mathbb{R}^n/\mathbb{Z}^n$, where $\mathbb{Z}^n$ acts on $\mathbb{R}^n$ by usual integer translations. Thus, the complex $\mathscr{S}(\Gamma)$ is realized as a subcomplex of an n-dimensional torus.

Proposition 15.2.1 *The following are properties of $\mathscr{S}(\Gamma)$.*

1. *The fundamental group of $\mathscr{S}(\Gamma)$ is isomorphic to $A(\Gamma)$.*
2. *The universal cover of $\mathscr{S}(\Gamma)$ is contractible.*

That the fundamental group of $\mathscr{S}(\Gamma)$ is isomorphic to $A(\Gamma)$ is a straightforward calculation using Van Kampen's Theorem. That the universal cover of $\mathscr{S}(\Gamma)$ is contractible is much less obvious, and follows from the fact that $\mathscr{S}(\Gamma)$ admits the structure of a locally CAT(0) cube complex. To delve into the details would take us far afield, and we shall content ourselves to direct the reader to some references, such as [29, 62, 119]. The crucial point here is that the homology and cohomology of $\mathscr{S}(\Gamma)$ are in fact a invariants of $A(\Gamma)$, since $\mathscr{S}(\Gamma)$ is a $K(G, 1)$ for $G = A(\Gamma)$.

The homology of $\mathscr{S}(\Gamma)$ is easily calculated by a standard Mayer–Vietoris argument. In our construction of $\mathscr{S}(\Gamma)$ above, we obtain a distinguished k-subtorus of $\mathscr{S}(\Gamma)$ for every k-subclique of Γ. When two such distinguished subtori (corresponding to subcliques K_1 and K_2 of Γ) meet, they meet along the distinguished subtorus corresponding to the intersection $K_1 \cap K_2$ (which is just the basepoint in case this intersection is empty). Thus, we see that:

Proposition 15.2.2 *Let R be a ring. Then $H_k(A(\Gamma), R) \cong R^{N_k}$, where N_k denotes the number of k-cliques in Γ.*

Here and throughout, we always assume that the $A(\Gamma)$ action on the ring of coefficients is trivial, so that our homology and cohomology groups are always untwisted. Computation of the twisted groups is much more complicated; cf. [40, 73]. In particular, the rank of the abelianization of $A(\Gamma)$ is the number of vertices of Γ, and the dimension of the second homology coincides with the number of edges.

The cohomology groups of $A(\Gamma)$ have the same ranks as the homology groups, and the formal structure of $A(\Gamma)$ (or of $\mathscr{S}(\Gamma)$) allows one to give a satisfactory description of the cohomology algebra of $A(\Gamma)$. For this, we let T_k denote the k-dimensional torus. As is standard, the cohomology algebra of T_k with coefficients in R is $\bigwedge(R^k)$, the exterior algebra of R^k.

Proposition 15.2.3 *Let $\Lambda \subset \Gamma$ be a subgraph. Then the map $A(\Gamma) \longrightarrow A(\Lambda)$ defined by the identity for vertices $\lambda \in V(\Lambda)$ and by $v \mapsto 1$ otherwise is a retraction of groups.*

Of course, the fact that Salvetti complexes are classifying spaces for right-angled Artin groups means that Proposition 15.2.3 admits a dual statement for spaces. That is, there are natural retractions $\mathscr{S}(\Gamma) \longrightarrow \mathscr{S}(\Lambda)$ which induce the corresponding maps on fundamental groups whenever $\Lambda \subset \Gamma$ is a subgraph.

Specializing to the case where $K \subset \Gamma$ is a k-clique, we get a natural surjective map

$$A(\Gamma) \longrightarrow A(K) \cong \mathbb{Z}^k,$$

and thus an induced injective map on cohomology $\bigwedge(R^k) \longrightarrow H^*(A(\Gamma), R)$. Suppose we have a decomposition of graphs $\Gamma = \Lambda_1 \cup \Lambda_2$ and $\Theta = \Lambda_1 \cap \Lambda_2$. For technical reasons, we suppose that every edge between Λ_1 and Λ_2 is realized by an edge between Λ_i and Θ for $i \in \{1, 2\}$. We obtain a natural commutative diagram of retractions.

$$\begin{array}{ccc} A(\Gamma) & \longrightarrow & A(\Lambda_1) \\ \downarrow & & \downarrow \\ A(\Lambda_2) & \longrightarrow & A(\Theta) \end{array}$$

Replacing the retractions by inclusions of groups, $A(\Gamma)$ acquires the structure of a graph of groups with vertex groups $A(\Lambda_1)$ and $A(\Lambda_2)$ and edge group $A(\Theta)$ (cf. [105]). Without the assumption that every edge between Λ_1 and Λ_2 be realized by Θ, this previous assertion would no longer be true.

Dualizing, we get a commutative diagram on cohomology.

$$\begin{array}{ccc} H^*(A(\Gamma), R) & \longleftarrow & H^*(A(\Lambda_1), R) \\ \uparrow & & \uparrow \\ H^*(A(\Lambda_2), R) & \longleftarrow & H^*(A(\Theta), R) \end{array}$$

In category theory language, $H^*(A(\Gamma), R)$ is the pushout of the corresponding diagram. Again, the technical hypothesis on the decomposition of Γ is hidden in this last assertion, since the assertion follows from the Mayer–Vietoris sequence and would be false without this hypothesis (cf. for example when Γ is a complete graph and Λ_1 and Λ_2 are both proper subgraphs).

These considerations show that one can describe the cohomology algebra of $A(\Gamma)$ entirely in terms of exterior algebras by inductively building up Γ from its cliques. In particular, one can take an appropriate exterior algebra for each maximal clique in Γ, and identify exterior subalgebras corresponding to intersections of maximal cliques. The simplest cliques are the 1-cliques, and a retraction

$$A(\Gamma) \longrightarrow \langle v \rangle \cong \mathbb{Z}$$

for $v \in V(\Gamma)$ allows us to identify preferred generators $\{v^* \mid v \in V(\Gamma)\}$ for $H^1(A(\Gamma), R)$, which we will refer to as the *dual 1-classes* to the vertex generators.

These dual 1-classes can be interpreted as dual to certain natural subspaces of $\mathscr{S}(\Gamma)$, though we will not require this point of view here.

We clearly have that $H^*(A(\Gamma), R)$ is generated by its degree one part. Now let $v, w \in V(\Gamma)$ and let v^* and w^* be the corresponding dual 1-classes. There is a retraction $A(\Gamma) \longrightarrow \langle v, w\rangle$, and the target group is either $\mathbb{Z}^2$ or F_2, corresponding to the cases where $\{v, w\} \in E(\Gamma)$ and where $\{v, w\} \notin E(\Gamma)$ respectively. In the first case, the cup product $v^* \smile w^*$ is nontrivial and in the second case, the cup product vanishes.

The most important consequence of the previous discussion for us in the sequel is the following, which characterizes the degree one and degree two parts of the cohomology of $A(\Gamma)$ together with the cup product pairing:

Proposition 15.2.4 *Let Γ be a finite simplicial graph with $V(\Gamma) = \{v_1, \ldots, v_n\}$ and let $E(\Gamma) = \{e_1, \ldots, e_m\}$. Then there are bases $\{v_1^*, \ldots, v_n^*\}$ and $\{e_1^*, \ldots, e_m^*\}$ for $H^1(A(\Gamma), R)$ and $H^2(A(\Gamma), R)$ respectively, such that:*

(1) $v_i^* \smile v_j^* = 0$ *if* $\{v_i, v_j\} \notin E(\Gamma)$.
(2) $v_i^* \smile v_j^* = \pm e_\ell^*$ *if* $\{v_i, v_j\} = e_\ell$.

The description of $H^1(A(\Gamma), R)$ and $H^2(A(\Gamma), R)$ furnished by Proposition 15.2.4 will be essential in describing many of the correspondences between the group theoretic structure of $A(\Gamma)$ and the combinatorics of Γ.

15.2.2 Vector Spaces with a Vector-Space Valued Pairing

In the sequel, it is sometimes convenient to consider vector spaces equipped with a bilinear vector-space valued pairing. We will write $q \colon V \times V \longrightarrow W$ for such a pairing, where V and W are both finite dimensional vector spaces over the same field F. The pairing q is intended to generalize the cup product pairing

$$\smile \colon H^1(A(\Gamma), F) \times H^1(A(\Gamma), F) \longrightarrow H^2(A(\Gamma), F),$$

and so we will always adopt the assumption that q is either symmetric or anti-symmetric unless otherwise noted. This assumption on q is mostly for convenience, since relaxing some sort of symmetry assumption only adds unnecessary layers of complication that do not enrich the underlying theory in a meaningful way.

We will say that the triple (V, W, q) is *pairing-connected*, if for all nontrivial direct sum decompositions $V \cong V_0 \oplus V_1$, there are vectors $v_0 \in V_0$ and $v_1 \in V_1$ such that $q(v_0, v_1) \neq 0$. With this terminology, we can formulate and prove an entry in the algebra-combinatorics dictionary.

Proposition 15.2.5 (See [53]) *Let Γ be a finite simplicial graph, let*

$$V = H^1(A(\Gamma), F), \quad W = H^2(A(\Gamma), F),$$

and let q be the cup product pairing. Then Γ is connected if and only if (V, W, q) is pairing-connected.

Connectedness of Γ has another, simpler characterization in terms of $A(\Gamma)$, as we shall indicated below; namely, Γ is connected if and only if $A(\Gamma)$ is freely indecomposable; see Theorem 15.3.2. The (mostly complete) proof of Proposition 15.2.5 will illustrate the principle that many results that related the algebra of $A(\Gamma)$ with the combinatorics of Γ have an easy direction and a less easy direction.

Proof of Proposition 15.2.5 Suppose first that (V, W, q) is pairing-connected, and let $\Gamma = \Gamma_0 \cup \Gamma_1$ be a purported separation of Γ. Let V_i denote the span of the vertices $\{v_j^* \mid v_j \in V(\Gamma_i)\}$ for $i \in \{0, 1\}$. Pairing connectedness implies that there are vectors $w_i \in V_i$ such that $q(w_0, w_1) \neq 0$. Writing w_0 and w_1 in terms of the preferred basis vectors, we see that there are vertices $x_i \in V(\Gamma_i)$ such that $q(x_0^*, x_1^*) \neq 0$, which implies that $\{x_0, x_1\} \in E(\Gamma)$ by Proposition 15.2.4, a contradiction.

Suppose conversely that Γ is connected, and let $V \cong V_0 \oplus V_1$ be a nontrivial direct sum decomposition that witnesses the failure of (V, W, q) to be pairing-connected. Let $\{x_1, \ldots, x_m\}$ be a sequence of vertices of Γ such that every vertex of Γ appears on this list, and such that for all suitable i we have $\{x_i, x_{i+1}\} \in E(\Gamma)$. We allow this list to have repeats.

Let

$$w_0 = \sum_{i=1}^{n} \alpha_i v_i^* \in V_0, \;\; w_1 = \sum_{i=1}^{n} \beta_i v_i^* \in V_1$$

be expressions for nonzero vectors with respect to the standard dual basis for V. If $\{v_i, v_j\} \in E(\Gamma)$ then the expression $q(w_0, w_1) = 0$ implies that $\alpha_i \beta_j = \alpha_j \beta_i$. The two sides of this last equation are either both zero or both nonzero, and in the latter case we have that the pairs (α_i, α_j) and (β_i, β_j) are proportional. In this case, since $\{v_1^*, \ldots, v_n^*\}$ is a basis for V, we may perturb w_0 or w_1 within the respective vector spaces V_0 and V_1 in order to obtain vectors for which the coefficients corresponding to v_i^* and v_j^* are not proportional. Thus, the condition $q(w_0, w_1)$ implies that

$$\alpha_i \beta_j = \alpha_j \beta_i = 0.$$

With these observations, we can complete the proof. Let w_0 be as above. Relabeling if necessary, we have $v_1 = x_1$ and $v_2 = x_2$. Without loss of generality, we may assume that $\alpha_1 \neq 0$. Now let $w_1 \in V_1$ be expressed as above. If $\beta_2 \neq 0$ then $\alpha_1 \beta_2 \neq 0$, a conclusion that was ruled out by the considerations in the previous paragraph. Thus, $\beta_2 = 0$, and since w_1 was arbitrary, the coefficient of x_2^* vanishes for all vectors in V_1. Then, we may find a vector in V_0 whose coefficient α_2 is nonzero, and arguing symmetrically, we see that the coefficient β_1 is zero for all vectors in V_1. By induction on m and using the fact every vertex of Γ occurs on the list $\{x_1, \ldots, x_m\}$, we see that V_1 must be the zero vector space. This is a contradiction. □

15.2.3 The Cohomology Ring of $A(\Gamma)$ Determines Γ

We are now ready to state and prove a central fact about the cohomology of $A(\Gamma)$, namely that it determines the isomorphism type of Γ.

Theorem 15.2.6 *Let Γ be a finite simplicial graph, let $V = H^1(A(\Gamma), F)$, let $W = H^2(A(\Gamma), F)$, and let q be the cup product pairing. Then the triple (V, W, q) determines Γ up to isomorphism.*

One essential point in Theorem 15.2.6 is that the triple (V, W, q) is considered abstractly, without any further data such as bases. Before giving a proof of Theorem 15.2.6, we can make several observations about special instances of the result. First, the dimension of $V = H^1(A(\Gamma), F)$ coincides with $|V(\Gamma)|$, and the dimension of $W = H^2(A(\Gamma), F)$ coincides with $|E(\Gamma)|$, as is immediate from Proposition 15.2.4. Moreover, the first and second cohomology of $A(\Gamma)$ together with the cup product pairing identify complete graphs. To see this, it is convenient to introduce a map $V \longrightarrow \mathrm{Hom}(V, W)$, defined by $v \mapsto f_v$, and where $f_v(v') = v \smile v'$. The graph Γ is complete if and only if for all $v \in H^1(A(\Gamma), F)$, the rank of the image of f_v is $\dim H^1(A(\Gamma), F) - 1$. We leave the verification of this last claim as a straightforward exercise for the reader.

The fact that $A(\Gamma)$ determines the graph Γ uniquely is well-known. See [44, 85, 102] for several perspectives. The proof offered here that gives uniqueness of Γ via the cohomology algebra of $A(\Gamma)$ fits into the theory of *cohomological uniqueness*. In the context of cohomological uniqueness, one is often concerned with the question of whether or not a particular space (often decorated with adjectives such as p–completeness, where p is a prime) is determined up to homotopy equivalence by its cohomology (with various groups of coefficients). In our setting, Theorem 15.2.6 implies that among Salvetti complexes associated to finite simplicial graphs, the integral (or rational) cohomology of the space determines the space up to homotopy equivalence, and its defining graph up to isomorphism. Moreover, only the ring structure on the cohomology algebra is required, and only in degrees one and two. The reader is directed to [42, 101, 115, 116] for a more detailed discussion of cohomological uniqueness.

Another perspective on the cup product pairing determining $A(\Gamma)$ and Γ uniquely is given by the theory of 1-formality in the sense of D. Sullivan, a property which is shared notably with Kähler groups, and into which we will not delve in deeply. The reader is directed to Chapter 3 of [4], for instance.

Proof of Theorem 15.2.6 We will actually prove a stronger statement. Suppose $\Gamma \longrightarrow \Lambda$ is obtained by deleting vertices (so that Λ is a subgraph of Γ), with an induced retraction $A(\Gamma) \longrightarrow A(\Lambda)$ defined by sending the vertices $V(\Gamma) \setminus V(\Lambda)$ to the identity. Thus, we obtain triples $(V_\Gamma, W_\Gamma, q_\Gamma)$ and $(V_\Lambda, W_\Lambda, q_\Lambda)$ corresponding to the cohomologies of these groups, and a map of triples

$$i_{\Lambda,\Gamma} \colon (V_\Lambda, W_\Lambda, q_\Lambda) \longrightarrow (V_\Gamma, W_\Gamma, q_\Gamma),$$

which is injective on the level of vector spaces, and q_Λ is extended by q_Γ.

Claim The triple

$$\{(V_\Lambda, W_\Lambda, q_\Lambda), (V_\Gamma, W_\Gamma, q_\Gamma), i_{\Lambda,\Gamma}\}$$

uniquely determines graphs Λ and Γ, together with an injection of graphs $\Lambda \longrightarrow \Gamma$. The theorem will then follow from the special case where $V(\Lambda) = \varnothing$.

We proceed by induction on $(|V(\Gamma)|, |V(\Lambda)|)$, ordered lexicographically, the cases where $|V(\Gamma)| \in \{1, 2\}$ being easy consequences of the remarks preceding the proof. We now suppose the claim has been established for all graphs with at most n vertices, and we suppose that Γ has $n+1$ vertices. We consider the (possibly trivial) subspace

$$V_0 \subset V = H^1(A(\Gamma), F)$$

spanned by vectors for which f_v has rank zero. It is immediate from Proposition 15.2.4 that a vector $w \in V_0$ is in the span on vectors dual to vertices of degree zero in Γ. The quotient V/V_0 is isomorphic to $H^1(A(\Gamma'), F)$, where Γ' is the result of deleting all the vertices of Γ that have degree zero.

The natural map $A(\Gamma) \longrightarrow A(\Gamma')$ given by sending isolated vertices to the identity induces a map $H^1(A(\Gamma', F) \longrightarrow H^1(A(\Gamma), F)$, which identifies V/V_0 with a subspace of V. The cup product on $H^1(A(\Gamma), F)$ restricts to

$$q\colon H^1(A(\Gamma'), F) \times H^1(A(\Gamma'), F) \longrightarrow W.$$

Thus, if $V_0 \neq 0$ then Γ' satisfies the conclusion of the claim by induction, and Γ is obtained from Γ' by adding $\dim V_0$ many isolated vertices. We may therefore assume that Γ has no isolated vertices.

We now consider a vector $v \in V$ such that the rank of f_v is minimized.

Case 1 Suppose first that the linear span U of v coincides with the span of a vector dual to a vertex x of Γ, as furnished by Proposition 15.2.4. Then V/U coincides with the first cohomology of $A(\Gamma')$, where Γ' is obtained from Γ by deleting x. The map $V \longrightarrow V/U$ is induced by the inclusion $\Gamma' \longrightarrow \Gamma$ and the corresponding injection $A(\Gamma') \longrightarrow A(\Gamma)$.

Writing Z for the image of f_v, we have that W/Z coincides with the second cohomology of $A(\Gamma')$, and the cup product pairing descends to a bilinear map

$$\overline{q}\colon V/U \times V/U \longrightarrow W/Z,$$

which coincides with the cup product pairing on the cohomology of $A(\Gamma')$. By induction, the triple $(V/U, W/Z, \overline{q})$ determines Γ' uniquely.

Let $N \subset V$ be the kernel of f_v. Then N is spanned by the dual vector v associated to the vertex x and the duals of the vertices which are not adjacent of

x. If $N = U$ then Γ is the join of x and Γ'. If not, then we pass to the quotient V/N, which coincides with $H^1(A(\mathrm{Lk}(x)), F)$. Again, the map $V \longrightarrow V/N$ is induced by the inclusion of $A(\mathrm{Lk}(x)) \longrightarrow A(\Gamma)$. Passing to a suitable quotient W/Y of W as above, we can recover the cup product pairing on the cohomology of $A(\mathrm{Lk}(x))$, and thus recover $\mathrm{Lk}(x)$, by induction. Finally, we use the full strength of the induction hypothesis to obtain an injection $i_x\colon \mathrm{Lk}(x) \longrightarrow \Gamma'$. The graph Γ is now reconstructed by attaching x to each vertex in the image of i_x.

To complete the induction, let

$$\{(V_\Lambda, W_\Lambda, q_\Lambda), (V_\Gamma, W_\Gamma, q_\Gamma), i_{\Lambda,\Gamma}\}$$

be a triple satisfying the hypotheses of the claim. We quotient out the degree one part of the cohomology V_Λ and V_Γ by U, and the map $i_{\Lambda,\Gamma}$ descends to the quotients by hypothesis. By induction, we obtain an injection of graphs $\Lambda' \longrightarrow \Gamma'$, where the primed graphs are obtained by deleting the vertex x. The links of x in Γ' and Λ' can be determined as above, whence we can reconstruct Γ.

Case 2 Suppose that $v \in V$ is arbitrary such that the rank k of f_v is minimized, and suppose that v is supported on the duals of two or more vertices, so that

$$v = \sum_{i=1}^{m} \alpha_i x_i^*,$$

where all indices have nonzero coefficients and $m \geq 2$. It is clear that for all i, the degree of x_i must coincide with k, by an easy application of Proposition 15.2.4. Consider the vertices x_1 and x_2. Observe that Proposition 15.2.4 again implies that there cannot be a vertex that is distinct from both x_1 and x_2 and that is adjacent to x_1 but not to x_2. Thus, every vertex that is adjacent to x_1 and distinct from x_2 is also adjacent to x_2. By symmetry, the same statement holds after switching the roles of x_1 and x_2. The argument now bifurcates into two subcases, according to whether x_1 and x_2 are adjacent or not.

Subcase 1 Suppose first that x_1 and x_2 are adjacent, and suppose that $m \geq 3$. Suppose that x_3 is not adjacent to x_1. Then since the degrees of x_2 and x_3 are the same and coincide with the rank k of f_v, we have that

$$|\,\mathrm{Lk}(x_2) \cup \mathrm{Lk}(x_3)| \geq k + 1.$$

This violates the minimality of the choice of v, since then Proposition 15.2.4 implies that the rank of f_v is at least $k + 1$. It follows that x_1 and x_3 are adjacent, and by symmetry we have that x_2 and x_3 are adjacent. By a straightforward induction, we have that $\{x_1, \ldots, x_m\}$ form a clique, and for all i, a vertex $y \notin \{x_1, \ldots, x_m\}$ adjacent to x_i is adjacent to all the vertices $\{x_1, \ldots, x_m\}$. Observe that if

$$v' = \sum_{i=1}^{m} \beta_i x_i^*$$

is another linear combination of dual vectors, then nonzero linear combinations w of v and v' also satisfy that the rank of f_w is equal to k. We set $V_{\min}$ to be a maximal vector subspace of V that contains v and such that for all $0 \neq w \in V_{\min}$, the rank of f_w is equal to k.

It is straightforward now to show that $V_{\min}$ is generated by $\{x_1^*, \ldots, x_\ell^*\}$, where $\{x_1, \ldots, x_\ell\}$ form an ℓ-clique such that

$$\mathrm{Lk}(x_i) \setminus \{x_j\} = \mathrm{Lk}(x_j) \setminus \{x_i\}$$

for all i and j.

We may now proceed as in Case 1 above, treating $\{x_1, \ldots, x_\ell\}$ as a single vertex, and replacing the subspace U by the subspace $V_{\min}$.

Subcase 2 We now have that x_1 and x_2 are not adjacent. If $m \geq 3$, then the argument in Subcase 1 above implies that x_3 is adjacent to neither x_1 nor x_2. We thus conclude that $\{x_1, \ldots, x_m\}$ form a totally disconnected subgraph of Γ, and $\mathrm{Lk}(x_i) = \mathrm{Lk}(x_j)$ for all i and j. We construct a vector space $V_{\min}$ as in Subcase 1 and conclude that it is generated by $\{x_1^*, \ldots, x_\ell^*\}$, where $\{x_1, \ldots, x_\ell\}$ form a totally disconnected graph and such that the links of any two vertices on this list coincide. We again reduce to Case 1.

□

Some remarks about Theorem 15.2.6 are in order. For one, one need only consider the degree one and degree two parts of the cohomology and not the full cohomology algebra, and this is not surprising since a graph is determined by its vertices and its edges, and a graph determines the corresponding right-angled Artin group. Second, in Case 2 of the proof, the vertices $\{x_1, \ldots, x_\ell\}$ are indistinguishable from each other, in the sense of graph automorphisms. That is, every permutation of $\{x_1, \ldots, x_\ell\}$ is realized by a graph automorphism of Γ, and therefore it is reasonable that one can treat this collection of vertices as a single vertex. Moreover, in the two subcases, $\{x_1, \ldots, x_\ell\}$ generates either an abelian or a free subgroup of $A(\Gamma)$. The full group of automorphisms of $\mathbb{Z}^\ell$ or of F_ℓ embeds in the group $\mathrm{Aut}(A(\Gamma))$ (cf. Sect. 15.3.5 below). Finally, in the proofs of Subcases 1 and 2, we obtain a vector space $V_{\min}$, which either comes from a clique or a totally disconnected subgraph. These two cases can be checked linear algebraically by whether the cup product pairing is trivial or not on $V_{\min}$.

15.3 Translating Between Group Theory and Combinatorics

In this section, we will describe some of the results and ideas that go into translation between the algebraic structure of $A(\Gamma)$ and the combinatorics of Γ. As we have remarked already, the abstract structure of $A(\Gamma)$ determines completely the nature of Γ, passing perhaps through cohomology (Theorem 15.2.6). We will seek clean, definitive results characterizing aspects of the combinatorial structure of Γ in terms of the algebra of $A(\Gamma)$. In the process, we will gain insight into both structures.

15.3.1 Elementary Properties

We begin with some of the first properties of graphs, and how these properties are reflected in $A(\Gamma)$. In Proposition 15.2.5, we have that pairing-connectedness of the triple (V, W, q) characterizes the connectedness of Γ. One can characterize the connectedness of Γ and its complement directly from the group theory of $A(\Gamma)$, without reference to the cohomology algebra, as follows.

Theorem 15.3.1 ([106]) *The group $A(\Gamma)$ splits as a nontrivial direct product if and only if Γ splits as a nontrivial join.*

Recall that a graph Γ splits as a nontrivial join if and only if the complement of Γ is disconnected. Dually, we have the following fact:

Theorem 15.3.2 ([26]) *The group $A(\Gamma)$ splits as a nontrivial free product if and only if Γ is disconnected.*

Both Theorems 15.3.1 and 15.3.2 are easy in one direction. If Γ is disconnected, then $A(\Gamma)$ admits a presentation of the form

$$A(\Gamma) = \langle V(\Gamma_1) \cup V(\Gamma_2) \mid \mathrm{R}_1 \cup \mathrm{R}_2 \rangle,$$

where Γ_1 and Γ_2 are nonempty and disjoint subgraphs of Γ, and where R_i only contains generators from Γ_i for $i \in \{1, 2\}$. It follows then immediately that $A(\Gamma) \cong A(\Gamma_1) * A(\Gamma_2)$.

If Γ splits as a join $\Gamma_1 * \Gamma_2$, then every vertex of Γ_1 is adjacent to every vertex of Γ_2. We have that $A(\Gamma_1)$ and $A(\Gamma_2)$ are subgroups of $A(\Gamma)$, and together generate the whole group. Moreover, they normalize each other and have trivial intersection (this last point is not completely trivial and requires some argument if one wishes to be pedantic, but we shall sweep it under the rug). It follows that $A(\Gamma_1)$ and $A(\Gamma_2)$ generate a direct product.

The converse directions are more complicated, and we outline the main ideas for the convenience of the reader.

Sketch of Proof of Theorem 15.3.2 We use the characterization of free products that follows from the work of Stallings [109, 110]. Let G be a finitely generated group with Cayley graph X. Recall that the set of *ends* of G is the inverse limit of $\pi_0(X \setminus K)$, where K ranges over all compact subgraphs of X. A group has zero, one, two, or infinitely many ends. As right-angled Artin groups are torsion-free (as follows from Proposition 15.2.1 for instance), we have that a right-angled Artin group $A(\Gamma)$ splits as a nontrivial free product if and only if it has infinitely many ends. It thus suffices to argue that a connected graph Γ yields a group with finitely many ends. For a graph with a single vertex, we have $A(\Gamma)$ is $\mathbb{Z}$ and hence has two ends. A straightforward argument shows that if G and H are both infinite groups then $G \times H$ has one end. Thus, we have that all nontrivial joins of graphs yield right-angled Artin groups with one end, and by induction we suppose that all connected graphs with at most n vertices yield groups with at most two ends. Let

$v \in V(\Gamma)$. Then there is a proper subgraph Λ of Γ such that

$$\Gamma = \Lambda \cup_{\mathrm{Lk}(v)} \mathrm{St}(v).$$

We have that $\mathrm{Lk}(v)$ is not empty since Γ is connected. Thus, we have that

$$A(\Gamma) = A(\Lambda) *_{A(\mathrm{Lk}(v))} A(\mathrm{St}(v)).$$

If Λ is connected then $A(\Gamma)$ is an amalgamated product of two finite-ended groups over an infinite subgroup (cf. [105]), whence one can prove directly that $A(\Gamma)$ is one-ended. If Λ is disconnected, then one can argue component-by-component of Λ to obtain the same conclusion. □

For Theorem 15.3.1, we require a basic result about the structure of centralizers of elements in $A(\Gamma)$. Let w be a word in the vertices of Γ and their inverses. We say that w is *reduced* if w cannot be shortened by applications of free reductions and moves of the form $[v_1^{\pm 1}, v_2^{\pm 1}]$ for $\{v_1, v_2\} \in E(\Gamma)$. We say that w is *cyclically reduced* if it remains reduced after allowing cyclic permutations of the letters occurring in w. It is true but not trivial that the moves of free reduction and commutation solve the word problem in right-angled Artin groups, and that cyclic reduction solves the conjugacy problem (see especially [39], cf. [31, 65, 114, 120]).

The *support* of w is written $\mathrm{supp}(w)$ and is defined to be the set of vertices which are required (possibly inverted) to express w. It is not completely trivial but true that the support of w is well-defined in the sense that for reduced words, $w_1 = w_2$ in $A(\Gamma)$ implies that $\mathrm{supp}(w_1) = \mathrm{supp}(w_2)$.

Theorem 15.3.3 ([106]) *Let* $1 \neq w \in A(\Gamma)$ *be cyclically reduced. Then the centralizer of* w *lies in* $\langle \mathrm{supp}(w) \cup \mathrm{Lk}(\mathrm{supp}(w)) \rangle$. *If the centralizer of* w *is not cyclic then either* $\mathrm{Lk}(\mathrm{supp}(w))$ *is nonempty, or* $\mathrm{supp}(w)$ *decomposes as a nontrivial join.*

Armed with Theorem 15.3.3, we can illustrate the other direction of Theorem 15.3.1.

Proof of Theorem 15.3.1 Suppose that $A(\Gamma) \cong G \times H$ for nontrivial groups G and H. Then since $A(\Gamma)$ is torsion-free, we have that every nontrivial element of $A(\Gamma)$ contains a copy of $\mathbb{Z}^2$ in its centralizer. Writing $V(\Gamma) = \{v_1, \ldots, v_n\}$, we have that $w = v_1 \cdots v_n$ is cyclically reduced and has noncyclic centralizer. Moreover, $\mathrm{Lk}(\mathrm{supp}(w)) = \emptyset$, so that Theorem 15.3.3 implies that $\mathrm{supp}(w) = \Gamma$ splits as a nontrivial join. □

Theorem 15.3.3 has several other important consequences that relate the algebraic structure of $A(\Gamma)$ to the combinatorics of Γ. First, we have the following.

Theorem 15.3.4 *The cohomological dimension of* $A(\Gamma)$ *coincides with the size of the maximal clique in* Γ.

Theorem 15.3.4 follows from standard ideas about cohomological dimension (cf. [30]), using the description of the Salvetti complex as a union of tori together with the fact that it is aspherical by Proposition 15.2.1. We have that the maximal dimensional cells in $\mathscr{S}(\Gamma)$ have the same dimension as the maximal size of a clique in Γ, say k. Moreover, this k-cell is the top dimensional cell in a subtorus of dimension k, which has nontrivial cohomology in degree k. Finally, the retraction $\mathscr{S}(\Gamma) \longrightarrow (S^1)^k$ implies that the degree k cohomology of $\mathscr{S}(\Gamma)$ is also nontrivial. It follows that k is also the cohomological dimension of $A(\Gamma)$.

The cohomological dimension and maximal clique size also describe the rank of a maximal abelian subgroup.

Theorem 15.3.5 *The maximal clique size of Γ coincides with the rank of a maximal abelian subgroup of $A(\Gamma)$.*

Theorem 15.3.5 is also a consequence of general facts about cohomological dimension. Clearly, if the maximal clique size of Γ is k then $A(\Gamma)$ contains a copy of $\mathbb{Z}^k$. Since $\mathscr{S}(\Gamma)$ is k-dimensional and aspherical, it follows that no cover of $\mathscr{S}(\Gamma)$ can have fundamental group $\mathbb{Z}^{k+1}$, so there are no abelian subgroups of rank exceeding k.

For another perspective, suppose Γ is connected and $G < A(\Gamma)$ is an abelian subgroup of rank $k \geq 2$. Conjugating if necessary, at least one nontrivial element of G is cyclically reduced, so that Theorem 15.3.3 implies that all nontrivial elements of G are supported on a subgraph J of Γ that splits as a nontrivial join. Writing $A(J) \cong A(J_1) \times A(J_2)$, we may restrict the projections $A(J) \longrightarrow A(J_i)$ for each i to G.

Now, suppose first that Γ has no triangles (i.e. 3-cliques). Then J_1 and J_2 cannot have any edges, since otherwise Γ would have a triangle. It follows then that $A(J_i)$ is free for $i \in \{1, 2\}$, and so the image of G in $A(J_i)$ is cyclic for each i. It follows that G has rank at most two. Thus, we may assume by induction that if the maximal clique size of Γ is at most $k \geq 2$ then the maximal abelian subgroup has rank at most k. Supposing Γ has maximal clique size $k + 1$, then J_1 and J_2 have maximal clique sizes $k_1 > 0$ and $k_2 > 0$, which satisfy $k_1 + k_2 \leq k + 1$. It follows by induction that the ranks of the images of G in $A(J_1)$ and $A(J_2)$ are at most k_1 and k_2, so that G has rank at most $k + 1$. This gives an alternate proof of Theorem 15.3.5.

The final elementary combinatorial property of graphs we will discuss is the maximal degree of a vertex. This property is essential in the theory of expander graphs, which will be discussed below. In the sequel we will use a different characterization of the maximal degree that is understood through cohomology, though the following is a significantly cleaner statement.

Proposition 15.3.6 *Let Γ be a graph and let d denote the maximum valence of a vertex of Γ. Then the rank of the centralizer of a nontrivial element of $A(\Gamma)$ is at most $d + 1$. Conversely, if for all elements $1 \neq g \in A(\Gamma)$ the centralizer of g has rank at most $d + 1$, then the maximum degree of a vertex of Γ is at most d.*

The proof of Proposition 15.3.6 is a fairly straightforward application of Theorem 15.3.3, and we leave it as an exercise for the reader.

15.3.2 k-Colorability

From the point of computational complexity, one of the most basic and difficult questions one can pose about a graph Γ is about its colorability. A *(vertex) coloring* of a graph Γ is a function $\kappa\colon V(\Gamma) \longrightarrow X$, where X is a finite set of *colors*, and where $\{v, w\} \in E(\Gamma)$ implies that $\kappa(v) \neq \kappa(w)$. A classical result of Brooks [43] says that the minimal size of X is at most the maximal degree of a vertex of Γ plus one. If Γ is not an odd length cycle or a clique then the bound can be improved to the maximal degree of a vertex. The minimal size of X is called the *chromatic number* of Γ, and we say that Γ is $|X|$-colorable.

A graph that is 2-colorable is called *bipartite*. Determining if a graph is bipartite is easy from a computational point of view, and can be accomplished by a sorting algorithm that runs in a period of time that is bounded by a polynomial in the size of the set of vertices. However, the problem of determining if a graph is 3-colorable is extremely difficult from a computational standpoint, and is *NP-complete* (see Sect. 15.3.7 below).

We remark that there is a related notion of *edge coloring*, which is a function $\epsilon\colon E(\Gamma) \longrightarrow X$ such that if $v \in V(\Gamma)$ is incident to both e_1 and e_2, then $\epsilon(e_1) \neq \epsilon(e_2)$. It is clear that the minimal size of X for a valid edge coloring is bounded below by the maximal degree of a vertex of Γ. A result of Vizing [43] shows that Γ admits an edge coloring with $|X|$ the maximal degree of a vertex of Γ plus one. Thus, giving sharp or almost sharp estimates on edge colorability of a graph is an essentially local problem, whereas determining vertex colorability is an essentially global problem.

Let Γ be a k-colorable graph. Choose a k-coloring of Γ, and add an edge to Γ for every pair of vertices with different colors, naming the result Λ. Observe that the vertices of Λ are partitioned as

$$V(\Lambda) = V_1 \cup \cdots \cup V_k,$$

where there are no edges between vertices in V_i for each i, and where for $i \neq j$, each vertex of V_i is adjacent to each vertex of V_j. It follows that $A(\Lambda)$ is a product of free groups, and that $A(\Lambda)$ is a quotient of $A(\Gamma)$. It turns out that these elementary considerations characterize k-colorable graphs.

Theorem 15.3.7 *Let Γ be a finite graph with N vertices. Then Γ is k-colorable if and only if there is a surjective map*

$$A(\Gamma) \longrightarrow \prod_{i=1}^{k} F_{n_i},$$

where for each i the group F_{n_i} is free of rank n_i, and where

$$\sum_{i=1}^{k} n_i = N.$$

We have already established the "only if" direction, which is easy. The reverse direction is somewhat more substantial, owing to the fact that the surjective homomorphism need not send vertex generators of Γ to a free factor of one of the free groups occurring on the right hand side.

Sketch of Proof of Theorem 15.3.7 We identify the product of free groups with $A(\Lambda)$, where Λ is a k-fold join of completely disconnected graphs, say $\{\Lambda_1, \ldots, \Lambda_k\}$. If $g \in A(\Lambda)$, then g can be written uniquely as a product of $g_1 \cdots g_k$, where $g_i \in A(\Lambda_i)$. One then shows that if $g = g_1 \cdots g_k$ and $h = h_1 \cdots h_k$ are elements of $A(\Lambda)$ that commute, then for each i, the images of g_i and h_i in the abelianization of $A(\Lambda_i)$ are rational multiples of each other.

Now, the surjective map $A(\Gamma) \longrightarrow A(\Lambda)$ induces an isomorphism

$$\phi \colon H_1(A(\Gamma), \mathbb{Q}) \longrightarrow H_1(A(\Lambda), \mathbb{Q}),$$

which can be expressed as a matrix A with respect to the vertex generators of both graphs. We will view the rows of A as expressions for $\phi(v)$ for $v \in V(\Gamma)$, in terms of the vertex generators of Λ. We arrange the columns so that the first $|V(\Lambda_1)|$ columns correspond to vertices of Λ_1, followed by the vertices of Λ_2, and so on.

Write $A = (A_1 \mid \cdots \mid A_k)$, where the columns of A_i correspond to the vertices of Λ_i, and therefore the column space of A_i has dimension $|V(\Lambda_i)|$. Note that since A is invertible, the row space of A_1 has dimension $|V(\Lambda_1)| = n_1$.

It is an exercise in linear algebra to show that there is an $n_1 \times n_1$ minor B_1 of A_1 and a $(N - n_1) \times (N - n_1)$ minor C of $(A_2 \mid \cdots \mid A_k)$ such that both B_1 and C are invertible.

By induction, one permutes the rows of A to obtain a block matrix $B = (B_1 \mid \cdots \mid B_k)$ such that the diagonal $n_i \times n_i$ blocks $\{C_1, \ldots, C_k\}$ of B are invertible. This row permutation is simply a permutation of the vertices of Γ. One defines a coloring of the vertices by setting $\kappa(v_i) = j$ if in the matrix expression B of ϕ, we have that the row $\phi(v_i)$ meets the block C_j. That is, the vertices corresponding to the first n_1 rows are assigned color 1, the next n_2 are assigned color 2, and so on.

To check that this is a valid coloring, suppose v and w are adjacent in Γ. Then $[v, w] = 1$ in $A(\Gamma)$. For each block B_i, we may consider the restriction of the rows $\phi(v)$ and $\phi(w)$ to the columns in B_i. In B_i, these two rows are rational multiples of each other. If v and w were assigned the same color then in some block B_i, the rows both meet B_i in the diagonal sub-block C_i. Since C_i is invertible, this is a contradiction. Thus, we see that adjacent vertices of Γ are assigned different colors, and so the coloring of Γ is valid. □

Unpacking the final check that κ is a valid coloring in the proof of Theorem 15.3.7, it is not difficult to see that in fact one can relax the condition that the homomorphism $A(\Gamma) \longrightarrow A(\Lambda)$ be surjective, and replace it with the condition that it be surjective on the level of rational homology. From a practical point of view, this is a useful observation. Indeed, checking that a linear map is surjective is relatively easy, but maps to direct products of free groups are much less well-behaved, since the subgroup structure of the latter is very complicated [97].

15.3.3 Hamiltonicity

In addition to computing the chromatic number of a finite graph, a classical NP-complete problem in graph theory is deciding whether a given connected graph admits a Hamiltonian cycle. Here, a *Hamiltonian cycle* is a circuit in Γ that visits every vertex of Γ exactly once. A graph that admits a Hamiltonian cycle is simply called Hamiltonian. Much like vertex colorability versus edge colorability, there is a notion of a circuit in Γ that traverses every edge exactly once, called an *Eulerian cycle*. It is a standard fact that a connected graph admits an Eulerian cycle if and only if each vertex has even degree. Thus, determining whether a graph admits an Eulerian cycle is a purely local question, and the existence of a Hamiltonian cycle is a global question, impervious to local methods. We direct the reader to [43] for background on Eulerian and Hamiltonian paths and cycles in graphs.

Let $(x_0, \ldots, x_n)$ denote a Hamiltonian cycle in Γ, and let

$$\{x_0^*, \ldots, x_n^*\} \subset V = H^1(A(\Gamma), F)$$

denote the corresponding dual classes. Proposition 15.2.4 implies that

$$x_i^* \smile x_{i+1}^* \neq 0$$

for all i, where the indices are considered cyclically modulo n. This is the fundamental observation when it comes to characterizing Hamiltonicity of Γ in terms of the intrinsic algebra of $A(\Gamma)$.

Let (V, W, q) be a triple consisting of a vector space V equipped with a vector-space-valued (i.e. W-valued) (anti)-symmetric bilinear pairing. We will assume that V is finite dimensional. We say that (V, W, q) is *Hamiltonian* if for all bases $\{v_0, \ldots, v_n\}$ of V, there is a permutation $\sigma \in S_{n+1}$ such that for all i, we have $q(v_{\sigma(i)}, v_{\sigma(i+1)}) \neq 0$.

Setting

$$V = H^1(A(\Gamma), F), \quad W = H^2(A(\Gamma), F), \quad q = \smile,$$

suppose that (V, W, q) is Hamiltonian. Then there is a basis $\{x_0^*, \ldots, x_n^*\}$ consisting of classes dual to the vertices of Γ. The Hamiltonicity of the triple immediately implies the existence of a permutation σ such that

$$x_{\sigma(i)}^* \smile x_{\sigma(i+1)}^* \neq 0$$

for all relevant indices, which immediately implies that Γ admits a Hamiltonian cycle.

Theorem 15.3.8 (See [55]) *Let Γ and (V, W, q) be as above. Then Γ admits a Hamiltonian cycle if and only if (V, W, q) is Hamiltonian.*

The reader may check as an easy exercise that the Hamiltonicity of a triple (V, W, q) implies that the triple is in fact pairing-connected, so that if (V, W, q) is Hamiltonian then Γ is automatically connected by Proposition 15.2.5.

Establishing Theorem 15.3.8 is tricky, and requires significantly more insight than Theorem 15.3.7, for instance. We will attempt to briefly convey the main ideas to the reader in the remainder of this subsection. The reader is directed to [55] for a full account.

In order to establish Theorem 15.3.8, it is clearly sufficient to show that if Γ is Hamiltonian then the triple (V, W, q) is also Hamiltonian. One may begin with the standard dual basis $\{x_0^*, \ldots, x_n^*\}$ for V and hope to bootstrap it to show that (V, W, q) is Hamiltonian. One can begin with a change of basis matrix A, which transforms $\{x_0^*, \ldots, x_n^*\}$ to a given basis $\{v_0, \ldots, v_n\}$ for V. We write $A = (a_i^j)$, where the subscript refers to the row and the superscript refers to the column of a given entry.

We leave it as an easy exercise for the reader to show the following:

Lemma 15.3.9 *The triple (V, W, q) is Hamiltonian if for all $A \in GL_{n+1}(A)$, there is a permutation $\sigma \in S_{n+1}$ such that for all $0 \le i \le n$, there exists a $0 \le j \le n$ such that*

$$A_i^j = \begin{pmatrix} a_{\sigma(i)}^j & a_{\sigma(i)}^{j+1} \\ a_{\sigma(i+1)}^j & a_{\sigma(i+1)}^{j+1} \end{pmatrix}$$

is invertible, where all indices are considered cyclically.

Lemma 15.3.9 gives rise to a natural definition that one can associate to matrices (which need not be invertible, or even square). The *two-row graph* $\mathscr{G}(A)$ of a matrix A is a graph whose vertices are the rows $\{r_0, \ldots, r_n\}$ of A, and whose columns are given by the relation $\{r_i, r_j\} \in E(\mathscr{G}(A))$ if the matrix

$$A_{i,j}^k = \begin{pmatrix} a_i^k & a_i^{k+1} \\ a_j^k & a_j^{k+1} \end{pmatrix}$$

is invertible for some k.

It is clear from Lemma 15.3.9 that (V, W, q) is Hamiltonian provided that $\mathscr{G}(A)$ is itself Hamiltonian for all suitable matrices A. To get a feel for $\mathscr{G}(A)$, the reader is encouraged to prove directly that $\mathscr{G}(A)$ is connected whenever A is invertible. The heart of the proof of Theorem 15.3.8 is the following:

Lemma 15.3.10 *Let A be an invertible matrix. Then $\mathscr{G}(A)$ is Hamiltonian.*

Lemma 15.3.10 is a curious fact in its own right, and its proof is fairly involved. Producing a Hamiltonian cycle directly in $\mathscr{G}(A)$ appears to be a difficult problem itself, and which has the feel of an NP-complete problem (though this is by no means a theorem). Thus, one needs to use more indirect methods to find a Hamiltonian cycle in $\mathscr{G}(A)$.

The key idea is to analyze block submatrices of a matrix A which consist of nonzero entries with one-dimensional row spaces. One can consider maximal such blocks, which give rise to a partition of the products of entries of A which contribute to the determinant of A, according to the standard Leibniz formula. Using certain symmetries, one can then argue that if no Hamiltonian cycle exists in $\mathscr{G}(A)$ then all summands in the determinant of A cancel in pairs, and hence the determinant of A is zero.

15.3.4 Graph Expanders

In this subsection, we leave behind individual graphs, and concentrate on families of graphs known as graph expanders. Graph expanders are sequences of connected graphs that are uniformly sparse and uniformly difficult to separate. Expander families find applications in a myriad of different fields, such as knot theory, spectral graph theory and spectral geometry, probabilistic computation, and network theory. We direct the reader to [3, 23, 24, 67, 89, 92–94] for references relevant to this section.

A sequence of finite graphs $\{\Gamma_i\}_{i\in\mathbb{N}}$ is called a *graph expander family* if the following conditions are satisfied:

(1) There is a d such that for all i, the maximum degree of a vertex of Γ_i is at most d.
(2) We have $|V(\Gamma_i)| \longrightarrow \infty$.
(3) The Cheeger constant of Γ_i is uniformly bounded away from zero, independently of i.

Here, the *Cheeger constant* of a graph Γ is defined by considering subsets $A \subset V(\Gamma)$ such that $|A| \leq |V(\Gamma)|/2$, and by looking at ∂A, which is defined to be the set of vertices of $V(\Gamma) \setminus A$ that are adjacent to a vertex of A. The *isoperimetric constant* of A is defined to be

$$c_A = \frac{|\partial A|}{|A|},$$

and the Cheeger constant c is the minimum of c_A as A ranges over all admissible subsets of $V(\Gamma)$. From this point of view, it is clear why the Cheeger constant measures the difficulty in separating Γ: in order to completely cut a set $A \subset V(\Gamma)$ out of Γ, one has to sever at least $c \cdot |A|$ edges. The reader may check that if $\{\Gamma_i\}_{i\in\mathbb{N}}$ forms a graph expander family then the Cheeger constant inequality implies that each graph in the family is connected.

By associating the standard cohomology triple

$$V_i = H^1(A(\Gamma_i), F),\ W_i = H^2(A(\Gamma_i), F),\ q_i = \smile,$$

some of the defining properties of graph expanders translate almost immediately. Namely, we have $|V(\Gamma_i)| \longrightarrow \infty$ if and only if $\dim V_i \longrightarrow \infty$, and Γ_i is connected if and only if (V_i, W_i, q_i) is q_i-pairing-connected.

The remaining conditions for defining graph expanders require some new ideas. The degree of a vertex is already characterized in terms of centralizers via Proposition 15.3.6 above. Since centralizers of elements are less transparently cohomological objects, we first translate this notion of degree into linear algebra. Let (V, W, q) be a vector space with a vector-space-valued bilinear pairing. If $\emptyset \neq S \subset V$ and B is a basis for V, we write

$$d_B(S) = \max_{s \in S} |\{b \in B \mid q(s, b) \neq 0\}|.$$

To get rid of the dependence on B, we set $d(S)$ to be the minimum of $d_B(S)$, taken over all possible bases. To get rid of the dependence on S, we set $d(V)$ to be the minimum of $d(S)$, taken over all S which span V. The quantity $d(V)$ is called the *q-valence* of V.

A reader who has understood the ideas in the proof of Theorem 15.2.6 will have no trouble proving the following fact:

Proposition 15.3.11 *Let (V, W, q) be the usual cohomological triple associated to $A(\Gamma)$, and let d be the maximum degree of a vertex of Γ. Then $d(V) = d$.*

It remains to properly define the Cheeger constant of the triple (V, W, q). Again, a reader who has absorbed the ideas in the proof of Theorem 15.2.6 could probably guess the definition. Let $Z \subset V$ be a vector space with $0 \neq \dim Z \leq (\dim V)/2$. We will write C for the orthogonal complement of Z, which is to say the set of vectors $v \in V$ such that $q(v, z) = 0$ for all $z \in Z$. The isoperimetric constant of Z is defined to be

$$c_Z = \frac{\dim V - \dim Z - \dim C + \dim(C \cap Z)}{\dim Z}.$$

The Cheeger constant c_V of the triple (V, W, q) is taken to be the infimum of c_Z as Z varies over all nonzero subspaces of V of dimension at most half of that of V.

Let $\{x_1, \ldots, x_n\}$ denote the vertices of Γ and $\{x_1^*, \ldots, x_n^*\}$ be the dual generators of $H^1(A(\Gamma), F)$. If $B \subset \{x_1, \ldots, x_n\}$, write B^* for the corresponding subset of $\{x_1^*, \ldots, x_n^*\}$. The following is an exercise for the reader:

Proposition 15.3.12 *Let $\emptyset \neq B \subset V$, and let $Z \subset V$ be generated by B^*. Then*

$$c_Z = \frac{|\partial B|}{|B|}.$$

Thus, the Cheeger constant of Γ is bounded below by c_V. *A priori*, there are many more subspaces of V than there are subgraphs of Γ, so that in principle c_V could be strictly smaller than the Cheeger constant of Γ.

Theorem 15.3.13 (See [53]) *Let* $\{\Gamma_i\}_{i\in\mathbb{N}}$ *be a sequence of graphs, and let*

$$\{(V_i, W_i, q_i)\}_{i\in\mathbb{N}}$$

be the corresponding cohomological triples. We have that $\{\Gamma_i\}_{i\in\mathbb{N}}$ *forms a family of expanders if and only if:*

(1) There is a d *such that the* q_i*-valence of* V_i *is bounded above by* d.
(2) We have $\dim V_i \longrightarrow \infty$.
(3) There is an $\epsilon > 0$ *such that for all* i*, we have* $c_{V_i} \geq \epsilon$.

As in the case of graph expander families, pairing-connectedness in the sequence of cohomological triples may be assumed but is actually redundant since it is a consequence of the Cheeger constant bound. An abstract sequence of triples $\{(V_i, W_i, q_i)\}_{i\in\mathbb{N}}$ is called a family of *vector space expanders* (not to be confused with *dimensional expanders*, cf. [23, 24, 93]). In light of the preceding discussion, in order to establish Theorem 15.3.13, it suffices to show that for each i, the Cheeger constant c_{V_i} coincides with the Cheeger constant of Γ_i. Unfortunately, the author does not know a conceptually simple proof of this fact. The proof given in [53] involves a rather technical sorting argument, and so we will not comment on it any further.

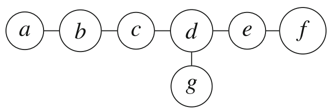
15.3.5 Graph Automorphisms

One of the most basic questions one can ask about a graph (and indeed about a relation) is how symmetric it is. Symmetry is measured by the richness of the automorphism group, and the smaller the size of the automorphism group, however it is measured, the less symmetric the object.

The automorphisms of graphs are of great interest in graph theory [17, 43, 58], and in complexity theory as well [7]. Many finite graphs are highly symmetric. For instance, the automorphism group of a k-clique is the full symmetric group on k letters. Many other graphs have no nontrivial automorphisms. For instance, take a path of length five, with vertices labeled linearly as $\{a, b, c, d, e, f\}$, and add another vertex g which is adjacent only to d. The resulting graph Γ has no nontrivial automorphisms, as is readily verified by an exhaustive check. See Fig. 15.1.

Observe that a nontrivial automorphism of a graph Γ gives rise to a non-inner automorphism of $A(\Gamma)$. Moreover, if $v \in V(\Gamma)$, then the function $v \mapsto v^{-1}$ extends to a non-inner automorphism of $A(\Gamma)$ via the identity on the remaining vertices. It

Fig. 15.1 A graph that has no nontrivial automorphisms



is easy to see that the group $\mathrm{Aut}(A(\Gamma))$, and in fact $\mathrm{Out}(A(\Gamma))$, contains a subgroup isomorphic to

$$\mathrm{Aut}(\Gamma) \ltimes (\mathbb{Z}/2\mathbb{Z})^{|V(\Gamma)|}.$$

Thus, if Γ admits a nontrivial automorphism, then $\mathrm{Out}(A(\Gamma))$ contains a nonabelian finite subgroup.

Theorem 15.3.14 (See [51]) *Let Γ be a finite simplicial graph. We have that Γ admits a nontrivial automorphism if and only if* $\mathrm{Out}(A(\Gamma))$ *contains a finite nonabelian subgroup.*

The "only if" direction follows from the discussion preceding Theorem 15.3.14. The converse is significantly harder and requires a more careful analysis of $\mathrm{Out}(A(\Gamma))$.

A result of M. Laurence ([90], cf. [106]) says that $\mathrm{Aut}(A(\Gamma))$ is generated by automorphisms of the following type.

(1) Vertex inversions.
(2) Graph automorphisms.
(3) Partial conjugations.
(4) Dominated transvections.

Graph automorphisms have already been discussed, and *vertex inversions* have been mentioned above as arising from the map $v \mapsto v^{-1}$ for some $v \in V(\Gamma)$. A *partial conjugation* is given by considering a vertex $v \in V(\Gamma)$ whose star $\mathrm{St}(v)$ separates Γ. The automorphism acts by conjugation by v on one component of $\Gamma \setminus \mathrm{St}(v)$ and by the identity on the remaining components of Γ.

To define *dominated transvections*, we say that a vertex $v \in V(\Gamma)$ *dominates* a vertex $w \in V(\Gamma)$ if $\mathrm{Lk}(w) \subset \mathrm{St}(v)$. Then, the map $w \mapsto wv$ extends to an automorphism of $A(\Gamma)$ via the identity on the remaining vertices. Domination is clearly a relation on vertices of Γ that can be determined from the combinatorics of Γ.

Sketch of Proof of Theorem 15.3.14 We suppose that Γ admits no nontrivial automorphisms. A theorem of Toinet [113] implies that if $\phi \in \mathrm{Aut}(A(\Gamma))$ has finite order then ϕ acts nontrivially on $H_1(A(\Gamma), \mathbb{Z}) \cong \mathbb{Z}^{|V(\Gamma)|}$. Thus, it suffices to consider the action of automorphisms on $H_1(A(\Gamma), \mathbb{Z})$, and the effect of partial conjugations is then trivial.

Next, one shows that if there is a cycle $\{v_1, v_2, \ldots, v_k, v_1\}_{k\geq 2}$ where v_i dominates v_{i+1} (with the indices considered cyclically), then Γ admits a nontrivial automorphism, specifically an automorphism that exchanges two vertices of Γ. It follows that no such cycles exist. We may therefore order the vertices of Γ in such a way that if $v_i < v_j$ then v_j cannot dominate v_i. If we then write the image of $\mathrm{Out}(A(\Gamma))$ in $\mathrm{GL}_n(\mathbb{Z})$ with respect to the corresponding ordered basis for $H_1(A(\Gamma), \mathbb{Z})$, the result is a group of upper triangular integer matrices. Such a group has only abelian finite subgroups (coming from diagonal matrices with entries ± 1).

Thus, if Γ has no automorphisms then $\mathrm{Out}(A(\Gamma))$ has only abelian finite subgroups. □

We remark that in the proof of Theorem 15.3.14, one of the key observations is that a graph with a domination cycle admits a nontrivial automorphism. The converse of this statement is false. The 5-cycle C_5 admits many automorphisms, but no two vertices dominate each other.

15.3.6 Some Further Entries in the Combinatorics–Algebra Dictionary

There are a number of other results relating the combinatorics of graphs to the algebraic structure of groups which we will not discuss in detail for the sake of space. We briefly mention two results appearing in [66]. Recall that a group G is *poly-free* if there is a finite length subnormal filtration of G by subgroups such that successive quotients are free. Hermiller–Šunić proved that a right-angled Artin group is always poly-free, and that the length of the poly-free filtration is bounded above by the chromatic number of the defining graph. In the same paper, they established that $A(\Gamma)$ is a semidirect product of two finitely generated free groups if and only if Γ is a tree or a *complete bipartite graph*, which is to say a join of two completely disconnected graphs. Moreover, for a connected graph Γ with at least two vertices, the poly-free length of $A(\Gamma)$ is exactly two if and only if there is a subset $D \subset V(\Gamma)$ such that no pair of elements of D spans an edge, and every circuit in Γ meets D in at least two vertices. It is an interesting direction for future research to investigate the relationship between the normal structure of $A(\Gamma)$ and the combinatorics of Γ, and it appears that this subject is largely unexplored.

15.3.7 Usefulness Beyond Group Theory and Combinatorics

The various correspondences between combinatorics of graphs and algebraic structures of groups have theoretical and practical applications beyond the structural framework of Question 15.1.1 and its refinements. Here, we record some specific examples.

15.3.7.1 Complexity of Problems in Combinatorial Group Theory

One of the main applications of the foregoing discussion is in the domain of complexity theory, which is hardly surprising in light of the fact that many computationally difficult problems (i.e. NP-complete problems, cf. [6, 56, 98]) are formulated in a finitistic way, with reference to only combinatorial structures.

Consider a right-angled Artin group $A(\Gamma)$, and a homomorphism

$$\phi\colon A(\Gamma) \longrightarrow F_{n_1} \times F_{n_2} \times F_{n_3},$$

where F_{n_i} denotes a free group of rank n_i, and where

$$n_1 + n_2 + n_3 = |V(\Gamma)|.$$

If Γ is specified (e.g. by a list of vertices and pairs of adjacent vertices) and ϕ is specified in terms of the image of each vertex of Γ with respect to a fixed free basis of each of the free group factors in the target of ϕ, then it is easy to check if ϕ is a homomorphism that is surjective on the level of first rational homology. Indeed, it suffices to check first that ϕ is well-defined, meaning that adjacent vertices in Γ are sent to commuting elements of $F_{n_1} \times F_{n_2} \times F_{n_3}$, which can be performed efficiently. The latter claim results from the fact that centralizers of elements in $F_{n_1} \times F_{n_2} \times F_{n_3}$ are straightforward to describe, and because the word problem is efficiently solvable. Then, one must check that ϕ is surjective on the level of first rational homology, which is an easy linear algebra problem. In light of Theorem 15.3.7, the data specifying the homomorphism ϕ forms a (short) certificate of the fact that Γ is 3-colorable. Since the 3-colorability of Γ and the existence of this homomorphism are equivalent, the problem of deciding whether such a homomorphism exists is NP-complete. To state this conclusion formally:

Proposition 15.3.15 *Let Γ be a finite graph with $|V(\Gamma)| = N$, and let*

$$\{F_{n_1}, \ldots, F_{n_k}\}$$

be free groups such that

$$\sum_{i=1}^{k} n_i = N.$$

Write $G = \prod_i F_{n_i}$.

(1) If $k = 2$ then the problem of deciding whether or not there exists a homomorphism $A(\Gamma) \longrightarrow G$ that is surjective on first rational homology is in P.
(2) If $k = 3$ then the problem of deciding whether or not there exists a homomorphism $A(\Gamma) \longrightarrow G$ that is surjective on first rational homology is NP-complete.
(3) The problem of finding the minimal k for which there exist free groups $\{F_{n_1}, \ldots, F_{n_k}\}$ as above and a homomorphism $A(\Gamma) \longrightarrow G$ that is surjective on first rational homology is NP-complete.

Finding explicit examples of NP-complete problems is always of interest in complexity theory, and given the profusion of them in graph theory, Proposition 15.3.15 is just a taste of the richness of the available theory arising in the context of groups.

15.3.7.2 Hamiltonicity Testing

Continuing in the theme of NP-complete problems, it is well-known that deciding if a finite graph admits a Hamiltonian path or Hamiltonian cycle is NP-complete, as we have mentioned above. The ideas surrounding Theorem 15.3.8 can be used to certify that certain graphs are not Hamiltonian, in a purely finitistic linear algebraic way.

To expand on this a bit, first note that the field over which cohomology is considered is arbitrary. In particular, we may assume that the underlying field is just the field with two elements. Under this assumption, all the relevant vector spaces become finite sets, and are hence amenable to combinatorial techniques.

Consider then the standard cohomological triple (V, W, q) for a right-angled Artin group $A(\Gamma)$. In order to show that Γ is not Hamiltonian, it suffices to find a single basis for V which witnesses the claim that (V, W, q) is not Hamiltonian. Thus, such a basis can be used as a short certificate that a graph contains no Hamiltonian circuit.

15.3.7.3 Linear Algebraic Detection of Graph Expanders

Considering cohomology with coefficients in a field with two elements allows a finitary and algebraic way to check if a sequence of graphs is a family of expanders. Moreover, it is shown in [53] that there are families of vector space expanders that do not arise from the cohomology of families of graph expanders. Thus, the theory of vector space expanders is *a priori* richer than the theory of graph expanders. Some practical applications of expanders can be found in [35, 59], for instance.

15.3.7.4 Interactive Proof Systems

Many interactive proof systems function as a way for a prover to demonstrate a proposition to a skeptical verifier. Using an unbiased random bit sent by the verifier, the prover sends a response that is conditioned on the value of the random bit. In this way, the verifier's ignorance of the prover's private information is balanced by the prover's ignorance of the value of the bit that will be sent by verifier, and this balance can be used to communicate the existence of knowledge without revealing its content. This is, for instance, the idea behind *zero-knowledge proof protocols*, in which the prover holds a certificate for an instance of an NP-complete problem, and convinces the verifier of the fact that she is in possession of a valid certificate without revealing the certificate itself. Any NP-complete problem can be used as a platform. Thus, linear algebraic versions of Hamiltonicity as in Theorem 15.3.8 and Proposition 15.3.15 are suitable for formulating a zero-knowledge proof protocol. A detailed explanation of a platform using Theorem 15.3.8 is given in [55]. For general background on interactive proofs and zero-knowledge proof protocols, we refer the reader to [6, 8, 22, 60, 100].

15.3.7.5 Group-Based Cryptosystems

Many cryptosystems rely on computational problems that are difficult to solve directly, which is why many modern cryptographic protocols assume $P \neq NP$. The theme of this section has been the translation of combinatorial properties of graphs, and especially computationally interesting ones, into algebraic language. This immediately suggests numerous potential group-based cryptosystems, a topic which has been developing rapidly in recent decades. Explicit cryptosystems using right-angled Artin groups as a platform have been proposed in [50], for example. Translating the graph homomorphism problem (which is NP-complete) into an instance of the subgroup homomorphism problem for right-angled Artin groups, one can formulate a secure authentication scheme, for instance. For further discussion of specific cryptosystems and for a biased sample of the literature, we direct the reader to [49, 51, 52, 75, 84, 99].

15.4 The Extension Graph and Its Properties

We now leave the world of the finite graph Γ and its relationship with $A(\Gamma)$, and turn to the (usually) infinite extension graph Γ^e. We recall that Γ^e is a development of Γ into a graph on which $A(\Gamma)$ acts by conjugation. So, we fix an identification of the vertices of Γ with generators for $A(\Gamma)$, set the vertices of Γ^e to be the collection of all conjugates of $V(\Gamma)$ by elements of $A(\Gamma)$, and set the edge relation to be commutation inside of $A(\Gamma)$. The reader will find that the ideas here, though still fundamentally relating combinatorics to algebra, are quite different from those in Sect. 15.3.

15.4.1 Basic Properties of the Extension Graph

Some properties of the extension graph are easy to prove. For instance:

Proposition 15.4.1 *The extension graph Γ^e is finite if and only if Γ is complete.*

Others are somewhat less obvious. We note some which will be useful in the sequel, and which otherwise will give the reader a better idea of how the extension graph functions.

Proposition 15.4.2 (See [78]) *The extension graph Γ^e enjoys the following properties:*

(1) The graph Γ^e is connected if and only if Γ is connected.
(2) The graph Γ^e is connected and of infinite diameter if and only if Γ is connected, has at least two vertices, and is not a join.

(3) The size of a maximal clique in Γ *and* Γ^e *coincide.*
(4) If Λ *is a subgraph of* Γ *then* Λ^e *is a subgraph of* Γ^e.
(5) The graph Γ^e *is k-colorable if and only if* Γ *is k-colorable.*

The proof of item (2) of Proposition 15.4.2 we will provide probably illustrates the diversity of methods that can be used in investigating right-angled Artin groups.

Sketch of Proof of Proposition 15.4.2, (2) Consider a collection of disjoint compact annuli $\{A_v \mid v \in V(\Gamma)\}$, one for each vertex of Γ. Glue two such annuli A_v and A_w together along a disk if the vertices v and w are *not* adjacent in Γ. We do this in such a way that the result is an orientable surface Σ with boundary. A key observation is that since Γ is not a join, its complement graph X is connected. Therefore, Σ is a connected surface. Since Σ was built out of at least two annuli, an easy Euler characteristic computation shows that Σ is of hyperbolic type (i.e. admits a complete hyperbolic metric of finite volume). We will name the core curves of the annuli in the construction $\{\gamma_1, \ldots, \gamma_n\}$.

The (isotopy class of the) homeomorphism of Σ given by cutting Σ open along γ_i and re-gluing with a full right-handed twist is called a (right-handed) *Dehn twist* about γ_i, and is denoted by T_i. Recall that the group of isotopy classes of (orientation preserving) homeomorphisms of Σ is called the *mapping class group* of Σ, and is written $\mathrm{Mod}(\Sigma)$ [47]. A result of the author [85] shows that there is an $N > 0$ such that for all $k \geq N$, the subgroup of $\mathrm{Mod}(\Sigma)$ generated by $\{T_1^k, \ldots, T_n^k\}$ is isomorphic to $A(\Gamma)$.

The surface Σ has an associated *curve graph* $\mathscr{C}(\Sigma)$, which is of infinite diameter. This curve graph consists of isotopy classes of embedded, essential, nonperipheral loops on Σ, with the edge relation being disjoint realization. There are certain mapping classes ψ which have the property that for any vertex c of $\mathscr{C}(\Sigma)$, the graph distance between c and $\psi^k(c)$ tends to infinity as k tends to infinity [112]. These mapping classes are called *pseudo-Anosov*, and are typical inside of $\mathrm{Mod}(\Sigma)$.

In particular, realizing $A(\Gamma) < \mathrm{Mod}(\Sigma)$ as above, there is an element $g \in A(\Gamma)$ whose realization as a mapping class is pseudo-Anosov. Moreover, the realization $A(\Gamma) < \mathrm{Mod}(\Sigma)$ is compatible with a realization of $\Gamma^e \subset \mathscr{C}(\Sigma)$. Specifically, if $v \in V(\Gamma)$ is associated to a Dehn twist about γ_i and if $h \in A(\Gamma)$ corresponds to the mapping class group ψ_h, then the vertex v^h is sent to $\psi_h(\gamma_i)$.

Now, since we have a map $\Gamma^e \longrightarrow \mathscr{C}(\Sigma)$ which respects the edge relation, general facts about graph homomorphisms imply that it cannot be distance increasing. Thus, if $d_{\mathscr{C}(\Sigma)}(\gamma_i, \psi_h^k(\gamma_i))$ tends to infinity then $d_{\Gamma^e}(v, v^{h^k})$ also tends to infinity. The conclusion now follows. □

It turns out that mapping class groups of surfaces are extremely useful tools for probing right-angled Artin groups, and that many of their properties can be paired analogously. This is a theme that will recur in this section, and we will comment more on it below.

15.4.2 The Extension Graph and Subgroups

One useful property of the extension graph, and for which it was developed in the first place, is that the extension graph classifies right-angled Artin subgroups of a right-angled Artin group. Classically, we know that subgroups of finitely generated free abelian groups are again free abelian (by the classification of finitely generated modules over a principal ideal domain) and subgroups of free groups are always free (by the Nielsen–Schreier Theorem). Since right-angled Artin groups interpolate between these two extremes, it is therefore a natural question whether (finitely generated) subgroups of right-angled Artin groups are again right-angled Artin groups, and if so what sorts of right-angled Artin groups they are.

It is not true that subgroups of right-angled Artin groups are again right-angled Artin groups. There are many different subgroups of right-angled Artin groups, ranging from surface groups [38, 107] to hyperbolic 3-manifold groups [1, 2, 117–119] to many arithmetic lattices in rank one Lie groups [18], all the way to groups with various exotic finiteness properties [19]. It is in fact known that every finitely generated subgroup of $A(\Gamma)$ is again a right-angled Artin group if and only if Γ has no subgraph isomorphic to a square or to a path of length three, by a result of Droms [45].

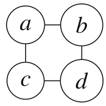
It is difficult to characterize all subgroups of right-angled Artin groups, even finitely presented ones (see [28]). Some general known facts are that every nonabelian subgroup of a right-angled Artin group contains a nonabelian free group by a result of Baudisch [10], and in fact any such subgroup surjects to a nonabelian free group by a result of Antolín–Minasyan [5]. A nonabelian subgroup of a right-angled Artin group must surject to $\mathbb{Z}^2$ [46, 87]. Solvable subgroups of right-angled Artin groups are automatically finitely generated and free abelian, by the Flat Torus Theorem [29].

Given the difficulty of understanding general subgroups of right-angled Artin groups, it is therefore interesting and natural to wonder which subgroups of $A(\Gamma)$ are of the form $A(\Lambda)$, and what sorts of graphs Λ can occur. To the author's knowledge, there is no clean, complete answer available, though the partial answers are satisfying and useful for many applications.

Theorem 15.4.3 *Let $\Lambda < \Gamma^e$ be a finite subgraph. Then there is an injective homomorphism $A(\Lambda) \longrightarrow A(\Gamma)$.*

The injection in Theorem 15.4.3 is quite explicit; one simply views vertices of Λ as elements in $A(\Gamma)$ and passes to a sufficiently high power. Theorem 15.4.3 first appeared in a paper of Kim and the author [78], though apparently this fact was already known to experts in combinatorial group theory. One approach to proving Theorem 15.4.3 does not require ideas beyond those that go into item (2) of Proposition 15.4.2. Once the extension graph has been embedded in the curve graph in a way that preserves both adjacency and non-adjacency, the author's result from [85] about powers of mapping classes applies and gives the desired result.

Fig. 15.2 The square



Unfortunately, Theorem 15.4.3 does not admit an easy converse. The first examples disproving the obvious naïve converse appeared in the work of Casals-Ruiz–Duncan–Kazachkov [33], and a large class of examples was produced by Kim and the author [81]. With some further assumptions on Γ, one can formulate a converse to Theorem 15.4.3.

Theorem 15.4.4 (See [78]) *Suppose* Γ *has no* 3*-cliques, and suppose that* $A(\Lambda) < A(\Gamma)$*. Then* Λ *is a subgraph of* Γ^e*.*

Theorem 15.4.4 is a corollary of a more general result, which is the most general converse to Theorem 15.4.3 that is known to the author.

Theorem 15.4.5 (See [78]) *Suppose that* $A(\Lambda) < A(\Gamma)$*. Then* Λ *is a subgraph of the clique graph* $(\Gamma^e)_k$*.*

The basic idea behind Theorem 15.4.5 is again to use mapping class groups, though it is significantly more complicated than Theorem 15.4.3 and Proposition 15.4.2. One builds certain "partial" pseudo-Anosov mapping classes in the image of $A(\Lambda)$ and builds an embedding of a larger graph X into Γ^e, which contains Λ in its clique graph. Incidentally, Theorem 15.4.5 has a natural analogue for mapping class groups: if a right-angled Artin group $A(\Gamma)$ embeds in a mapping class group $\mathrm{Mod}(\Sigma)$, then Γ embeds as a subgraph of $\mathscr{C}(\Sigma)_k$, the clique graph of the curve graph ([80], cf. [82]). We will avoid giving further details here.

Theorem 15.4.5 admits several other corollaries that can serve as converses to Theorem 15.4.3, and also allows one to prove many results that relate the combinatorics of Γ to the structure of $A(\Gamma)$. Given the conclusion of Theorem 15.4.5, we leave the following result (originally due to Kambites [76], who offered a combinatorial argument that is very different in flavor from the ideas expounded here) as an exercise for the reader:

Proposition 15.4.6 *Let* Γ *be a finite graph. Then* Γ *contains a square if and only if* $F_2 \times F_2 < A(\Gamma)$*.*

Here, by *square* we mean a graph with four vertices and a cyclic adjacency relation (Fig. 15.2).

15.4.3 A Characterization of Cographs via Right-Angled Artin Groups and the Geometry of the Extension Graph

An important class of graphs that occurs naturally in graph theory is the class of *cographs*, or P_4-free graphs (see [74, 103, 111] for some early references

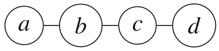


Fig. 15.3 The graph P_4

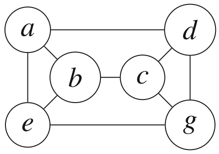


Fig. 15.4 The graph X_6

introducing cographs). These are simply the graphs that do not have the path P_4 of length three as a subgraph (Fig. 15.3).

Right-angled Artin groups on cographs can be characterized algebraically, and right-angled Artin groups provide a perspective on cographs that insight into one of their most fundamental properties, i.e. recursive definition.

Theorem 15.4.7 ([78]) *Let Γ be a finite connected graph. The following are equivalent:*

(1) The graph Γ has no (full) subgraph isomorphic to P_4.
(2) The graph Γ^e has no (full) subgraph isomorphic to P_4.
(3) The graph Γ is either a single vertex or splits as a nontrivial join.

Corollary 15.4.8 *The graph Γ is a cograph if and only if $A(\Gamma)$ does not contain a copy of $A(P_4)$.*

In particular, Theorem 15.4.7 shows that a right-angled Artin group cannot contain "hidden" copies of $A(P_4)$. If $A(\Gamma)$ contains $A(P_4)$ then one can decide simply from looking at the graph Γ. This is in contrast to other classes of graphs. For instance, $A(P_4)$ contains a copy of $A(P_5)$, where P_5 denotes the path of length four. Thus, there can be hidden copies of $A(P_5)$. For a more striking example, one may consider X_6, the complement graph of the hexagon, also known as the triangular prism. This graph contains no cycle C_5 of length 5, though Kim proved that $A(C_5) < A(X_6)$ [77]; also, C_5 is a subgraph of the extension graph X_6^e, and so Kim's result follows from Theorem 15.4.3 (Fig. 15.4).

We leave the proofs of Theorem 15.4.7 and Corollary 15.4.8 as an exercise for the reader, as they follow from Proposition 15.4.2 and some elementary combinatorial group theory considerations.

Let Γ be a connected graph such that Γ^e has finite diameter. By Theorem 15.4.7 (or even just by Proposition 15.4.2), the graph Γ splits as a nontrivial join. If Γ is a cograph and Λ is a join factor of Γ, then Λ must also be a cograph and hence Λ^e also has finite diameter, whence it follows that Λ must also split as a nontrivial join.

Let $\mathscr{K}_0$ denote a singleton vertex. For $i > 0$, we set $\mathscr{K}_{2i-1}$ to be the collection of all finite graphs obtained as (possibly trivial) joins of elements of $\mathscr{K}_{2i-2}$. We set $\mathscr{K}_{2i}$ to be the collection of all finite graphs obtained as disjoint unions of elements of $\mathscr{K}_{2i-1}$. Clearly for $i \le j$ we have $\mathscr{K}_i \subset \mathscr{K}_j$, and we set

$$\mathscr{K} = \bigcup_{i \ge 0} \mathscr{K}_i.$$

Clearly, if $\Gamma \in \mathscr{K}$ then Γ is a cograph. Conversely, the preceding remarks and an easy induction on $|V(\Gamma)|$ show that if Γ is a cograph then $\Gamma \in \mathscr{K}$. This coincides with the recursive description of cographs.

Since $\mathscr{K}$ is built up recursively, we can give the following characterization of $A(\Gamma)$ for $\Gamma \in \mathscr{K}$, which results immediately from the preceding discussion:

Corollary 15.4.9 *We have $\Gamma \in \mathscr{K}$ if and only if $A(\Gamma)$ is an element of the smallest class of groups that:*

(1) Contains $\mathbb{Z}$;
(2) Is closed under finite direct products;
(3) Is closed under finite free products.

For example, note that $\Gamma \in \mathscr{K}_0$ if and only if $A(\Gamma) \cong \mathbb{Z}$. We have $\Gamma \in \mathscr{K}_1$ if and only if $A(\Gamma) \cong \mathbb{Z}^n$ for some n. We have $\Gamma \in \mathscr{K}_2$ if and only if $A(\Gamma)$ is a free product of free abelian groups. A graph Γ lies in $\mathscr{K}_3$ if and only if $A(\Gamma)$ is a direct product of free products of free abelian groups. The following characterizes $\mathscr{K}_i$ for $i \le 3$:

Proposition 15.4.10 (See [83]) *A graph Γ lies in $\mathscr{K}_i$ for $i \le 3$ if and only if $A(\Gamma)$ has no subgroup isomorphic to $(F_2 \times \mathbb{Z}) * \mathbb{Z}$.*

As an aside, we note that the hierarchy $\mathscr{K}$ and the associated right-angled Artin groups is closely related to the theory of right-angled Artin group actions on the interval and on the circle. It turns out that $A(P_4)$ does not act faithfully by C^2 diffeomorphisms on I or S^1 [9], so any right-angled Artin group admitting such an action must have its underlying graph in $\mathscr{K}$. By a result of Kim and the author [83], a right-angled Artin group $A(\Gamma)$ admits a faithful C^2 action on I or S^1 if and only if it admits a faithful C^∞ such action, if and only if $\Gamma \in \mathscr{K}_3$.

15.4.4 More on the Geometry of the Extension Graph

As we have suggested in this section, and in particular in the discussion about Proposition 15.4.2, the extension graph of Γ plays a role analogous to that of the curve graph $\mathscr{C}(\Sigma)$ of a surface, with the role of the mapping class group in the latter context played by the group $A(\Gamma)$ in the former context.

The graph $\mathscr{C}(\Sigma)$ is very complicated in both its local and its global structure. One of the most important foundational results about the global structure of $\mathscr{C}(\Sigma)$ is a result of Masur and Minsky which asserts that $\mathscr{C}(\Sigma)$ is δ-*hyperbolic* [96], see also [57, 61]. That is, there is a $\delta \ge 0$ so that in any geodesic triangle in $\mathscr{C}(\Sigma)$, a δ-neighborhood of two of the sides of the triangle contains the third.

Perhaps the easiest example of an infinite diameter δ-hyperbolic metric space is an infinite diameter tree, which is 0-hyperbolic. There are many other examples of δ-hyperbolic spaces that are not trees, such the usual hyperbolic spaces. For most surfaces, the curve graph $\mathscr{C}(\Sigma)$ is far from being a tree; it has one end, whereas for example a locally finite tree that admits a proper and cocompact action by an infinite group will have at least two ends, as follows from Bass–Serre Theory [105].

The geometry of the extension graph is something in between the curve graph and a tree. To state a precise result, we need the notion of a quasi-isometry. Let $f\colon X \longrightarrow Y$ be a function between metric spaces. Then we say that f is a *quasi-isometry* if there are constants $\lambda \geq 1$ and $C \geq 0$ such that for all $x, z \in X$, we have

$$\frac{1}{\lambda} \cdot d_X(x, z) - C \leq d_Y(f(x), f(z)) \leq \lambda \cdot d_X(x, z) + C,$$

and where for all $y \in Y$ there exists an $x \in X$ such that

$$d_Y(f(x), y) \leq C.$$

Here, the distance functions are all interpreted in the relevant spaces. A quasi-isometry can be thought of a function that is bi-Lipschitz on a large scale. For instance, the integers equipped with the metric induced from the real line are quasi-isometric to the real line, and any two finite-diameter metric spaces are quasi-isometric to each other, but an infinite-diameter metric space is not quasi-isometric to a finite-diameter metric space.

The relation induced by quasi-isometry is an equivalence relation on metric spaces, and so one often speaks of the quasi-isometry class of a metric space. The quasi-isometry class of a finitely generated group is the quasi-isometry class of its Cayley graph, equipped with the graph metric; see [41] for more details, for example.

In coarse geometry, one often searches for properties of metric spaces that are invariant under quasi-isometry. Examples of such properties include δ-hyperbolicity and the number of ends.

A metric space is called a *quasi-tree* if it contains a 0-hyperbolic metric space in its quasi-isometry class. Whereas simplicial trees are 0-hyperbolic, the converse is not quite true: a geodesic metric space is 0-hyperbolic if and only if it is an $\mathbb{R}$-tree. We will not discuss $\mathbb{R}$-trees any further, since they are not necessary for our discussion. We specialize the definition of a quasi-tree slightly: if Γ is a graph equipped with the graph metric, we call it a quasi-tree if it contains a simplicial tree in its quasi-isometry class.

Theorem 15.4.11 (See [78]) *Let Γ be a connected graph. Then Γ^e is a quasi-tree, and is in particular δ-hyperbolic. More precisely:*

(1) If Γ splits as a nontrivial join, then Γ^e has finite diameter and is hence quasi-isometric to a point.

(2) If Γ does not split as a nontrivial join then Γ^e is quasi-isometric to a regular simplicial tree of countable degree.

More interesting than the mere description of the quasi-isometry type of the extension graph is the interaction between group elements in $A(\Gamma)$ and Γ^e. Here, the analogy between the mapping class group and $A(\Gamma)$ develops further, with the natural isometric action of $A(\Gamma)$ on Γ^e mirroring many of the properties of the natural isometric action of $\mathrm{Mod}(\Sigma)$ on $\mathscr{C}(\Sigma)$.

The classical *Nielsen–Thurston classification* [47, 112] says that a mapping class is either finite order, reducible (i.e. some power fixes the homotopy class of an essential nonperipheral loop on the surface Σ), or pseudo-Anosov. As discussed around Proposition 15.4.2, this lattermost type of mapping class is characterized by the fact that every orbit of its action on $\mathscr{C}(\Sigma)$ is unbounded. Finite order and reducible mapping classes are characterized by every orbit in $\mathscr{C}(\Sigma)$ being bounded (and in fact having a periodic point in $\mathscr{C}(\Sigma)$). Algebraically, a reducible mapping class has a copy of $\mathbb{Z}^2$ in its centralizer [21], whereas a pseudo-Anosov mapping classes have virtually cyclic centralizers [48].

Further insight into the action of $\mathrm{Mod}(\Sigma)$ is provided by a result of Bowditch [25], which says that the action of $\mathrm{Mod}(\Sigma)$ on $\mathscr{C}(\Sigma)$ is *acylindrical*. Acylindricity is a notion of proper discontinuity for group actions on non-proper metric spaces which are not properly discontinuous. Following Bowditch (cf. [86, 104]) we say that an action of a group G on a metric space X is acylindrical if for all $r > 0$ there exist constants R and N such that for all pairs of points $x, y \in X$ with $d(x, y) \geq R$, we have

$$|\{g \in G \mid d(gx, x), d(gy, y) \leq r\}| \leq N.$$

In other words, the r-quasi-stabilizer of R-separated points is uniformly finite. Bowditch showed that if X is a δ-hyperbolic graph and G acts isometrically and acylindrically on X then each $g \in G$ is either *elliptic* or *loxodromic*. The former of these means that some (equivalently every) orbit of G on X is bounded. A loxodromic element is characterized by having a positive asymptotic translation distance in X. Moreover, the asymptotic translation length is bounded away from zero by a constant that depends only on the hyperbolicity and acylindricity constants. The Nielsen–Thurston classification can be thus recast in terms of acylindricity: a mapping class is pseudo-Anosov if and only if it is loxodromic as an isometry of $\mathscr{C}(\Sigma)$.

For extension graphs, one has a picture that is analogous to curve graphs.

Theorem 15.4.12 (See [79]) *Let Γ be a connected graph with at least two vertices. The action of $A(\Gamma)$ on Γ^e is acylindrical. An element $1 \neq g \in A(\Gamma)$ is elliptic if and only if g is conjugate into a subgroup $A(J)$, where J is a subgraph of Γ that is a nontrivial join. Equivalently, g is elliptic if and only if its centralizer in $A(\Gamma)$ is noncyclic.*

An element $1 \neq g \in A(\Gamma)$ *is loxodromic if and only if its centralizer is cyclic. An element* g *is cyclically reduced and loxodromic if and only if* $\mathrm{supp}(g)$ *is not contained in a subgraph of* Γ *that splits as a nontrivial join.*

The join/non-join dichotomy for graphs and their associated right-angled Artin groups runs deep, and analogies between $A(\Gamma)$ and Γ^e with $\mathrm{Mod}(\Sigma)$ and $\mathscr{C}(\Sigma)$ are extensive. Many (but not all; see [88]) of the instances of these analogies can be and have been incorporated into the theory of hierarchically hyperbolic groups.

A further equivalence in Theorem 15.4.12 is given by a result of Behrstock–Charney [11], which asserts that a nontrivial element of $A(\Gamma)$ is loxodromic if and only if, when viewed as a deck transformation of the universal cover of the Salvetti complex $\mathscr{S}(\Gamma)$, it acts as a *rank one isometry*. That is, the corresponding deck group element has an axis that does not bound a half-plane (cf. [29]).

15.4.5 The Extension Graph as a Quasi-Isometry and Commensurability Invariant

A basic problem in geometric group theory is to sort groups into quasi-isometry classes. For right-angled Artin groups, the natural question is to decide when two right-angled Artin groups $A(\Gamma)$ and $A(\Lambda)$ are quasi-isometric. Much progress on understanding the quasi-isometric classification of right-angled Artin groups has been made, for instance by Behrstock–Neumann [12], Behrstock–Januszkiewicz–Neumann [13], Bestvina–Kleiner–Sageev [20], Huang [70], and Margolis [95] (see also [32, 71]). Thus, we can consider the following equivalence relation on finite graphs: Γ is equivalent to Λ if $A(\Gamma)$ and $A(\Lambda)$ are quasi-isometric to each other. Other than the cases we have cited, understanding this equivalence relation in full is still unresolved.

Certainly two right-angled Artin groups that are isomorphic to each other will be quasi-isometric to each other, and from Theorem 15.2.6, we know that if $A(\Gamma)$ and $A(\Lambda)$ are isomorphic to each other then Γ and Λ are isomorphic as graphs. There is yet another equivalence relation on finite graphs that is coarser than isomorphism and yet finer than quasi-isometry.

If $H < G$ are groups with G finitely generated and $[G : H] < \infty$, then with respect to any finite generating sets for G and H, the inclusion of H into G is a quasi-isometry on the level of Cayley graphs, as is readily verified. It follows that if G and H are finitely generated groups, and both G and H contain a finite index subgroup isomorphic to K, then G and H are quasi-isometric. In this case, we say that G and H are *commensurable*. Like quasi-isometry, commensurability is an equivalence relation on groups. It is well known that commensurability of groups is a strictly finer equivalence relation than quasi-isometry. For instance, one can take closed hyperbolic 3-manifolds whose volumes are not rational multiples of each other. Then, the corresponding fundamental groups are both quasi-isometric to hyperbolic space, but are not commensurable [57]. Even among right-angled Artin groups, commensurability is a strictly finer equivalence relation (see [34, 70]).

It is easy to produce pairs of non-isomorphic graphs which give rise to commensurable right-angled Artin groups. For instance, consider a graph Γ and $v \in V(\Gamma)$. There is a surjective homomorphism $A(\Gamma) \longrightarrow \mathbb{Z}/2\mathbb{Z}$ that sends v to the nontrivial element in $\mathbb{Z}/2\mathbb{Z}$ and sends the remaining vertices to the identity. It is an exercise in combinatorial group theory for the reader to prove that the kernel of this homomorphism is isomorphic to $A(\Lambda)$, where Λ is obtained by taking two copies of Γ and identifying them along $\mathrm{St}(v)$. If v is not central in $A(\Gamma)$ then it is easy to see that Λ and Γ fail to be isomorphic graphs, but $A(\Gamma)$ and $A(\Lambda)$ are clearly commensurable. This construction can be repeated *ad infinitum*, generally producing infinite families of non-isomorphic graphs whose associated right-angled Artin groups are all commensurable.

There are pairs of graphs which give rise to commensurable right-angled Artin groups, but for which a commensuration between them is less obvious. The reader is challenged to prove for themself that the groups $A(P_4)$ and $A(P_5)$ are commensurable, where as before P_4 and P_5 denote the paths of length three and length four respectively (cf. [34]). The fact that $A(P_4)$ and $A(P_5)$ are commensurable also shows that the extension graph is hopeless as a complete commensurability invariant. Again, the reader is encouraged to convince themself that the extension graphs of P_4 and P_5 are not isomorphic to each other. It turns out that in both cases, the corresponding extension graphs are trees, and what distinguishes them in their isomorphism type is the location of degree one vertices.

So, let us consider a connected graph Γ with no degree one vertices. In order to identify the extension graph algebraically and in an unambiguous way, it would help to be able to identify vertices and their conjugates, up to powers. For this, it helps to assume that Γ is connected, has no triangles, and has no squares. Under these assumptions, if v is a vertex of Γ then v contains a nonabelian free group in its centralizer. Conversely, suppose that $g \in A(\Gamma)$ has a nonabelian free group in its centralizer. Then, since Γ has no triangles and no squares, every nontrivial join in Γ is merely the star of a vertex of Γ, and the structure of such a star is the join of a single vertex and a completely disconnected graph. It follows that if g has a nonabelian free group in its centralizer, then g is conjugate to a nonzero power of a vertex generator of Γ. It follows that maximal cyclic subgroups of $A(\Gamma)$ whose centralizers contain nonabelian free groups are in bijection with conjugates of vertex generators of $A(\Gamma)$. Since the adjacency relation in Γ^e is just commutation in $A(\Gamma)$, we immediately obtain:

Theorem 15.4.13 (See [79]) *Let Γ be a finite connected graph with no degree one vertices, no triangles, and no squares. Then the extension graph Γ^e is a commensurability invariant for $A(\Gamma)$. That is, if $A(\Gamma)$ is commensurable with $A(\Lambda)$ then $\Gamma^e \cong \Lambda^e$.*

Incidentally, the analogy between right-angled Artin groups and mapping class groups persists here as well, since the curve graph can be obtained from the mapping class group in the same way that the extension graph is obtained from $A(\Gamma)$. Specifically, let T be a Dehn twist about a simple closed curve on Σ. Then T is centralized by two maximal rank torsion-free abelian subgroups of $\mathrm{Mod}(\Sigma)$ which intersect in a copy of $\mathbb{Z}$. This can be used to algebraically characterize a (nonzero power of a) Dehn twist as an element of $\mathrm{Mod}(\Sigma)$. A Dehn twist unambiguously identifies the homotopy class of a simple closed curve on Σ, and the adjacency relation in $\mathscr{C}(\Sigma)$ coincides with commutation of Dehn twists in $\mathrm{Mod}(\Sigma)$. Thus, the curve graph can be recovered algebraically from $\mathrm{Mod}(\Sigma)$. It follows in particular that automorphisms of $\mathrm{Mod}(\Sigma)$ induce automorphisms of $\mathscr{C}(\Sigma)$, a fact which can be used to prove various rigidity results (see [27, 72, 91], for instance).

As we have seen, we can have commensurable right-angled Artin groups with non-isomorphic extension graphs. It is also possible to have two right-angled Artin groups whose extension graphs are isomorphic and yet the groups are not quasi-isometric to each other (see Example 5.22 in [68]). So, there is similarly no hope that extension graphs form a complete quasi-isometry invariant for right-angled Artin groups.

Recall from Sect. 15.3.5 that a full set of generators for $\mathrm{Aut}(A(\Gamma))$ is known, and from the description of these generators, it is immediate that $\mathrm{Out}(A(\Gamma))$ is finite if and only if $\mathrm{Aut}(A(\Gamma))$ admits no nontrivial partial conjugations and no dominated transvections. Graphs for which $\mathrm{Out}(A(\Gamma))$ is finite can thus be identified through a finitary combinatorial analysis, since it suffices to check that there are no separating stars of vertices and no pairs of vertices where one dominates the other (see [37] for a discussion of the genericity of this phenomenon).

The following result was established by Huang [69]:

Theorem 15.4.14 *Suppose* Γ *is a graph for which* $\mathrm{Out}(A(\Gamma))$ *is finite. The following are equivalent:*

(1) The group $A(\Lambda)$ *is quasi-isometric to* $A(\Gamma)$.
(2) The group $A(\Lambda)$ *is isomorphic to a finite index subgroup of* $A(\Gamma)$.
(3) The graphs Λ^e *and* Γ^e *are isomorphic.*

Thus, in the case of finite groups of outer automorphisms, quasi-isometry, commensurability, and isomorphism of extension graphs are equivalent conditions to place on a right-angled Artin group. Here again, the analogy with mapping class groups persists. If two mapping class groups of surfaces are quasi-isometric, then except for some sporadic cases, the resulting mapping class groups are in fact isomorphic to each other [14]. Thus again excluding some sporadic cases, quasi-isometry, commensurability, and isomorphism of mapping class groups are equivalent. Finally, aside from some sporadic cases, isomorphism of curve graphs is equivalent to isomorphism of mapping class groups [108].

15.5 Further Directions

Much remains to be understood in the relationship between combinatorics and algebra via the lens of right-angled Artin groups. As the reader has certainly come to understand, it is not just some property of groups that one seeks to analogize a property of graphs; one wants it to be a clean and natural statement about groups that reflects the particular flavor of the property in question. Therefore, it is not likely one could produce a satisfactory omnibus result, since some subjective notions of beauty and philosophical considerations enter into the picture.

With these musings, we close by giving some particular open questions of interest. Some are well-known open problems, and we make no claim to having been the first to pose them.

Question 15.5.1 *What is the full quasi-isometric classification of right-angled Artin groups? What about the commensurability classification of right-angled Artin groups? What sorts of combinatorial objects serve as complete invariants for these equivalence relations?*

Some specific natural combinatorial properties we have not discussed are of interest in graph theory.

Question 15.5.2 *What algebraic property of $A(\Gamma)$ is equivalent to the planarity of Γ?*

Closely related to Question 15.5.2 is the problem of determining whether a graph Λ is a subdivision of a graph Γ by examining the relationship between the groups $A(\Lambda)$ and $A(\Gamma)$, which to the knowledge of the author is also open.

A graph is *self-complementary* if it is isomorphic to its complement graph. A singleton vertex is self-complementary, as are the path P_4 of length three and the cycle C_5 of length five. A question that is a particular favorite of the author is the following:

Question 15.5.3 *What algebraic property of $A(\Gamma)$ is equivalent to the statement that Γ is self-complementary?*

Following the remarks in Sect. 15.3.6 above and the results of [66], we have the following.

Question 15.5.4 *What is the relationship between the normal subgroup structure of $A(\Gamma)$ and the combinatorics of Γ?*

Finally, we have the following more open-ended question.

Question 15.5.5 *Is there a synthesis between the ideas in Sect. 15.3 and algebraic graph theory? How can one formulate spectral graph theory in terms of right-angled Artin groups?*

Some of the discussion in this survey is a step towards an answer to Question 15.5.5. For one, the Cheeger constant c of a graph can be viewed as a spectral invariant of a graph, as it controls the spectral gap of the graph via the Cheeger inequality due to Dodziuk and Alon–Milman (see [89] for a detailed discussion): if λ_2 is the second largest eigenvalue of a d–regular connected graph Γ then

$$\frac{1}{2}(d - \lambda_2) \leq c \leq \sqrt{2d(d - \lambda_2)}.$$

The content of Theorem 15.3.13 is that the Cheeger constant of a graph Γ can be read off from the cohomology algebra of $A(\Gamma)$. It is natural to ask how one might recover more information about the eigenvalues of the adjacency matrix of Γ from the group theory of $A(\Gamma)$.

We hope that this survey will encourage further investigations in these directions.

Acknowledgements The author is partially supported by an Alfred P. Sloan Foundation Research Fellowship, by NSF Grant DMS-1711488, and by NSF Grant DMS-2002596. The author thanks K. Ohshika and A. Papadopoulos for inviting him to write this survey, and to R. Flores for providing numerous invaluable comments on an earlier draft.

References

1. I. Agol, Criteria for virtual fibering. J. Topol. **1**(2), 269–284 (2008)
2. I. Agol, The virtual Haken conjecture. Doc. Math. **18**, 1045–1087 (2013). With an appendix by I. Agol, D. Groves, J. Manning
3. N. Alon, Eigenvalues and expanders. Combinatorica **6**, 83–96 (1986). *Theory of Computing* (Singer Island, FL, 1984)
4. J. Amorós, M. Burger, K. Corlette, D. Kotschick, D. Toledo, *Fundamental Groups of Compact Kähler Manifolds*. Mathematical Surveys and Monographs, vol. 44 (American Mathematical Society, Providence, RI, 1996)
5. Y. Antolín, A. Minasyan, Tits alternatives for graph products. J. Reine Angew. Math. **704**, 55–83 (2015)
6. S. Arora, B. Barak, *Computational Complexity: A Modern Approach* (Cambridge University Press, Cambridge, 2009)
7. L. Babai, Graph isomorphism in quasipolynomial time [extended abstract], in *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing* (ACM, New York, 2016), pp. 684–697
8. L. Babai, S. Moran, Arthur-Merlin games: a randomized proof system, and a hierarchy of complexity classes. J. Comput. Syst. Sci. **36**, 254–276 (1988). *17th Annual ACM Symposium on the Theory of Computing* (Providence, RI, 1985), pp. 254–276
9. H. Baik, S. Kim, T. Koberda, Unsmoothable group actions on compact one-manifolds. J. Eur. Math. Soc. **21**(8), 2333–2353 (2019)
10. A. Baudisch, Subgroups of semifree groups. Acta Math. Acad. Sci. Hungar. **38**(1–4), 19–28 (1981)
11. J. Behrstock, R. Charney, Divergence and quasimorphisms of right-angled Artin groups. Math. Ann. **352**(2), 339–356 (2012)
12. J.A. Behrstock, W.D. Neumann, *Quasi-isometric classification of graph manifold groups*, Duke Math. J. **141**(2), 217–240 (2008)
13. J.A. Behrstock, T. Januszkiewicz, W.D. Neumann, Quasi-isometric classification of some high dimensional right-angled Artin groups. Groups Geom. Dyn. **4**(4), 681–692 (2010)
14. J. Behrstock, B. Kleiner, Y. Minsky, L. Mosher, Geometry and rigidity of mapping class groups. Geom. Topol. **16**(2), 781–888 (2012)
15. J. Behrstock, M.F. Hagen, A. Sisto, Hierarchically hyperbolic spaces, I: curve complexes for cubical groups. Geom. Topol. **21**(3), 1731–1804 (2017)
16. J. Behrstock, M.F. Hagen, A. Sisto, Hierarchically hyperbolic spaces II: combination theorems and the distance formula. Pac. J. Math. **299**(2), 257–338 (2019)
17. L.W. Beineke, R.J. Wilson (eds.), *Topics in Algebraic Graph Theory, Encyclopedia of Mathematics and Its Applications*, vol. 102 (Cambridge University Press, Cambridge, 2004)

18. N. Bergeron, F. Haglund, D.T. Wise, Hyperplane sections in arithmetic hyperbolic manifolds. J. Lond. Math. Soc. (2) **83**(2), 431–448 (2011)
19. M. Bestvina, N. Brady, Morse theory and finiteness properties of groups. Invent. Math. **129**(3), 445–470 (1997)
20. M. Bestvina, B. Kleiner, M. Sageev, The asymptotic geometry of right-angled Artin groups. I. Geom. Topol. **12**(3), 1653–1699 (2008)
21. J.S. Birman, A. Lubotzky, J. McCarthy, Abelian and solvable subgroups of the mapping class groups. Duke Math. J. **50**(4), 1107–1120 (1983)
22. M. Blum, How to prove a theorem so no one else can claim it, in *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Berkeley, Calif., 1986)* (Amer. Math. Soc., Providence, RI, 1987), pp. 1444–1451
23. J. Bourgain, Expanders and dimensional expansion. C. R. Math. Acad. Sci. Paris **347**(7–8), 357–362 (2009)
24. J. Bourgain, A. Yehudayoff, Expansion in $SL_2(\mathbb{R})$ and monotone expanders. Geom. Funct. Anal. **23**(1), 1–41 (2013)
25. B.H. Bowditch, Tight geodesics in the curve complex. Invent. Math. **171**(2), 281–300 (2008)
26. N. Brady, J. Meier, Connectivity at infinity for right angled Artin groups. Trans. Am. Math. Soc. **353**(1), 117–132 (2001)
27. T.E. Brendle, D. Margalit, Commensurations of the Johnson kernel. Geom. Topol. **8**, 1361–1384 (2004)
28. M. Bridson, On the subgroups of right-angled Artin groups and mapping class groups. Math. Res. Lett. **20**(2), 203–212 (2013)
29. M.R. Bridson, A. Haefliger, *Metric Spaces of Non-positive Curvature, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*, vol. 319 (Springer, Berlin, 1999)
30. K.S. Brown, *Cohomology of Groups, Graduate Texts in Mathematics*, vol. 87 (Springer, New York, 1982)
31. P. Cartier, D. Foata, *Problèmes combinatoires de commutation et réarrangements*. Lecture Notes in Mathematics, No. 85 (Springer, Berlin, 1969)
32. M. Casals-Ruiz, Embeddability and quasi-isometric classification of partially commutative groups. Algebr. Geom. Topol. **16**(1), 597–620 (2016)
33. M. Casals-Ruiz, A. Duncan, I. Kazachkov, Embedddings between partially commutative groups: two counterexamples. J. Algebra **390**, 87–99 (2013)
34. M. Casals-Ruiz, I. Kazachkov, A. Zakharov, On commensurability of right-angled Artin groups I: RAAGs defined by trees of diameter 4. Rev. Mat. Iberoam. **35**(2), 521–560 (2019)
35. D.X. Charles, K.E. Lauter, E.Z. Goren, Cryptographic hash functions from expander graphs. J. Cryptol. **22**(1), 93–113 (2009)
36. R. Charney, An introduction to right-angled Artin groups. Geom. Dedicata **125**, 141–158 (2007)
37. R. Charney, M. Farber, Random groups arising as graph products. Algebr. Geom. Topol. **12**(2), 979–995 (2012)
38. J. Crisp, B. Wiest, Embeddings of graph braid and surface groups in right-angled Artin groups and braid groups. Algebr. Geom. Topol. **4**, 439–472 (2004)
39. J. Crisp, E. Godelle, B. Wiest, The conjugacy problem in subgroups of right-angled Artin groups. J. Topol. **2**(3), 442–460 (2009)
40. M.W. Davis, The cohomology of a Coxeter group with group ring coefficients. Duke Math. J. **91**(2), 297–314 (1998)
41. P. de la Harpe, *Topics in Geometric Group Theory*. Chicago Lectures in Mathematics (University of Chicago Press, Chicago, IL, 2000)
42. A. Díaz, A. Ruiz, A. Viruel, Cohomological uniqueness of some p-groups. Proc. Edinb. Math. Soc. **56**(2), 449–468 (2013)
43. R. Diestel, *Graph Theory*, 5th edn. Graduate Texts in Mathematics, vol. 173 (Springer, Berlin, 2017)
44. C. Droms, Isomorphisms of graph groups. Proc. Am. Math. Soc. **100**(3), 407–408 (1987)

45. C. Droms, Subgroups of graph groups. J. Algebra **110**(2), 519–522 (1987)
46. G. Duchamp, D. Krob, The lower central series of the free partially commutative group. Semigroup Forum **45**(3), 385–394 (1992)
47. B. Farb, D. Margalit, *A Primer on Mapping Class Groups*. Princeton Mathematical Series, vol. 49 (Princeton University Press, Princeton, NJ, 2012)
48. A. Fathi, F. Laudenbach, V. Poénaru (eds.), *Travaux de Thurston sur les surfaces* (Société Mathématique de France, Paris, 1991). Reprint of *Travaux de Thurston sur les surfaces* (Soc. Math. France, Paris, 1979) [MR0568308 (82m:57003)]. Astérisque No. 66–67 (1991)
49. B. Fine, D. Kahrobaei, G. Rosenberger (eds.), *Computational and Combinatorial Group Theory and Cryptography*. Contemporary Mathematics, vol. 582 (American Mathematical Society, Providence, RI, 2012)
50. R. Flores, D. Kahrobaei, Cryptography with right-angled Artin groups (2016). Preprint, arXiv:1610.06495
51. R. Flores, D. Kahrobaei, T. Koberda, Algorithmic problems in right-angled Artin groups: complexity and applications. J. Algebra **519**, 111–129 (2019)
52. R. Flores, D. Kahrobaei, T. Koberda, A cryptographic application of the Thurston norm. Int. J. Comput. Math. Comput. Syst. Theory **5**(1), 15–24 (2020)
53. R. Flores, D. Kahrobaei, T. Koberda, Expanders and right-angled Artin groups. J. Topol. Anal. (2020). to appear
54. R. Flores, D. Kahrobaei, T. Koberda, An algebraic characterization of k–colorability. Proc. Am. Math. Soc. **149**(5), 2249–2255 (2021)
55. R. Flores, D. Kahrobaei, T. Koberda, Hamiltonicity via cohomology of right-angled Artin groups. Linear Algebra Appl. **631**, 94–110 (2021).
56. M.R. Garey, D.S. Johnson, *Computers and Intractability* (W. H. Freeman, San Francisco, CA, 1979). A Guide to the Theory of NP-Completeness, A Series of Books in the Mathematical Sciences
57. É. Ghys, P. de la Harpe (eds.), *Sur les groupes hyperboliques d'après Mikhael Gromov*. Progress in Mathematics, vol. 83 (Birkhäuser, Boston, MA, 1990). Papers from the Swiss Seminar on Hyperbolic Groups held in Bern, 1988
58. C. Godsil, G. Royle, *Algebraic Graph Theory*. Graduate Texts in Mathematics, vol. 207 (Springer, New York, 2001)
59. O. Goldreich, R. Impagliazzo, L. Levin, R. Venkatesan, D. Zuckerman, Security preserving amplification of hardness, in *31st Annual Symposium on Foundations of Computer Science, Vol. I, II (St. Louis, MO, 1990)* (IEEE Comput. Soc. Press, Los Alamitos, CA, 1990), pp. 318–326
60. O. Goldreich, S. Micali, A. Wigderson, Proofs that yield nothing but their validity, or All languages in NP have zero-knowledge proof systems. J. Assoc. Comput. Mach. **38**(3), 691–729 (1991)
61. M. Gromov, Hyperbolic groups, in *Essays in Group Theory*. Math. Sci. Res. Inst. Publ., vol. 8 (Springer, New York, 1987), pp. 75–263
62. M. Gromov, *Metric Structures for Riemannian and Non-Riemannian Spaces*. Progress in Mathematics, vol. 152 (Birkhäuser, Boston, MA, 1999). Based on the 1981 French original, With appendices by M. Katz, P. Pansu, S. Semmes, Translated from the French by Sean Michael Bates
63. M.F. Hagen, Weak hyperbolicity of cube complexes and quasi-arboreal groups. J. Topol. **7**(2), 385–418 (2014)
64. A. Hatcher, *Algebraic Topology* (Cambridge University Press, Cambridge, 2002)
65. S. Hermiller, J. Meier, Algorithms and geometry for graph products of groups. J. Algebra **171**(1), 230–257 (1995)
66. S. Hermiller, Z. Šunić, Poly-free constructions for right-angled Artin groups. J. Group Theory **10**(1), 117–138 (2007)
67. S. Hoory, N. Linial, A. Wigderson, Expander graphs and their applications. Bull. Am. Math. Soc. (N.S.) **43**(4), 439–561 (2006)

68. J. Huang, Quasi-isometry classification of right-angled Artin groups II: several infinite out cases (2016). Preprint, arXiv:1603.02372
69. J. Huang, Quasi-isometric classification of right-angled Artin groups I: the finite out case. Geom. Topol. **21**(6), 3467–3537 (2017)
70. J. Huang, Commensurability of groups quasi-isometric to RAAGs. Invent. Math. **213**(3), 1179–1247 (2018)
71. J. Huang, B. Kleiner, Groups quasi-isometric to right-angled Artin groups. Duke Math. J. **167**(3), 537–602 (2018)
72. N.V. Ivanov, Automorphism of complexes of curves and of Teichmüller spaces. Int. Math. Res. Notices **14**, 651–666 (1997)
73. C. Jensen, J. Meier, The cohomology of right-angled Artin groups with group ring coefficients. Bull. Lond. Math. Soc. **37**(5), 711–718 (2005)
74. H.A. Jung, On a class of posets and the corresponding comparability graphs. J. Combinatorial Theory Ser. B **24**(2), 125–133 (1978)
75. D. Kahrobaei, V. Shpilrain (eds.), *Algorithmic Problems of Group Theory, Their Complexity, and Applications to Cryptography*. Contemporary Mathematics, vol. 633 (American Mathematical Society, Providence, RI, 2015)
76. M. Kambites, On commuting elements and embeddings of graph groups and monoids. Proc. Edinb. Math. Soc. (2) **52**(1), 155–170 (2009)
77. S. Kim, Co-contractions of graphs and right-angled Artin groups. Algebr. Geom. Topol. **8**(2), 849–868 (2008)
78. S. Kim, T. Koberda, Embedability between right-angled Artin groups. Geom. Topol. **17**(1), 493–530 (2013)
79. S. Kim, T. Koberda, The geometry of the curve graph of a right-angled Artin group. Int. J. Algebra Comput. **24**(2), 121–169 (2014)
80. S. Kim, T. Koberda, An obstruction to embedding right-angled Artin groups in mapping class groups. Int. Math. Res. Not. **14**, 3912–3918 (2014)
81. S. Kim, T. Koberda, Anti-trees and right-angled Artin subgroups of braid groups. Geom. Topol. **19**(6), 3289–3306 (2015)
82. S. Kim, T. Koberda, Right-angled Artin groups and finite subgraphs of curve graphs. Osaka J. Math. **53**(3), 705–716 (2016)
83. S. Kim, T. Koberda, Free products and the algebraic structure of diffeomorphism groups. J. Topol. **11**(4), 1054–1076 (2018)
84. K.H. Ko, S.J. Lee, J.H. Cheon, J.W. Han, J. Kang, C. Park, *New Public-Key Cryptosystem Using Braid Groups*. Advances in cryptology—CRYPTO 2000 (Santa Barbara, CA), Lecture Notes in Comput. Sci., vol. 1880 (Springer, Berlin, 2000), pp. 166–183
85. T. Koberda, Right-angled Artin groups and a generalized isomorphism problem for finitely generated subgroups of mapping class groups. Geom. Funct. Anal. **22**(6), 1541–1590 (2012)
86. T. Koberda, What is...an acylindrical group action?. Notices Am. Math. Soc. **65**(1), 31–34 (2018)
87. T. Koberda, Actions of right-angled Artin groups in low dimensions, in *Handbook of Group Actions*, ed. by L. Ji, A. Papadopoulos, S.-T. Yau, Chap 6, vol. V. ALM 48 (International Press, Boston; Higher Education Press, Beijing, 2020), pp. 227–260
88. T. Koberda, J. Mangahas, S.J. Taylor, The geometry of purely loxodromic subgroups of right-angled Artin groups. Trans. Am. Math. Soc. **369**(11), 8179–8208 (2017)
89. E. Kowalski, *An Introduction to Expander Graphs*. Cours Spécialisés [Specialized Courses], vol. 26 (Société Mathématique de France, Paris, 2019)
90. M.R. Laurence, A generating set for the automorphism group of a graph group. J. Lond. Math. Soc. (2) **52**(2), 318–334 (1995)
91. C.J. Leininger, D. Margalit, Abstract commensurators of braid groups. J. Algebra **299**(2), 447–455 (2006)
92. A. Lubotzky, *Discrete Groups, Expanding Graphs and Invariant Measures*. Modern Birkhäuser Classics (Birkhäuser, Basel, 2010). With an appendix by J. D. Rogawski, Reprint of the 1994 edition

93. A. Lubotzky, E. Zelmanov, Dimension expanders. J. Algebra **319**(2), 730–738 (2008)
94. A. Lubotzky, R. Phillips, P. Sarnak, Ramanujan graphs. Combinatorica **8**(3), 261–277 (1988)
95. A. Margolis, Quasi-isometry classification of right-angled Artin groups that split over cyclic subgroups. Groups Geom. Dyn. **14**(4), 1351–1417 (2020)
96. H.A. Masur, Y.N. Minsky, Geometry of the complex of curves. I. Hyperbolicity. Invent. Math. **138**(1), 103–149 (1999)
97. K.A. Mihaĭlova, The occurrence problem for direct products of groups. Dokl. Akad. Nauk SSSR **119**, 1103–1105 (1958)
98. M.L. Minsky, *Computation: Finite and Infinite Machines* (Prentice-Hall, Englewood Cliffs, NJ, 1967). Prentice-Hall Series in Automatic Computation.
99. A. Myasnikov, V. Shpilrain, A. Ushakov, *Group-Based Cryptography*. Advanced Courses in Mathematics (CRM Barcelona, Birkhäuser, Basel, 2008)
100. A. Rosen, *Concurrent Zero-Knowledge: With Additional Background and a Foreword by Oded Goldreich (Information Security and Cryptography)* (Springer, Berlin, 2006)
101. A. Ruiz, A. Viruel, Cohomological uniqueness, Massey products and the modular isomorphism problem for 2 -groups of maximal nilpotency class. Trans. Am. Math. Soc. **365**(7), 3729–3751 (2013)
102. L. Sabalka, On rigidity and the isomorphism problem for tree braid groups. Groups Geom. Dyn. **3**(3), 469–523 (2009)
103. D. Seinsche, On a property of the class of n-colorable graphs. J. Comb. Theory Ser. B **16**, 191–193 (1974)
104. Z. Sela, Acylindrical accessibility for groups. Invent. Math. **129**(3), 527–565 (1997)
105. J.-P. Serre, *Arbres, amalgames*, SL_2 (Société Mathématique de France, Paris, 1977). Avec un sommaire anglais, Rédigé avec la collaboration de Hyman Bass, Astérisque, No. 46
106. H. Servatius, Automorphisms of graph groups. J. Algebra **126**(1), 34–60 (1989)
107. H. Servatius, C. Droms, B. Servatius, Surface subgroups of graph groups. Proc. Am. Math. Soc. **106**(3), 573–578 (1989)
108. K.J. Shackleton, Combinatorial rigidity in curve complexes and mapping class groups. Pac. J. Math. **230**(1), 217–232 (2007)
109. J. Stallings, On torsion-free groups with infinitely many ends. Ann. Math. (2) **88**, 312–334 (1968)
110. J. Stallings, *Group Theory and Three-Dimensional Manifolds*. A James K. Whittemore Lecture in Mathematics given at Yale University, 1969, Yale Mathematical Monographs, 4 (Yale University Press, New Haven, 1971)
111. D.P. Sumner, Dacey graphs. J. Aust. Math. Soc. **18**, 492–502 (1974)
112. W.P. Thurston, On the geometry and dynamics of diffeomorphisms of surfaces. Bull. Am. Math. Soc. (N.S.) **19**(2), 417–431 (1988)
113. E. Toinet, Conjugacy p-separability of right-angled Artin groups and applications. Groups Geom. Dyn. **7**(3), 751–790 (2013)
114. L. VanWyk, Graph groups are biautomatic. J. Pure Appl. Algebra **94**(3), 341–352 (1994)
115. A. Viruel, Homotopy uniqueness of BG_2. Manuscripta Math. **95**(4), 471–497 (1998)
116. A. Viruel, Mod 3 homotopy uniqueness of BF_4. J. Math. Kyoto Univ. **41**(4), 769–793 (2001)
117. D.T. Wise, Research announcement: the structure of groups with a quasiconvex hierarchy. Electron. Res. Announc. Math. Sci. **16**, 44–55 (2009)
118. D.T. Wise, The structure of groups with a quasiconvex hierarchy (2011)
119. D.T. Wise, From riches to raags: 3-manifolds, right-angled Artin groups, and cubical geometry, in *CBMS Regional Conference Series in Mathematics*, vol. 117 (Published for the Conference Board of the Mathematical Sciences, Washington, DC, 2012)
120. C. Wrathall, The word problem for free partially commutative groups. J. Symb. Comput. **6**(1), 99–104 (1988)

Index

K. Ohshika, A. Papadopoulos (eds.), *In the Tradition of Thurston II*,
https://doi.org/10.1007/978-3-030-97560-9

R

S